

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

C. J. Lintott¹, K. W. Willett², S. P. Bamford³, K. L. Masters⁴, B. D. Simmons^{1,5}, etc.

¹*University of Oxford, UK*

²*School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA*

³*University of Nottingham, UK*

⁴*University of Portsmouth, UK*

⁵*Yale University, New Haven, USA*

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

ABSTRACT

Galaxy Zoo 2 data release paper.

Key words: galaxies

1 INTRODUCTION

2 PROJECT DESCRIPTION

3 DEBIASING THE DATA

3.1 Using Stripe 82

One possible method for debiasing the full GZ2 sample is to use the data from Stripe 82 to establish the unbiased level. Normal-depth images for Stripe 82 are selected in precisely the same manner as the main GZ2 sample, with a magnitude limit of $r_{\text{petro}} < 17.77$. The fiber size and redshift range are also identically chosen, resulting in a comparison sample of roughly 10% the size of the GZ2 original + extra samples.

Galaxies in Stripe 82 have both “normal-depth” images as well as deeper images created by co-adding exposures from the runs that SDSS did in the region prior to DR7 (Annis et al. 2011). With increased sensitivity to fainter magnitudes, this should mean that galaxies at higher redshifts will have their morphological classifications altered, presumably in the direction of higher accuracy. The spectroscopic incompleteness fraction is similar for the GZ2 original + extra (14%) and Stripe 82 normal depth (17%) samples, but is twice as high for the galaxies in the Stripe 82 coadded data (35%).

The hypothesis is that the change in vote fraction as a function of galaxy metadata (z, R_{50}, m_R, μ) would have the same effect on classifying GZ2 main sample galaxies at higher redshift in the same R_{50}, m_R, μ bins.

Firstly, we need to verify that the distribution of votes for the main sample is similar enough to the Stripe 82 normal-depth (with $r < 17.0$) that the same bias correction will apply for both. Table 1 shows the tasks in the GZ2 question tree and the mean weighted vote fraction for each response. Overall, the mean values for both the main and Stripe 82 galaxies are very similar, with almost all responses varying by $< 10\%$ between the two samples. The only excep-

tions 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 1 shows that the weighted vote fractions also behave similarly as a function of redshift, particularly in the $0.03 < z < 0.08$ 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. Similarly, the coadded data shows higher fractions of galaxies with bars (Task 03) and for possessing visible spiral structure (Task 04). A possible cause for this is that the new image pipeline in the coadded data allows viewers to see faint features or disks. The average seeing in the coadded data improves from 1.4” to 1.1” (Annis et al. 2011), which may need to be corrected for in the Stripe 82 comparisons.

What needs to be done:

- **Yes** Verify that the normal-depth Stripe 82 vote fractions are similar to the original + extra samples.
- **They do, even accounting for the magnitude cut.** See whether co-added images follow a different distribution, esp. as function of z, R_{50}, M_r, μ .
- **Only difference is a moderate increase in spiral fraction for disk galaxies in coadd1 data. The results of K-S tests on the weighted vote fractions are confusing. Very similar distributions (by eye) have very low probabilities according to the K-S test.** Decide which of the two co-added samples is more reliable for debiasing efforts
- optional: *Do a PCA on the normal vs. coadded versions. Figure out which of the four parameters. and in what weighted combination, produces the most scatter in the data.*

- How far out in redshift can we correct the normal data using coadd? Need enough galaxies per bin in M_r , R_{50} , z space *for each task* in the Stripe 82 data to establish the baseline. B09 set this at thirty galaxies per bin.

- Figure out how the vote fractions are going to be adjusted once a bias has been identified.

- Apply corrections to the vote fractions in the GZ2 main sample. See if the results are flat for certain tasks (and over what redshift range it applies)

4 THE CATALOG

5 COMPARISON OF GZ2 TO OTHER CLASSIFICATION METHODS

- Nair & Abraham (2010a)
- Huertas-Company et al. (2011)
- CAS studies (Conselice 2003, 2006)
- EFIGI (Baillard et al. 2011)

The main purpose of this section is to perform preliminary comparisons between the weighted, but not debiased data from GZ2 available in early 2012. This analysis will need to be repeated to account for biases in the data; hopefully with the groundwork laid here, it will be faster to repeat it later.

5.1 Nair & Abraham

Nair & Abraham (2010a, ; hereafter NA10) catalogued 14,034 galaxies from the SDSS DR4. The galaxy redshifts range from $0.01 < z < 0.1$, and go down to an extinction-corrected apparent magnitude limit of $g < 16$. Overlap with the GZ2 sample is substantial; joining on their objIDs, 12,480 galaxies are found in both catalogs. This comprises 89.9% of the NA10 catalog, but only 4.5% of the total GZ2 sample. The GZ2 sample is deeper ($r < 17$), spans a wider redshift range ($0.0005 < z < 0.25$), and contains more recent data releases (DR7) than the NA10 catalog.

Nair & Abraham (2010a) used classifications by a single professional (P. Nair) to carry out morphological classifications. She determined RC3 T-Types for the entire sample through visual inspection of g -band images, covering each source twice. There is no discussion as to the procedure if T-Types changed between her first and second classifications.

In addition to the T-Types, she also classified various “fine structure” morphological features in each galaxy. It is not stated how many times these classifications were performed. These include:

- bars (strong, weak, intermediate, ansae, “peanut”, nuclear, and/or unsure)
- rings (nuclear, inner, outer)
- lenses [regions of constant surface brightness; NOT gravitational lenses] (inner, outer)
- pairs of objects (close, projected, adjacent, overlapping, + flags for second object type)
- interaction (none, disturbed, warp, shells, short tail, medium tail, long tail, bridge)
- tails (number)

Several of the fine structure features can be directly compared to the GZ2 data: bars, rings, and pairs/interactions.

5.1.1 Bars

NA10 detect 2537 barred galaxies, which is 18% of their total sample. For objects with T-Types later than E/S0, this rises to 25% of the sample. This is much lower than the fraction found in local galaxies ($\sim 60\%$), but similar to the fraction found by (Masters et al. 2011) for the bar fraction in disk, face-on galaxies from GZ2 (29.4%).

Two parameters can be set that reduce the number of galaxies in the overlap between the samples, but result in a cleaner cut for comparisons. The first uses the Masters et al. (2011) cut for face-on galaxies ($\log(a/b) < 0.3$ using the `expab_r` parameter in the Galaxy table). The second is to only look at galaxies with at least 10 classifications for task 03 (*bar present?*) in GZ2. The reduced face-on sample has 7,121 galaxies from the original 12,480. All trends described below hold generally for both the full overlapping samples and the cleaner face-on sub-sample.

Of the objects NA10 identify as barred, 2348/2537 (93%) are objects in GZ2. Figure 2 shows correlations for face-on galaxies between the bar fractions of the NA10 and GZ2 catalogs. Galaxies in NA10 are classified by a single person, and so the presence of a bar is indicated with a series of flags. Bars in a galaxy can be classified according to 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 2 shows that the raw GZ2 weighted bar fraction (WBF) very closely agrees with the NA10 fraction of barred galaxies for each GZ2 bin. The two quantities are not identical; the x-axis plots galaxies with *individual classifications* at whatever fraction of votes chose a bar. The y-axis shows that for galaxies in that WBF bin from GZ, the average bar classification is similar to the weighted vote totals. The distribution shows a Spearman’s rho of $\rho = 0.984$, and closely follows a 1:1 relationship between the two lines. If supported by debiased GZ2 data, this means that the aggregate votes of Zooites closely reproduce the expert classification of NA10.

In the top right, Figure 2 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 either extrema, although there are a large number of galaxies in GZ2 with a fraction of 0.0 than one would expect. Possession of a bar is less straightforward; while the frequency of NA10 bars does increase with GZ2 fraction, 32% of NA-barred galaxies lie below the canonical GZ2 WBF of 0.5. Conversely, there are very few false positives (5.5% of non-barred NA10 galaxies) above this value.

In the bottom left of Figure 2, the distribution of GZ2 WBF as a function of NA10 bar strength is plotted. The distribution for all bars is the same as shown in the top right, increasing with GZ2 WBF. There is a clear difference in GZ classification between the three sets of bars; interestingly, 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

