Galaxy Zoo: Morphological Classifications for Galaxies in HST Legacy Imaging

Lead Author and other Galaxy Zoo science team members

* This publication has been made possible by the participation of more than 200,000 volunteers in the Galaxy Zoo project. Their contributions are individually acknowledged at http://authors.galaxyzoo.org/authors.html.

E-mail: lead.author@university.edu

25 February 2016

ABSTRACT

This will be the data release paper for GZ:Hubble. We present the classifications, the methodology for data reduction and corrections for redshift dependent biases in the observed morphologies.

1 INTRODUCTION

Usual due diligence for an intro science paper. Should cover:

(i) Motivation for studying morphology of galaxies

(ii) Particular scientific questions governed by galaxies at $z \sim 1$

(iii) Theoretical predictions for galaxy morphology evolution

(iv) Past imaging and morphology studies at $z \sim 1$

(v) Summary of GZ/citizen science work on galaxy morphology to date

2 SAMPLE AND DATA

2.1 Summary of HST Legacy Survey Imaging

• Hubble ACS imaging for the All-Wavelength Extended Groth Strip International Survey (AEGIS; Davis et al. 2007) covers a strip centered at $\alpha = 14^{\mathrm{h}}17^{\mathrm{m}}, \delta = +52°30'$. The strip was originally selected due to low extinction and Galactic/zodiacal emission, making it a prime target for multiwavelength observations by space-based observatories. The ACS images covered 63 separate tiles over a total area of ~ 710 arcmin^2. Images were in two bands, with exposure times of 2300 seconds in F606W (V_{606W}) and 2100 seconds in F814W (I_{814W}). The final mosaic images are dithered to a resolution of 0.03 $''$/pixel. For extended objects, the limiting magnitudes of sources in AEGIS are 26.23 (AB) in V_{606W} and 25.61 (AB) in I_{814W}.

• The Great Observatories Origins Deep Survey (GOODS; Giavalisco et al. 2004) covers two well-studied fields in the northern and southern hemispheres: the Hubble Deep Field-North ($\alpha = 12^{\mathrm{h}}36^{\mathrm{m}}, \delta = +62°14'$) and the Chandra Deep Field-South ($\alpha = 03^{\mathrm{h}}32^{\mathrm{m}}, \delta = -27°48'$). Data including Hubble ACS images are referred to as GOODS-N and GOODS-S, respectively. ACS imaged the GOODS fields in 4 filters – F435W (B_{435W}), V_{606W}, F775W

(I_{775W}), and F850LP (I_{850LP}). The mean exposure times for each epoch vary by band, from 1050–2100 seconds. The B_{435W} images were completed in a single epoch at the beginning of the survey, but the V_{606W}, I_{775W}, and I_{850LP} images were taken in five separate epochs separated by 40–50 days each. The ACS images are dithered to a pixel scale of 0.03 $''$/pixel and covers a total area of ~ 320 arcmin^2 (160 arcmin^2 per field). The 5σ limiting magnitudes for extended sources are 25.7 for V_{606W} and 25.0 for I_{775W}.

• The Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) covers an area of ~ 1.8 deg^2 centered at $\alpha = 10^{\mathrm{h}}00^{\mathrm{m}}28^{\mathrm{s}}, \delta = +02°12'21''$. Its location near the celestial equator was designed to enable coverage by ground-based telescopes in both the Northern and Southern Hemispheres, as well as the space-based observatories. The Hubble ACS data from COSMOS consists of 1 orbit with 2028 seconds per pointing in I_{814W}, consisting of 590 total pointings. The image resolution is dithered to 0.05 $''$/pixel. The 50% completeness magnitude for a galaxy with a half-light radius of $0''.50$ in I_{814W} is 24.7 mag.

• The Galaxy Evolution from Morphologies and SEDS (GEMS; Rix et al. 2004; Caldwell et al. 2008) survey is also centered on the Chandra Deep Field-South. The GEMS data covers ~ 800 arcmin^2, and surrounds the area covered by GOODS-S. Images from ACS in GEMS have 1 orbit per pointing for a total of 63 pointings. The exposure times are 2160 and 2286 seconds in V_{606W} and I_{850LP}, respectively. The image resolution has a pixel scale of 0.03 $''$/pixel. The 5σ limiting magnitudes for source detection are 25.7 AB in V_{606W} and 24.2 AB in I_{850LP}.

2.2 Image creation

The GOODS images in GZH use mosaics constructed from both 2-epoch and 5-epoch sets of data.