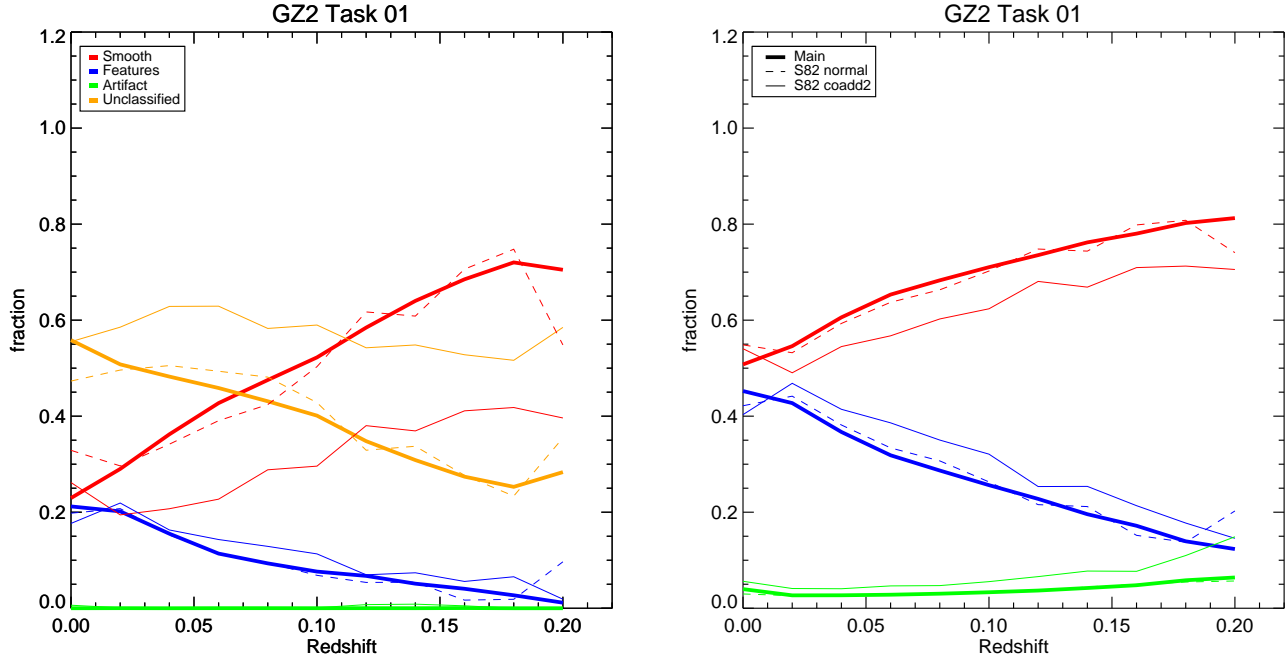


Figure 1. GZ2 weighted vote fractions for Task 01 (*smooth, features/disk, or star/artifact?*) as a function of spectroscopic redshift. The left graph shows the fraction of galaxies for which a category exceeded a threshold of 0.8, including galaxies which had no answer above the threshold. The right shows the mean of the vote fractions, weighted by the total number of responses to the task for each galaxy. Data are shown for the GZ2 original + extra (thick), Stripe 82 normal-depth (thin dotted), and Stripe 82 co-add depth (thin solid) samples. Stripe 82 data is only for galaxies with $r < 17.0$, the same magnitude limit applied to the GZ2 main sample.

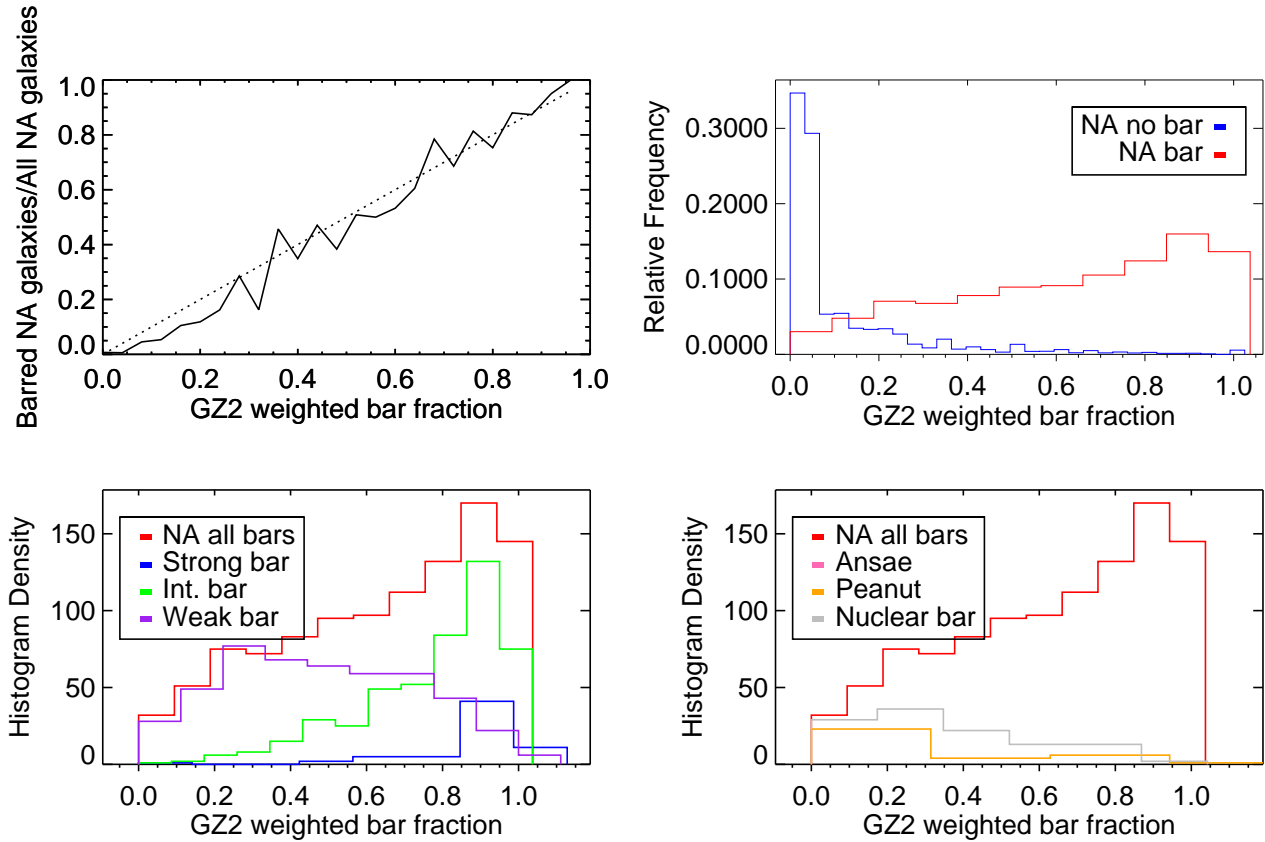


Figure 2. NA10 bar classifications compared to GZ2. Data are for the 7,121 galaxies which are face-on ($\log(a/b) < 0.3$) and with 10 or more GZ2 bar classifications.

barred population have high GZ2 WBFs, with 83% of strong bars and 56% of intermediate bars above a bar fraction of 0.8. Those numbers increase to 98% and 90%, respectively, if the criterion of 0.5 for the GZ2 WBF is used (Masters et al. 2011). Weakly-barred galaxies have only 13% of their GZ2 WBFs above 0.8 and 47% above 0.5.

NA10 identify three other fine-structure features related to bars: ansae, peanuts, and nuclear bars. None of the three correlate strongly with a high GZ2 WBF, with most galaxies in all three features at less than 0.5. Nuclear bars are the only feature that overlaps with the NA10 bar 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 face-on subsample of galaxies, due to the axial cut.

In the full GZ2 sample (original + extra), 87,651 galaxies had at least 10 responses to the question “*Is there a sign of a bar feature?*” - 22.1% of them had a WBF greater or equal to 0.5. In the clean face-on sample, this rises to 28.7%. This is in very 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).

In the published literature, (Masters et al. 2011) mention the NA10 catalog and note that both papers agree with a bar fraction depending on morphology with a minimum near the division between the blue and red sequences. There is no detailed comparison between the galaxies in the two samples.

5.1.2 Rings

The NA10 catalog also classifies ring galaxies in their sample. They include three basic types of rings in the catalog (Buta & Combes 1996). 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 six different options to choose from (ring, lens, disturbed, irregular, other, merger), the weighted ring fractions is less straightforward than for a binary tree.

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, 8.2% of galaxies have at least 10 votes for a ring, while 22% (10%) have a weighted fraction higher than 0.5 (0.8). In both catalogs, selecting only face-on galaxies makes no significant changes in the percentage of galaxies identified as having a ring.

In the top-left of Figure 3, the distribution of the number of GZ2 votes for a ring in face-on galaxies is shown, both for the total sample and for galaxies classified by NA10 as having a ring. The distributions grow closer as the number of “Yes” votes increases. The top-right panel of Figure 2 shows the CDF for the number of ring votes. Among all galaxies with at least 15 “Yes” votes, for example, $\sim 90\%$ of those galaxies are also identified by NA10 as having a ring; almost all of these are inner or outer rings.

In contrast, the weighted ring fraction (WRF) from the GZ2 votes is not a good match to the ring classifications of NA10. Half of all galaxies have a WRF of 0.0, indicating no votes for a ring-like structure in the image. For ringed galaxies identified by NA, the number of WRF=0.0 objects decreases dramatically, but results a generally flat distribution of ring fractions. No single cut on WRF is a good proxy for the NA10 classifications; at all WBF above 0.5, for example, only $\sim 45 - 65\%$ of the GZ2 ringed galaxies are identified as rings in NA10.

There is some evidence indicating that GZ2 classifications are sensitive only to certain types of rings. NA10 galaxies with nuclear rings, for example, have a large number of galaxies with no GZ2 ring votes. Several causes are possible: since nuclear rings are smaller, they are more difficult to discern in low surface brightness or bulge-dominated galaxies. In addition, the button in the GZ2 tree intended to show an example of a ring has a center dot (presumably a galactic bulge) surrounded by a ring. This could reasonably represent either an inner or outer ring, but would presumably not be associated with the intra-bulge nuclear rings by an untrained user. The bottom-right panel of Figure 3 shows that the number of NA10 galaxies with inner and/or outer rings does rise with WRF, but is still only 55% successful at WRF > 0.8.

5.1.3 Mergers/interacting galaxies

Galaxies in GZ2 can be labeled as a “merger” under the task “*Anything odd?*” NA10 classify possible mergers in two ways: by identifying pairs of objects in an image, and by identifying interacting galaxies. 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). If tidal tails are present, there is an additional flag counting the number of tails.

In the NA10 catalog, 22.3% of galaxies are labeled as paired; of these, the majority (72%) are close pairs. Interacting galaxies are a much smaller subset, comprising 7% of the NA10 sample. In the GZ2 catalog, only 3.6% (1.8%) of galaxies have a merger weighted fraction above 0.5 (0.8). If the vote totals for a possible merger are used instead, 5% (3%) of galaxies have at least 5 (10) votes for a merger. The large fraction of NA10 galaxies suggests that this is less likely to be a good proxy for a merger as identified by the GZ2 votes; interacting galaxies may be a more promising proxy.

Figure 4 shows the distributions of NA10 paired and interacting galaxies. Most galaxies in the overlapping sample have no votes for a merger; the same trends are seen for both pairs and interactions. If using the raw number of votes as a cutoff, GZ2 galaxies stabilize at an 80% match rate to the NA10 paired galaxies for 8 or more merger votes. The trend for interacting galaxies continues to increase with higher numbers of votes, only matching the 80% level at 25 or more merger votes, for which less than 100 galaxies are included.

Using the weighted merger fraction instead of the raw vote counts does not show a strong correlation with the NA10 classifications. Interestingly, both show peaks if a weighted merger fraction cutoff of 0.5 is chosen. The peak of

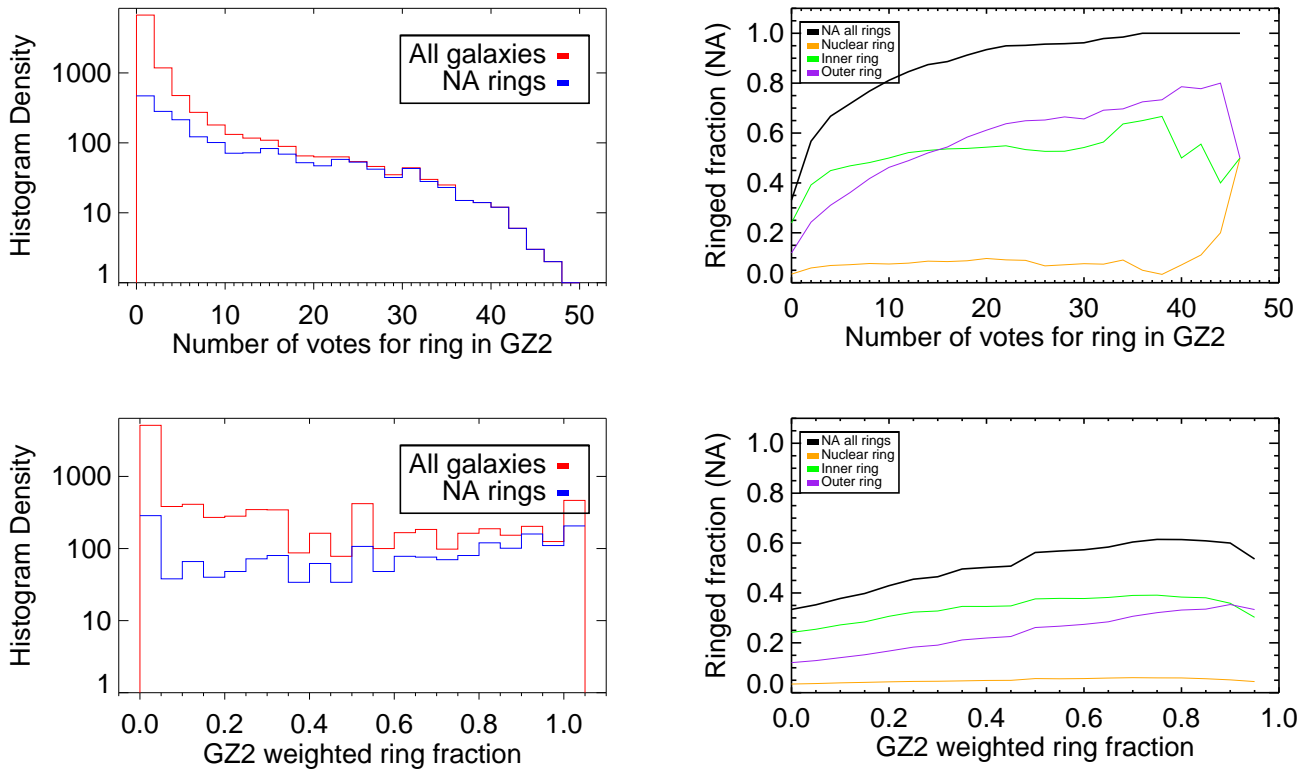


Figure 3. NA10 ring classifications compared to GZ2. Data are for the 9,746 galaxies in both samples which are face-on ($\log(a/b) < 0.3$).

the NA10 fraction, however, is significantly lower if matching on the weighted merger fraction; only 70% for paired galaxies and 40% for interacting galaxies. This shows no difference depending on the type of NA10 pair. Within the NA10 interacting classes, all actually show a decrease in number for a higher weighted merger fraction; the only exception is galaxies classified as bridges, the fraction of which increases slightly above a weighted merger fraction of 0.5.

Similar to the results above, there is no strong correlation between either the GZ2 weighted merger fraction or number of merger votes and the number of NA-identified tidal tails.

5.1.4 T-types

There has been no published discussion in the literature comparing large-scale morphologies of the NA10 and GZ catalogs. Nair & Abraham (2010a) was published after the first GZ1 results, (Lintott et al. 2008), but prior to the formal data release paper (Lintott et al. 2011). Huertas-Company et al. (2011) do compare automated classifications to both NA10 and GZ1, finding good agreement with both; this obliquely suggests that the GZ2 and NA10 classifications will also be consistent.

The left panel of Figure 5 shows the percentage of galaxies identified as having either a disk or features from the first question in the GZ2 tree, color-coded by their NA10 T-Types. There is a clear separation in the GZ2 fractions for galaxies classified as E vs. those with T-Types Sa and later. Late-type galaxies, including S0's, have a median weighted

fraction of the “features or disk question” of 0.796, with a standard deviation of 0.29. This distribution is bimodal, with one peak near 0.95 and a second at 0.1. Breaking down the late-type galaxies into more specific Hubble classifications, the late-type galaxies with low GZ2 feature votes are found to be primarily lenticular (S0; T-Type = -3 to 0) galaxies. If only galaxies with T-Types Sa or later are considered, the peak at lower GZ2 fractions disappears. The median GZ2 weighted fraction for these galaxies is 0.88, with a standard deviation of 0.23. The highest GZ2 weighted fraction for an elliptical galaxy in NA10 is 0.741; therefore, any cut above this limit includes *exclusively* identified by NA10 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 extremely well.

Since there are very few objects identified as stars or artifacts in the first GZ2 question, the weighted fraction for smooth galaxies is approximately $f_{smooth} = (1 - f_{features\ or\ disk})$. Elliptical galaxies (T-Type = -5) have a median weighted fraction of the “smooth” question of 0.86, with a standard deviation of 0.07. The GZ2 votes for the NA10 ellipticals are much more sharply peaked than the late-type galaxies, lacking the long tail seen even for 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 having features that are not edge-on disks, GZ2 users then vote on whether the galaxy