The filters that Griffith et al. (2012) uses for the colored

Table 1. Summary of Galaxy Zoo: Hubble imaging

Survey	t_{exp} [sec]	Filters	Resolution [$''$/pix]	Area [arcmin^2]	$N_{galaxies}$
AEGIS	$2100-2300$	V_{606W} and I_{814W}	0.03	710	8157
COSMOS	2028	I_{814W}	0.05	6480	88530
GEMS	$2160-2286$	V_{606W} and I_{850LP}	0.03	800	9143
GOODS	$1000-2100$	$B_{435W}, V_{606W}, I_{775W}, I_{850LP}$	0.03	320	7336
GOODS-N	$-$	$-$	$-$	$-$	*2551*
GOODS-S	$-$	$-$	$-$	$-$	*4785*
total	$-$	$-$	$-$	8310	113166

images were F606W and F775W for GOODS-N and F606W and F850LP for GOODS-S.

We use different filters for the north and south GOODS fields so that GEMS can be directly compared with GOODS-S (Figure 1).

Fake AGN

Stripe 82

Different treatment of colored noise in COSMOS; creating color gradients with Subaru data and using I_{814W} for illumination map.

FERENGI images.

2.3 Redshifts

We obtain redshifts from a variety of sources: spectroscopic and photometric redshifts from Griffith et al. (2012), spectroscopic, photometric, and grism redshifts from 3DHST (Momcheva et al. 2015), and photometric redshifts from UltraVISTA (Ilbert et al. 2013). The distribution is shown in Table 2, and are located in the 'Z BEST' column of the catalog. We first use the (Griffith et al. 2012) and (Momcheva et al. 2015) redshifts if those are available. If they are not, then we compare the 1σ errors of the photometric and grism redshifts from all three catalogs, and use the redshift with the smallest error. The source of the redshifts are displayed in the Z BEST SOURCE column of the catalog.

2.4 User weighting

The votes of individual users who classified galaxies in GZH are combined to make a vote fraction for each question on the classification tree. Users' votes are weighted slightly (in a method identical to that described in Willett et al. 2013) such that users who frequently disagree with all other users end up having very low weights. The majority of users have weights very close to $w = 1.0$ (**STEVEN: Is this true for GZH - do you have a plot of the distribution of user weights or consistencies we can include here?**).

3 GALAXY ZOO INTERFACE AND CLASSIFICATIONS

## 4 CORRECTING FOR REDSHIFT-DEPENDENT CLASSIFICATION BIAS

The previous versions of Galaxy Zoo morphology classifications (Lintott et al. 2008; Willett et al. 2013) were based on observations of galaxies in the Sloan Digital Sky Survey (SDSS) which are typically at $z < 0.1$. In these cases it was assumed that there was no cosmological evolution of the morphologies of galaxies and therefore any observed changes in the distribution of galaxies with different consensus morphologies was due to the effects of redshift on the image quality (*i.e..* the reduction in physical resolution, surface brightness dimming, etc). For both previous releases of GZ morphologies, we provided a correction for redshift-dependent bias based on matching the classification fractions at the highest redshfts with those at the lowest redshift. See Bamford et al. (2009) and Willett et al. (2013) for the details.

In the GZH samples, the redshift range is large enough that we expect to measure cosmological evolution of the types and morphologies of galaxies in the sample. As a result, the previous methods of correcting for redshift dependent bias will not work. In addition, the effects of band shifting will change the images even more across these redshift ranges.

In order to test and correct for the effects of redshift, we generated a set of calibration images. These images consist of the same galaxy as it would appear over a variety of redshifts. The input images are from the SDSS (York et al. 2000; Strauss et al. 2002) and are processed using the FERENGI code (Barden et al. 2008) to match the observational properties of the HST surveys out to $z = 1$. These images were classified in the Galaxy Zoo interface using the same classification scheme as the original HST images.