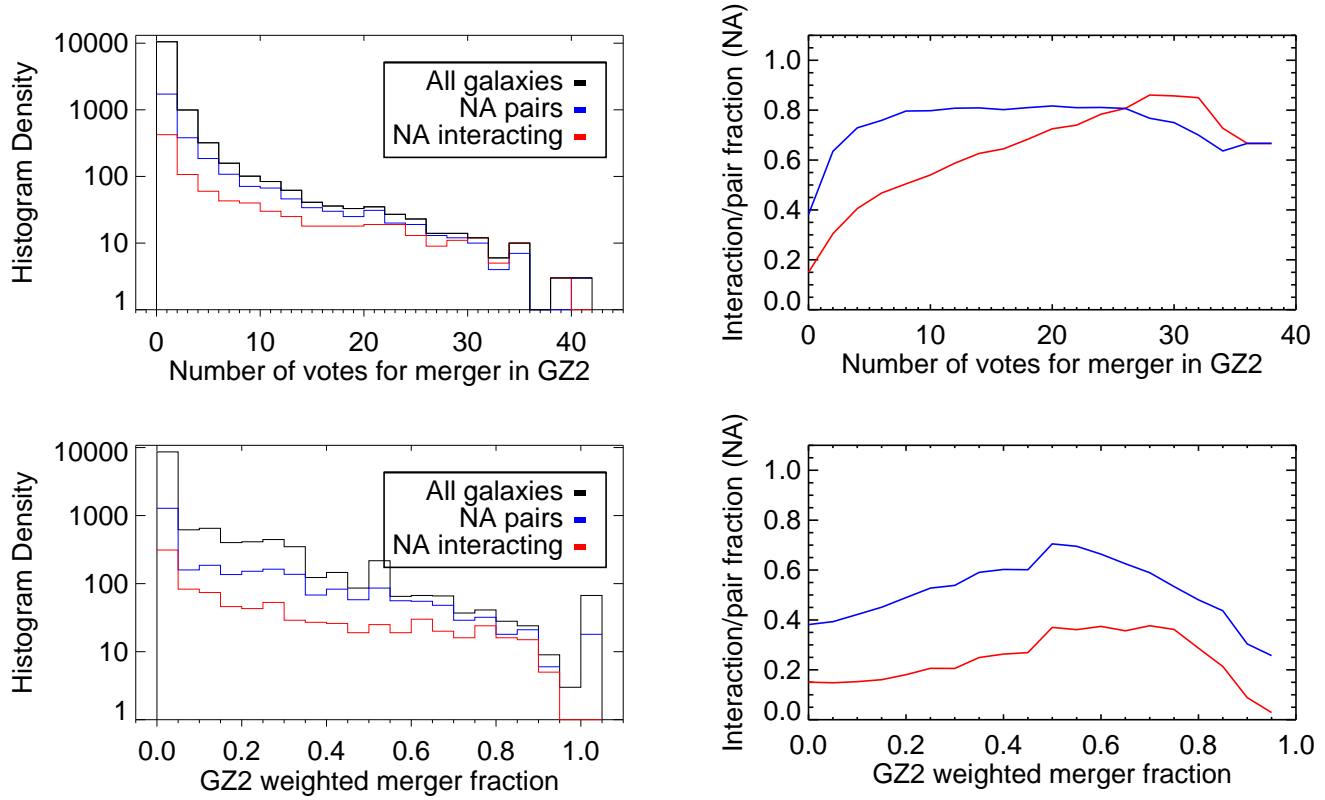


Figure 4. NA10 galaxies as a function of the votes/vote fraction for merger in GZ2. This includes all galaxies in the overlap sample (black), galaxies in pairs (blue), and interacting galaxies (red). Data are for the 12,480 galaxies found in both samples.

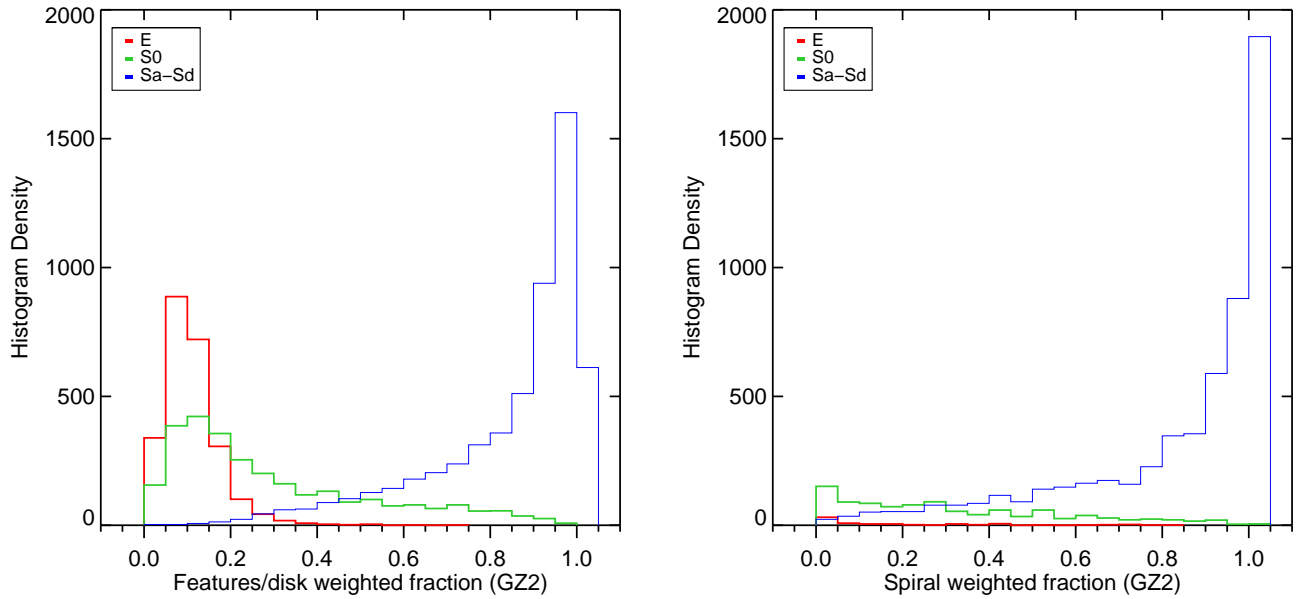


Figure 5. NA10 T-Type classifications compared to GZ2. Data on the left are for the 12,480 galaxies found in both samples; the right only shows the 5,683 galaxies with at least 10 responses to Task 04 (visible spiral structure) in GZ2.

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 above 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.

5.1.5 Spiral tightness

If a disk galaxy was identified as having spiral structure, Task 10 in GZ2 asked users to classify the “tightness” of the spiral arms. This task 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 6 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 disproportionately 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 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 clas-

sifications. 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.

5.1.6 Bulge dominance

Disk galaxies in GZ2 are also classified by the a visible dominance of a bulge (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 7 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 Sd spirals for non-zero weighted fractions. For weighted fractions above 0.1, 81% of galaxies are Sc or later; this rises to 100% for weighted fractions higher than 0.6. “Just noticeable” galaxies show a smooth change in T-Type distribution; low weighted fractions are dominated by S0 and Sa galaxies, while high weighted fractions are Sb-Sd. “Obvious” bulge galaxies are almost a mirror image of the “just noticeable” votes; low weighted fractions are Sb-Sd galaxies, and high weighted fractions are S0-Sa galaxies. Among galaxies classified as “dominant”, less than 10 galaxies have weighted fractions above 0.6 (which are a surprisingly diverse mix of S0, Sa, and Sd). Most remaining galaxies have dominant weighted 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 for which the bulge question was answered are most often “dominant”, but there is no obvious separation of ellipticals from disk galaxies based on this task.

Since Hubble types are based on both bulge dominance and the distance of their spirals, we explore whether the combination of Tasks 05 and 10 from GZ2 improve the separation between T-Types. Figure 8 shows the GZ2 bulge prominence vs. spiral tightness, color-coded by their NA10 T-Types. The GZ2 variables are a weighted average of the votes for all possible task responses, where:

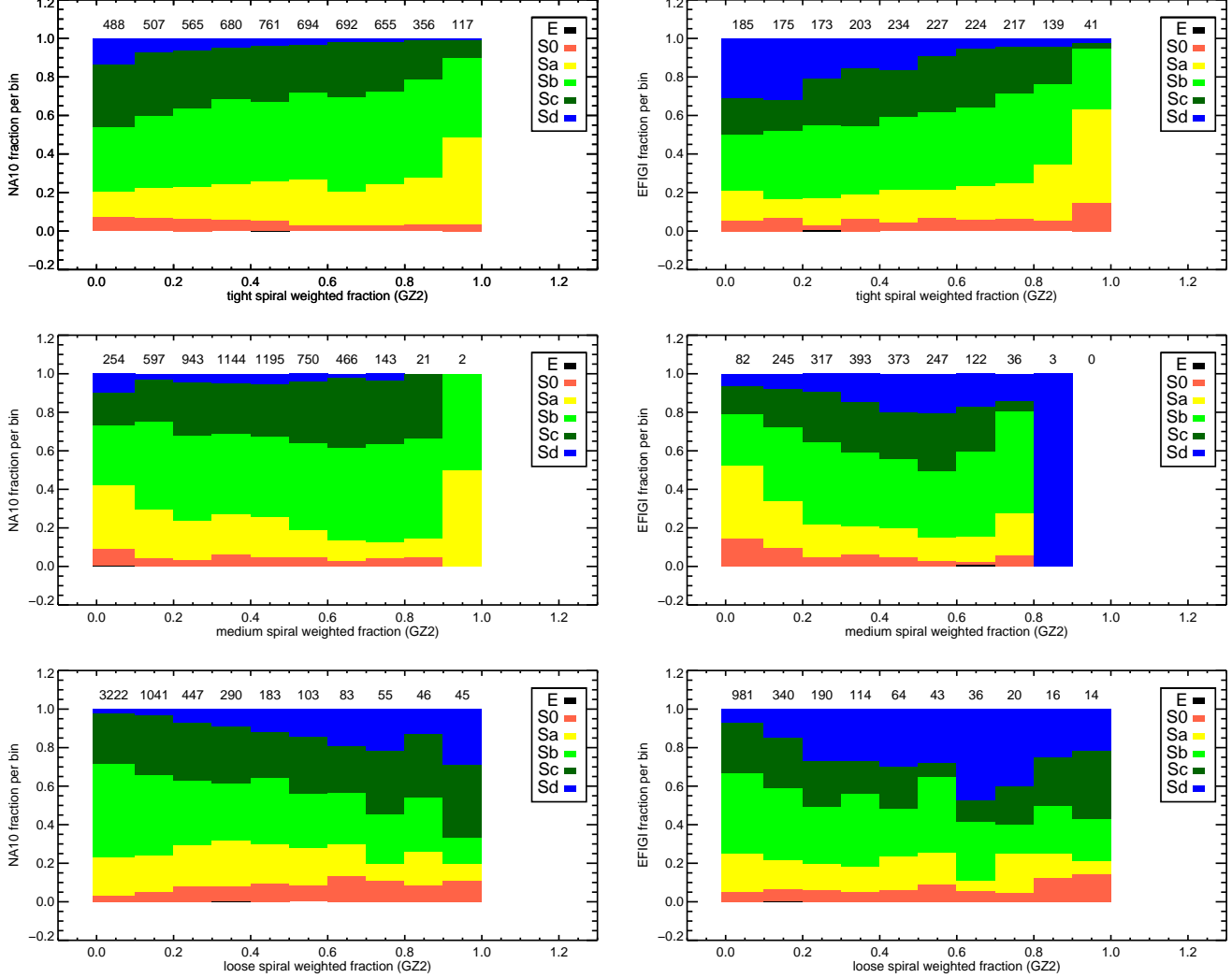


Figure 6. T-Type classifications compared to the GZ2 vote fractions for spiral tightness (Task 10). Left side is NA10 T-Types; right side is EFIGI T-Types. Data are for the 5,515 (NA10) and 1,907 (EFIGI) galaxies, respectively, with at least 10 GZ2 votes for Task 10. The number of galaxies per bin is indicated along the top of each panel.

$$p_{\text{bulge}} = (0f_{\text{nobulge}} + 1f_{\text{justnoticeable}} + 2f_{\text{obvious}} + 3f_{\text{dominant}})/3, \quad (1)$$

and

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

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.

5.1.7 T-Types to clicks

- How like a given T-Type is a galaxy?
- Take all GZ2 vote fractions, plus galaxy metadata (z, R_{50}, m_R, μ)
- Run a PCA on NA10 classifications to see which GZ2 parameters are most important

- Apply it and get the likelihood of a particular galaxy from the 20 axes

- This is the training set that is applied to the rest of Zoo2, accounting for evolution in the metadata properties on which the PCA can act.

5.2 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 astronomers, each of whom classified a subset of 445 galaxies. A training set of 100 galaxies was classified by all astronomers in the group to adjust for individual bias.

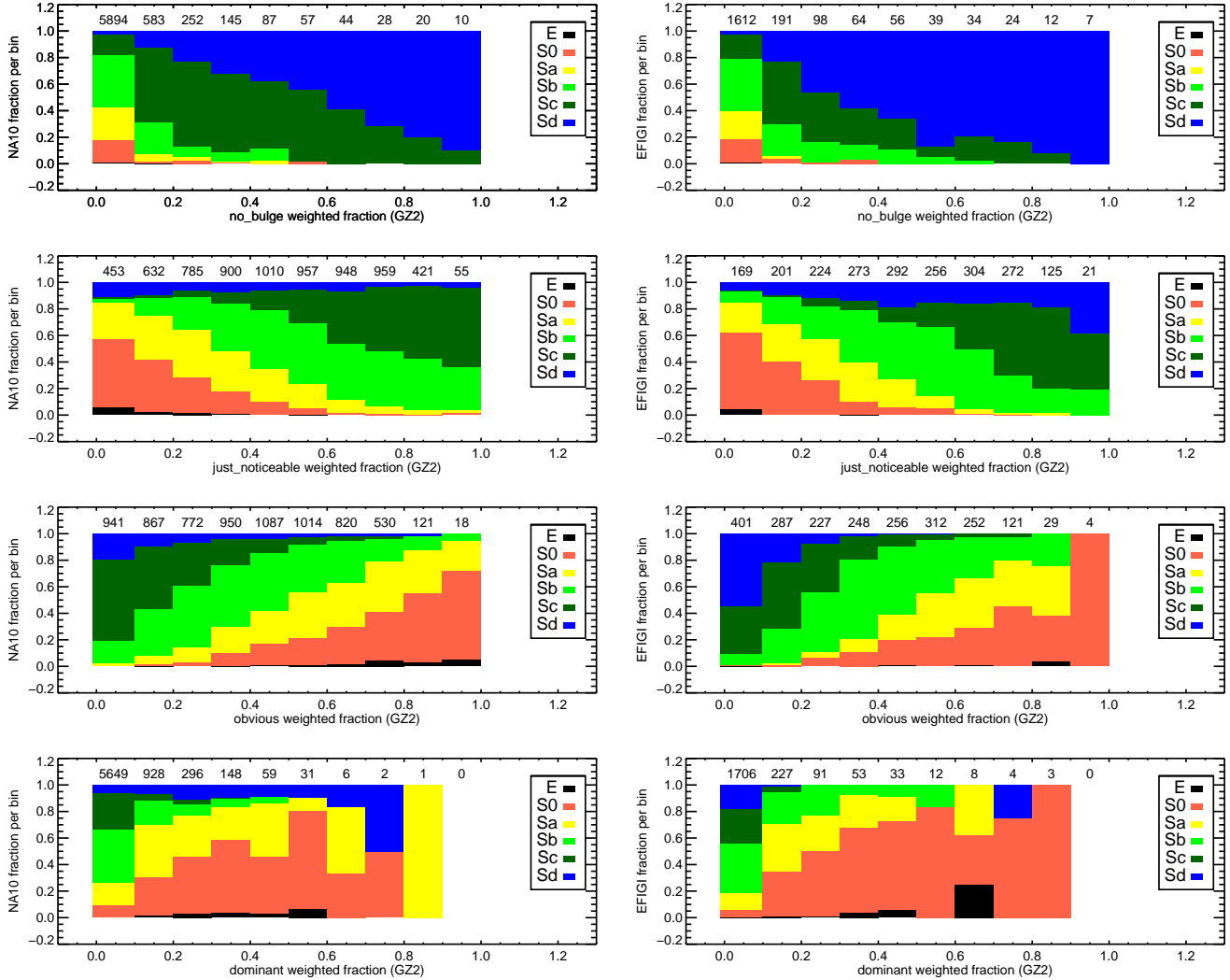


Figure 7. T-Type classifications compared to the GZ2 vote fractions for bulge prominence (Task 05). Left side is NA10 T-Types; right side is EFIGI T-Types. Data are for the 7,120 (NA10) and 2,321 (EFIGI) galaxies, respectively, with at least 10 GZ2 votes for Task 05. The number of galaxies per bin is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

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

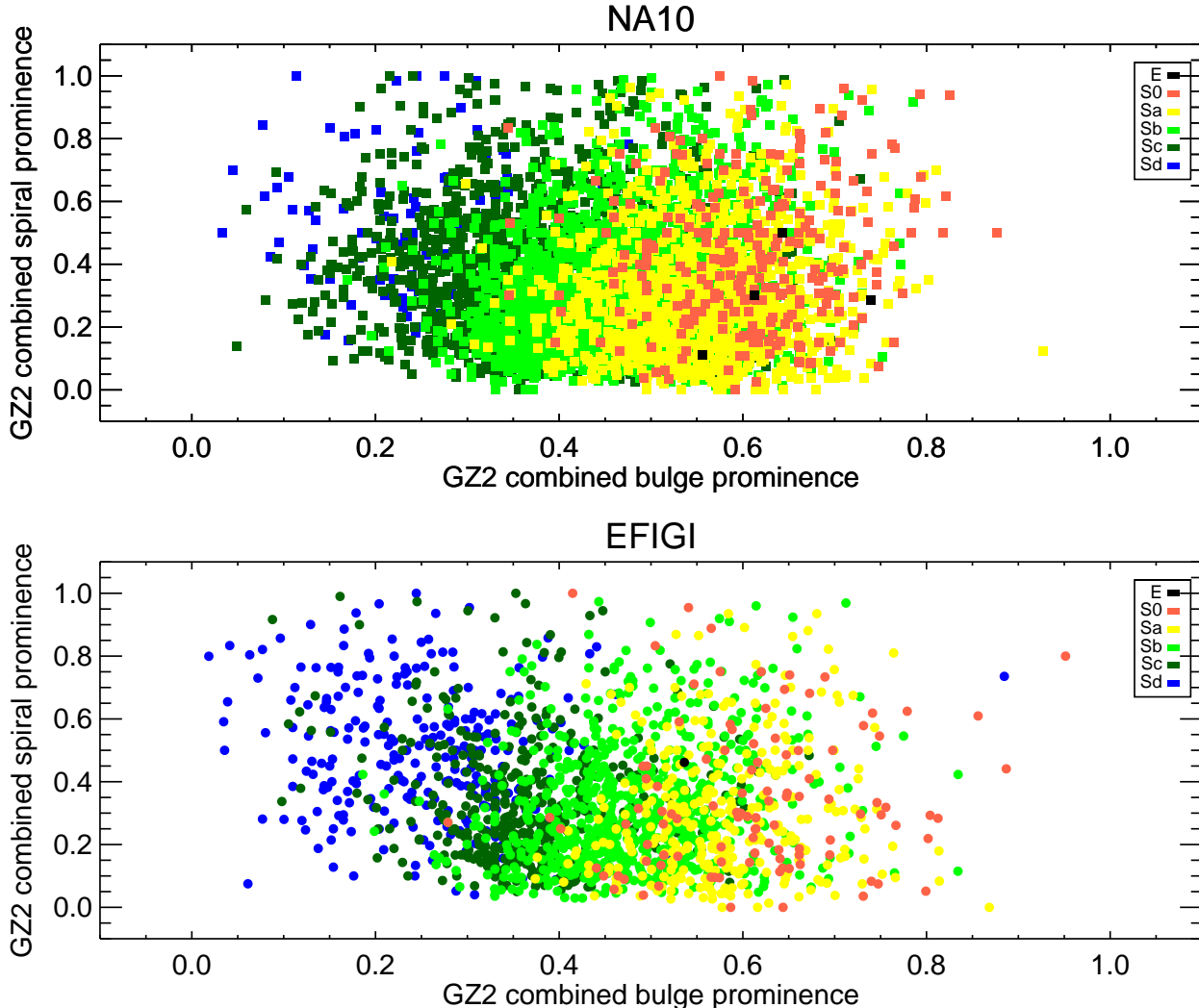


Figure 8. Weighted bulge classifications vs. weighted spiral classifications from GZ2. Galaxies are color-coded by their morphologies in EFIGI (*top*) and NA10 (*bottom*). Data are for galaxies with at least 10 classifications for both Task 05 (bulge dominance) and Task 10 (spiral structure) in GZ2.

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.

5.2.1 Bars

GZ2 asks users to identify whether a bar is present in the galaxy. EFIGI’s scale is based on bar length (not necessarily corresponding to strength) with respect to D_{25} , the decimal logarithm of the mean isophote diameter at a surface brightness of $\mu_B = 25 \text{ mag arcsec}^{-2}$. 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} .

There is a strong correlation between the GZ2 weighted fractions for bars and the bar attribute strength from EFIGI

(Figure 9). 65% of GZ2 galaxies in the overlap sample have no strong evidence for a bar (less than 0.3); of those, 77% had EFIGI bar attributes of 0.0 and 94% had 0.25 or less. For higher values of the GZ2 bar weighted fraction, the EFIGI attribute is slightly lower; the largest number of galaxies with GZ2 weighted fraction above 0.8 have EFIGI values of 0.75. The correlation coefficient between the variables is 0.51; if only face-on galaxies are considered, this increases to 0.75.

If the Masters et al. (2011) GZ2 weighted fraction of ≥ 0.5 , at least 10 bar votes, and face-on galaxies is applied, then 98% (646/660) galaxies overlapping with the EFIGI catalog have a bar attribute above 0. The mean EFIGI attribute (weighted by the confidence intervals) for galaxies barred by the Masters et al. (2011) criteria is 0.62, indicating a selection preference toward medium-length bars, between one-third and one-half of D_{25} .

The overall fraction of barred galaxies in EFIGI is 42% (1439/3354); this is essentially unchanged if only face-on galaxies are considered (915/2099 = 44%). This is signifi-

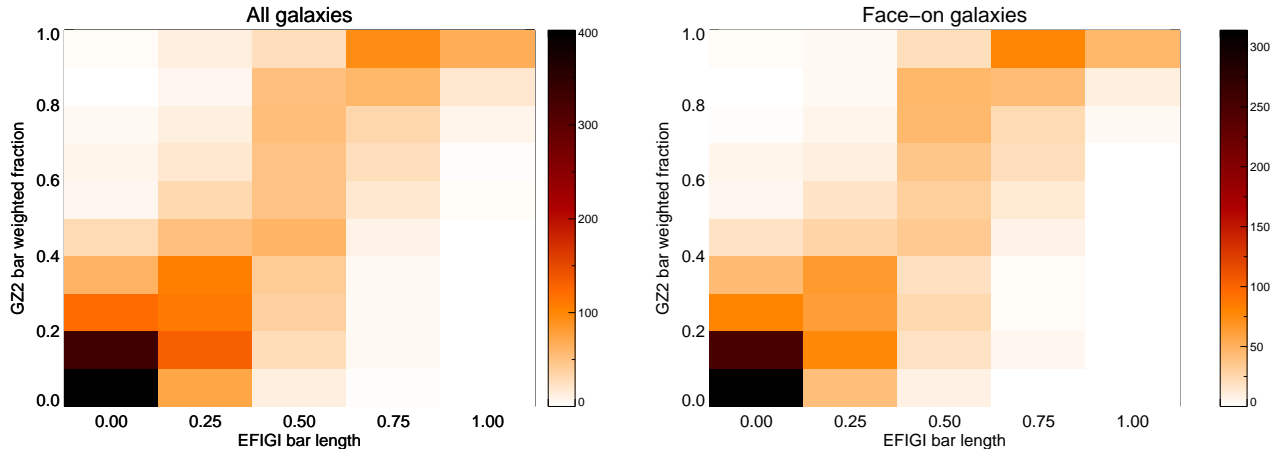


Figure 9. EFIGI bar length classifications compared to their GZ2 weighted fractions for the presence of a bar. Data on the left are for the 3,354 galaxies in both samples; the subset of 2,099 face-on galaxies is on the right. The dashed line is not a fit to the data, but is the one-to-one correlation.

cantly higher than the mean bar fraction of Masters et al. (2011), at 29.5%, but consistent with results using automated ellipse-fitting techniques (Barazza et al. 2008; Aguerri et al. 2009). The higher fraction in EFIGI is due to the contributions of galaxies with bar length attributes of 0.25, the majority of which have GZ2 weighted 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; since Baillard et al. (2011) defines this as a “barely visible” bar, it may be expected that less than half of GZ2 users would have detected it.

Masters et al. (2011) suggest in their paper that trends of bar fraction should ideally be discussed in the context of large or strong bars; an EFIGI cut of 0.5 or above would likely suit this philosophy best.

5.2.2 Arm curvature

EFIGI 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 weighted fractions/votes 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. The tight spiral weighted fraction follows the trend of the EFIGI arm curvature (Figure 10). The Spearman’s correlation coefficient for tight spirals is $\rho = 0.62$. The medium spiral weighted fraction is clustered in the middle of the EFIGI values, where galaxies with the highest GZ2 weighted 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 weighted fractions above 0.5, but those which do have low EFIGI arm curvature values (0.0–0.25).

Trends described above are quantitatively similar for both the full matching sample and for face-on galaxies only

(right side of Figure 10). Somewhat surprisingly, the distribution also appears similar if only considering galaxies with a minimum number of GZ2 votes on spiral winding arms. Lower limits of 5, 10, 20, and 30 votes produce the same 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.

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).

5.3 Huertas-Company - automated classifications

Huertas-Company et al. (2011) published a study in which they used training sets of galaxies to create artificial 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) are not suited for the fine structure questions such as bar and spiral number.

The classified sample of 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 the size of the GZ2 sample, which has a similar redshift range. The primary difference between the samples is that the HC sample goes to fainter depths,

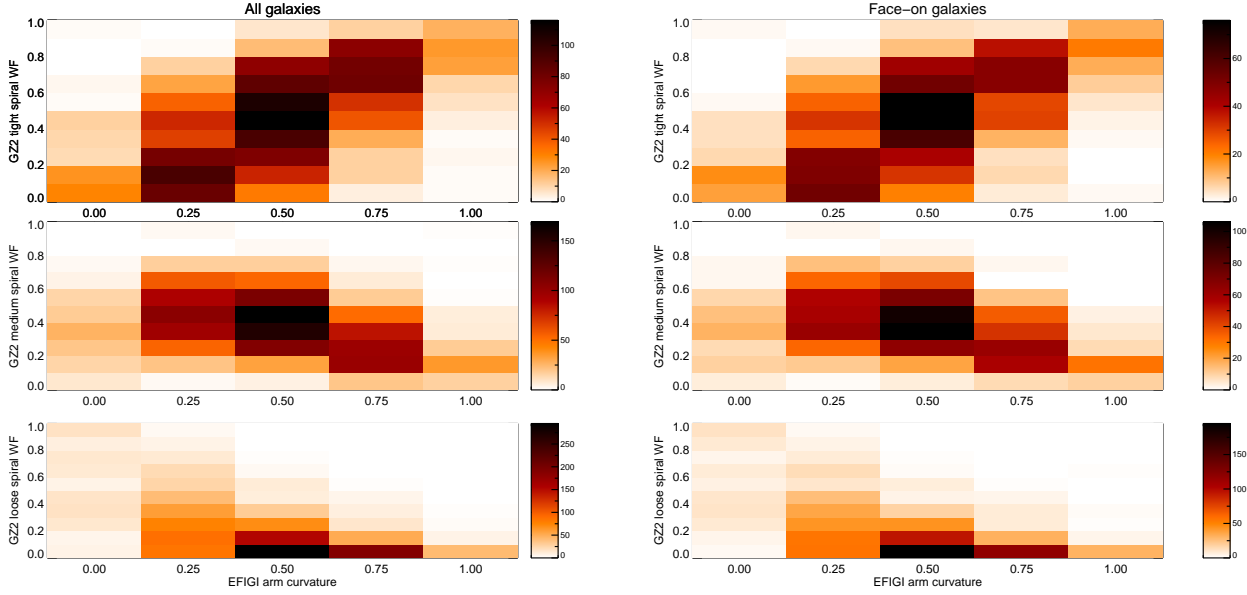


Figure 10. EFIGI arm curvature classifications compared to their GZ2 weighted fractions for the presence of a bar. Data on the left are for the 3,411 galaxies in both samples; the subset of 2,099 face-on galaxies is on the right. Dashed lines on the top and bottom pairs of plots show the one-to-one correlation, a Gaussian with $\mu = 0.5$ and $\sigma = 0.25$, and the one-to-one anti-correlation, respectively.

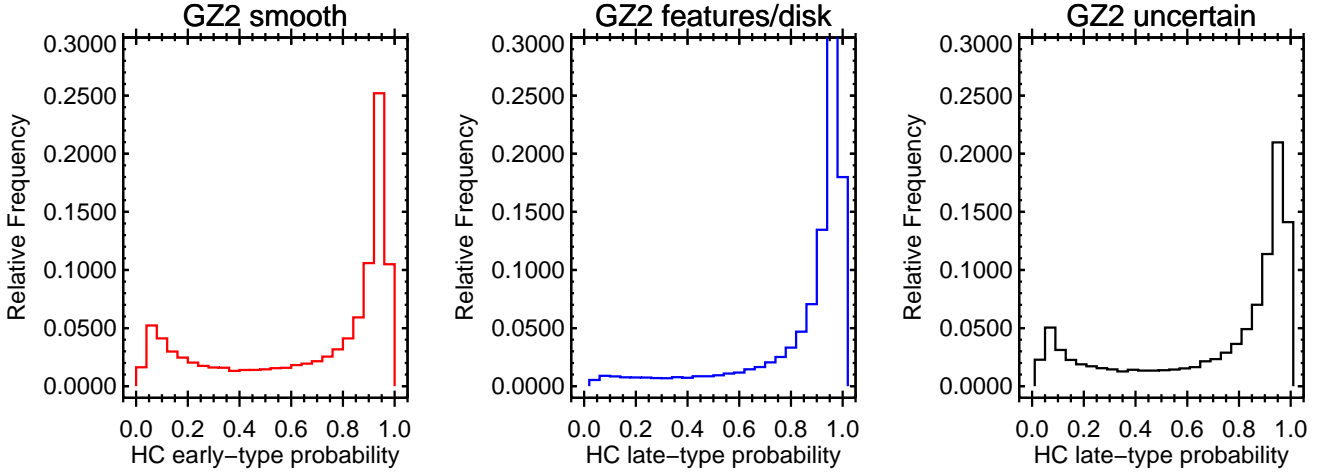


Figure 11. Left: distribution of HC early-type probabilities for galaxies with Task 01 “smooth” weighted fraction above 0.8. Middle: distribution of HC late-type probabilities for galaxies with Task 01 “features or disk” weighted fraction above 0.8. Right: HC late-type probabilities for uncertain galaxies ($p < 0.8$ for all responses to Task 01).

with more than 400,000 galaxies below the GZ2 limit of $r > 17$. The algorithm for classification is implemented using support vector machine software that tries to find boundaries between points in N -dimensional space. The training set for galaxy morphology is the 2,253 galaxies in Fukugita et al. (2007), classified by T-Type. Each galaxy has a probability of being in one of four subclasses: E, S0, Sab, and Scd.

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 lentic-

ulars, 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 of being S0 or Scd. The HC algorithm is thus very good at identifying galaxies at the extreme ends of the Hubble tuning fork, but have larger amounts of overlapping probabilities for intermediate states.

Since the GZ2 galaxies are a subset of the GZ1 sample, the results are expected to be similar to those described in Huertas-Company et al. (2011). Figure 11 shows the distributions of the HC early- and late-type probabilities for GZ2 galaxies robustly identified ($p > 0.8$) as either smooth or having features/disks. The median HC 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 HC (the bump on the left side of the first panel in Figure 11). The average 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-HC smooth galaxies have fewer round galaxies and many more “cigar-shaped” galaxies in this sample. A high axial ratio might have trained the HC 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 categorize 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 colors of $(g - r) = 0.67$ and $(g - r) = 0.97$. Since the SVM method does include SDSS colors as a parameter, we conjecture that the low HC early-type probability is in part due to the fact that they are blue, in addition to morphological features such as shape and concentration. It would be interesting to see what fraction of the blue ellipticals (Schawinski et al. 2009) fall in this peak.

The right panel of Figure 11 shows the distribution of “unclassified” galaxies, for which none of the responses for Task 01 had a weighted fraction above 0.8. The HC probability for these galaxies is bimodal, with the larger fraction classified as HC late-type and a smaller fraction as HC early-type.

Figure 12 plots the HC probability values against the GZ2 vote fraction for both smooth and feature/disk galaxies, similar to Figure 8 in Huertas-Company et al. (2011). Significant excesses are seen at the lower left and upper right corners for both samples, confirming the earlier result that robust classifications using both methods tend to agree. The relationship of the HC probability to GZ2 vote fraction, however, seems distinctly non-linear and significantly different from the relationship shown in Huertas-Company et al. (2011). Galaxies with low HC early-type probabilities show no strong correlation with the GZ2 weighted fraction, resulting in the vertical stripe in Figure 12. Intermediate values of the HC probabilities are typically classified by GZ2 users as “smooth”, with a concentration at the highest HC early-type probabilities. It should be noted that without GZ2 debias-

ing, the two panels in Figure 12 are essentially mirror images of each other (since $p_{HC,early} + p_{HC,late} \equiv 1$ and the GZ2 weighted fractions only have marginal contributions from either the star/artifact response or downweighting of inconsistent users).

Splitting the HC morphology types into the E, S0, Sab, and Scd subclasses, Figure 13 shows the correlations to the GZ2 weighted fractions. While the direction of the correlation is the same as seen in Figures 9 and 10 in Huertas-Company et al. (2011), the behavior at the extrema is quite different. The debiased GZ1 data shows a strong cluster of low-HC and low-GZ probabilities for all four subclasses; no such cluster is seen in any of the plots in Figure 13. Furthermore, high HC probabilities have the lowest fraction of galaxies in the GZ1 published data, where our plots show a concentration of galaxies along the high end for both P(E) and P(S0). It should be investigated whether the scaling on these sets of figures is truly plotting the same value.

- Check how the presence of a bar affects the HC classifications
- Since T-Types seem to be most strongly correlated with bulge prominence, plot these against each other.

6 GALAXY WARS

7 ASTRONOMY

8 CONCLUSIONS

REFERENCES

- Aguerre 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., Ivezić 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
- Barazza F. D., Jogee S., Marinova I., 2008, *ApJ*, 675, 1194
- Buta R., Combes F., 1996, *FCP*, 17, 95
- Conselice C. J., 2003, *ApJS*, 147, 1
- Conselice C. J., 2006, *MNRAS*, 373, 1389
- Fukugita M., Nakamura O., Okamura S., Yasuda N., Barntine J. C., Brinkmann J., Gunn J. E., Harvanek M., Ichikawa T., Lupton R. H., Schneider D. P., Strauss M. A., York D. G., 2007, *AJ*, 134, 579
- Huertas-Company M., Aguerri J. A. L., Bernardi M., Mei S., Sánchez Almeida J., 2011, *A&A*, 525, A157+
- Lintott C., Schawinski K., Bamford S., Slosar A., Land K., Thomas D., Edmondson E., Masters K., Nichol R. C., Raddick M. J., Szalay A., Andreescu D., Murray P., Vandenberg J., 2011, *MNRAS*, 410, 166
- Lintott C. J., Schawinski K., Slosar A., Land K., Bamford S., Thomas D., Raddick M. J., Nichol R. C., Szalay A., Andreescu D., Murray P., Vandenberg J., 2008, *MNRAS*, 389, 1179
- Masters K. L., Nichol R. C., Hoyle B., Lintott C., Bamford S. P., Edmondson E. M., Fortson L., Keel W. C., Schaw-

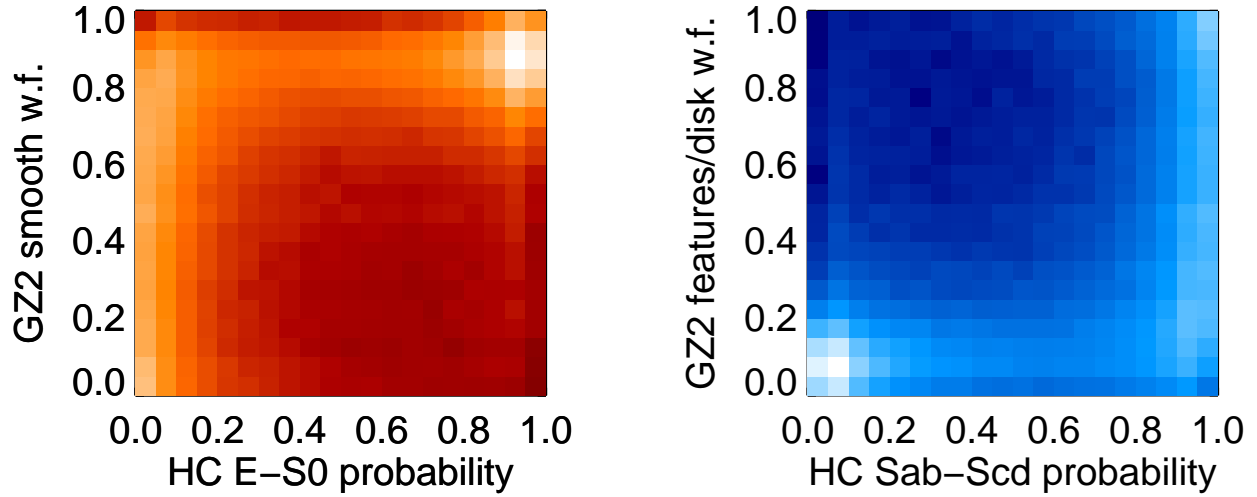


Figure 12. Left: GZ2 smooth weighted fraction as a function of Huertas-Company et al. (2011) early-type probability. Right: GZ2 features/disk weighted fraction as a function of HC late-type probability. Whiter values in both indicate a larger fraction of galaxies in that bin (logarithmic scale).

inski K., Smith A. M., Thomas D., 2011, MNRAS, 411, 2026
 Nair P. B., Abraham R. G., 2010a, ApJS, 186, 427
 Nair P. B., Abraham R. G., 2010b, ApJL, 714, L260
 Schawinski K., Lintott C., Thomas D., Sarzi M., Andreescu D., Bamford S. P., Kaviraj S., Khochfar S., Land K., Murray P., Nichol R. C., Raddick M. J., Slosar A., Szalay A., Vandenberg J., Yi S. K., 2009, MNRAS, 396, 818

This paper has been typeset from a \LaTeX file prepared by the author.

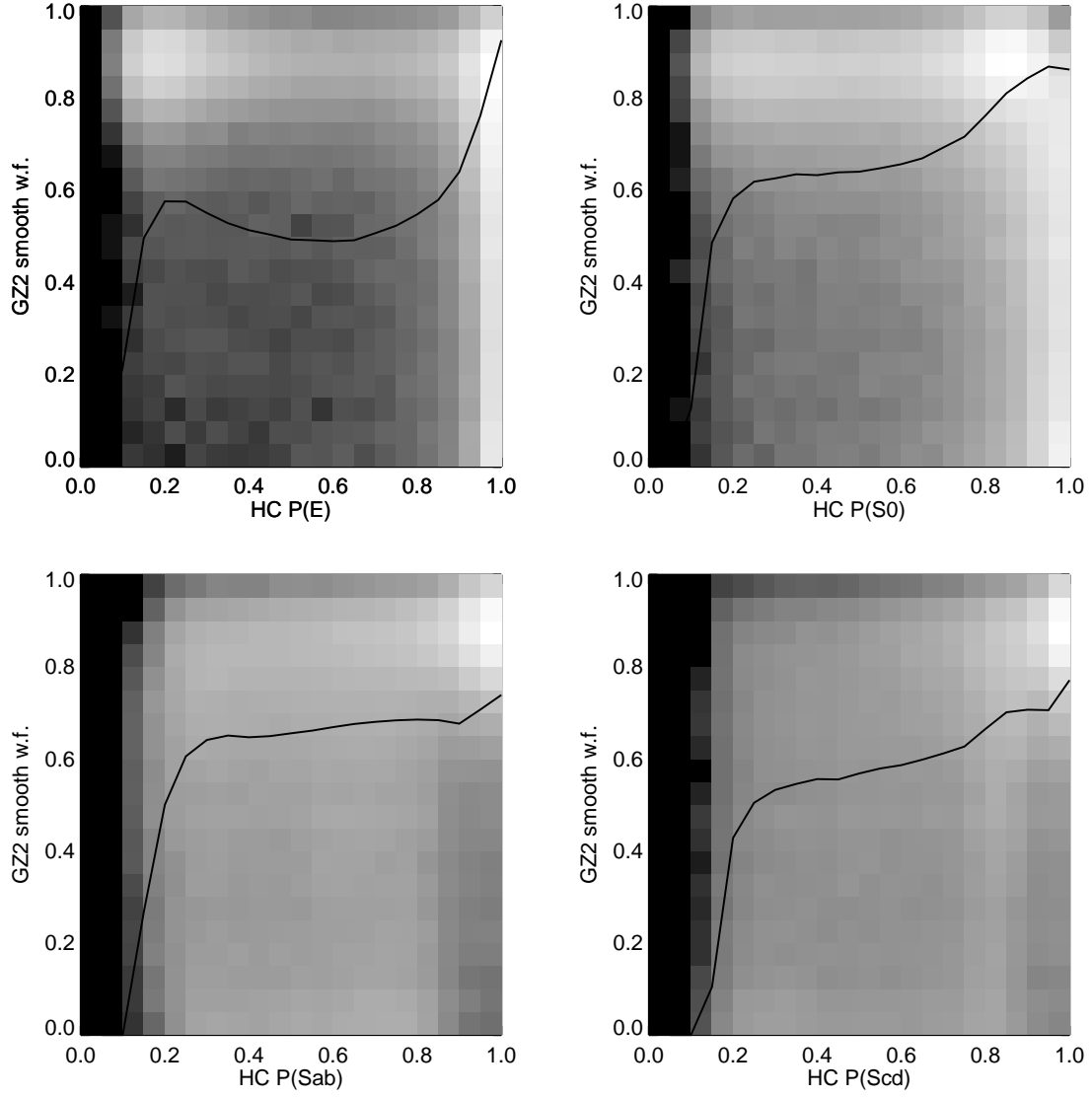


Figure 13. Left: GZ2 smooth weighted fraction as a function of Huertas-Company et al. (2011) early-type probability. Right: GZ2 features/disk weighted fraction as a function of HC late-type probability. Whiter values in both indicate a larger fraction of galaxies in that bin (logarithmic scale).

Table 1. GZ2 weighted vote fractions for the main and Stripe 82 normal-depth samples

User task and answer	Mean weighted vote fraction		Number of galaxies		Sample comparison	
	Main	Stripe 82	Main	Stripe 82	$f_m - f_s$	f_m/f_s
T01_SMOOTH_OR_FEATURES_A01_SMOOTH	0.642	0.613	273783	8437	0.029	1.048
T01_SMOOTH_OR_FEATURES_A02_FEATURES_OR_DISK	0.325	0.352	273783	8437	-0.027	0.924
T01_SMOOTH_OR_FEATURES_A03_STAR_OR_ARTIFACT	0.033	0.036	273783	8437	-0.003	0.924
T02_EDGEON_A04_YES	0.234	0.229	118147	4110	0.005	1.020
T02_EDGEON_A05_NO	0.766	0.771	118147	4110	-0.005	0.994
T03_BAR_A06_BAR	0.287	0.292	87647	3077	-0.004	0.986
T03_BAR_A07_NO_BAR	0.713	0.708	87647	3077	0.004	1.006
T04_SPIRAL_A08_SPIRAL	0.660	0.674	87621	3077	-0.014	0.980
T04_SPIRAL_A09_NO_SPIRAL	0.340	0.326	87621	3077	0.014	1.042
T05_BULGE_PROMINENCE_A10_NO_BULGE	0.129	0.114	87608	3076	0.015	1.129
T05_BULGE_PROMINENCE_A11_JUST_NOTICEABLE	0.479	0.491	87608	3076	-0.012	0.976
T05_BULGE_PROMINENCE_A12_OBVIOUS	0.321	0.333	87608	3076	-0.011	0.966
T05_BULGE_PROMINENCE_A13_DOMINANT	0.070	0.062	87608	3076	0.008	1.134
T06_ODD_A14_YES	0.199	0.185	273636	8391	0.014	1.073
T06_ODD_A15_NO	0.801	0.815	273636	8391	-0.014	0.983
T07_ROUNDED_A16_COMpletely_ROUND	0.330	0.326	229974	6956	0.005	1.014
T07_ROUNDED_A17_IN_BETWEEN	0.496	0.501	229974	6956	-0.004	0.991
T07_ROUNDED_A18_CIGAR_SHAPED	0.173	0.174	229974	6956	-0.000	0.999
T08_ODD_FEATURE_A19_RING	0.147	0.178	72439	2114	-0.031	0.827
T08_ODD_FEATURE_A20_LENS_OR_ARC	0.045	0.040	72439	2114	0.005	1.122
T08_ODD_FEATURE_A21_DISTURBED	0.121	0.125	72439	2114	-0.003	0.973
T08_ODD_FEATURE_A22_IRREGULAR	0.177	0.170	72439	2114	0.007	1.039
T08_ODD_FEATURE_A23_OTHER	0.282	0.245	72439	2114	0.037	1.152
T08_ODD_FEATURE_A24_MERGER	0.212	0.209	72439	2114	0.003	1.015
T08_ODD_FEATURE_A38_DUST_LANE	0.016	0.034	72439	2114	-0.018	0.475
T09_BULGE_SHAPE_A25_ROUNDED	0.538	0.567	21496	746	-0.029	0.948
T09_BULGE_SHAPE_A26_BOXY	0.103	0.096	21496	746	0.007	1.077
T09_BULGE_SHAPE_A27_NO_BULGE	0.359	0.338	21496	746	0.022	1.065
T10_ARMS_WINDING_A28_TIGHT	0.370	0.392	55800	2052	-0.022	0.943
T10_ARMS_WINDING_A29_MEDIUM	0.417	0.405	55800	2052	0.011	1.028
T10_ARMS_WINDING_A30_LOOSE	0.214	0.203	55800	2052	0.011	1.054
T11_ARMS_NUMBER_A31_1	0.074	0.067	55805	2053	0.008	1.117
T11_ARMS_NUMBER_A32_2	0.501	0.478	55805	2053	0.023	1.048
T11_ARMS_NUMBER_A33_3	0.093	0.098	55805	2053	-0.005	0.950
T11_ARMS_NUMBER_A34_4	0.038	0.043	55805	2053	-0.005	0.884
T11_ARMS_NUMBER_A36_MORE_THAN_4	0.030	0.033	55805	2053	-0.002	0.934
T11_ARMS_NUMBER_A37_CANT_TELL	0.263	0.281	55805	2053	-0.019	0.934

Note. — Data for each task is only for galaxies with at least 10 total responses to the classification question.