4.1 Selection of FERENGI input galaxies

We selected 288 unique galaxies from SDSS imaging to run through the FERENGI code. The selection spanned a variety of galaxy morphologies (as selected by GZ2 classifications) and r'-band surface brightnesses, and also spanned the redshift range of SDSS targets (in $N_z = 4$ bins) in order to be optimised for different target minimum redshifts in HST imaging.
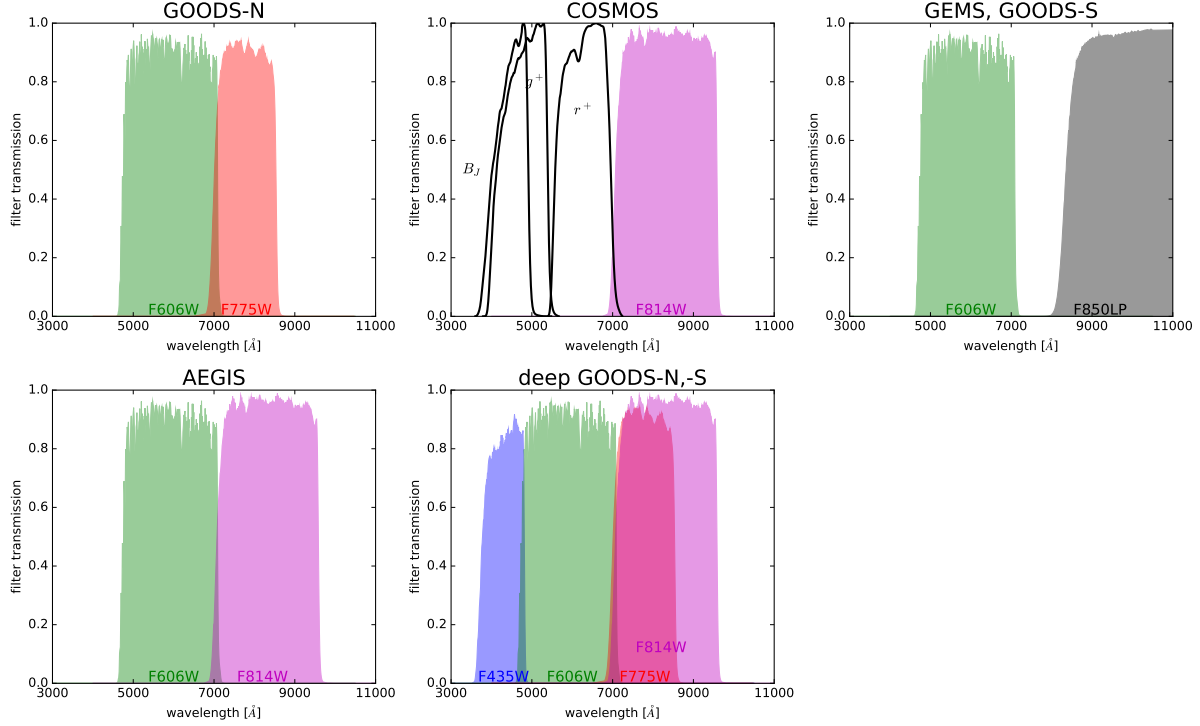
Figure 1. Transmission curves of the filters used by *HST* Advanced Camera for Surveys (ACS) in wide-field channel mode for the various surveys in Galaxy Zoo: Hubble. The unfilled black curves show the filters for the Suprime Camera on the *Subaru* telescope which were used to create color gradients in the composite images for COSMOS.

Table 2. Redshift sources of the Hubble data, by sample.

	AEGIS	COSMOS	GEMS	GOODS-N	GOODS-S	GOODS FULL	SDSS	Total
Griffith specz	3,656	7,201	387	1,997	3,016			16,257
3DHST specz	12	35	6	5	103			161
Griffith photoz	4,442	82,662	8,886	218	1,662			97,870
3DHST grismz	341	358	9	101	67			876
UltraVista photoz	0	2,473	0	0	0			2,473
3DHST photoz	56	79	16	230	65			446
Total	8,507	92,808	9,304	2,551	4,913	11,157	51,861	181,101

The selection criteria for the different morphological categories is summarised in Table 3. The surface brightness selection ($N_\mu = 3$) was (1) low: $\mu > 21.5$ mag arcsec^{-2}; (2) mid: $20.5 < \mu < 21.5$ mag arcsec^{-2}; and (3) high: $\mu < 20.5$ mag arcsec^{-2}. For each of the four "target redshifts" ($z = 0.3, 0.5, 0.8$ and 1.0), the images were redshifted in $\Delta z = 0.1$ bins up to $z = 1.0$.

In addition to the physical parameters of the input images, the FERENGI output depends on assumptions of the global galaxy evolution model. This evolution is a crude mechanism that mimics the brightness increase of galaxies with increasing redshift (out to at least $z \sim 1-2$). The effect on the redshifted images is simply an empirical addition to the magnitude of a galaxy of the form $M' = e \times z + M$, where M' is the corrected magnitude, and e is the evolutionary correction in magnitudes (i.e., $e = -1$ essentially brightens the galaxy by 1 magnitude by $z = 1$). We ran FERENGI for values of e starting from $e = 0$ and decreasing

Table 4. Summary of FERENGI artificial redshifting

z_{target}	$N_{z\text{bins}}$	$N_{\text{evolution}}$	e_{max}	N_{galaxies}	N_{images}
0.3	8	7	−3.0	72	4032
0.5	6	4	−1.5	72	1728
0.8	3	3	−1.0	72	648
1.0	1	3	−1.0	72	216

to $e = -3.5$ in increments of $\Delta e = 0.5$. Figure 3 shows several examples of the effects of "losing" spiral/disc features with increasing redshift for two galaxies with $e = 0$.

The final number of FERENGI images produced for each galaxy is ultimately a function of galaxy's redshift, since the new images cannot be resampled at better angular resolution than the original SDSS data, as well as the number of e values selected. Table 4 summarizes the total sample of redshifted images produced for GZH.
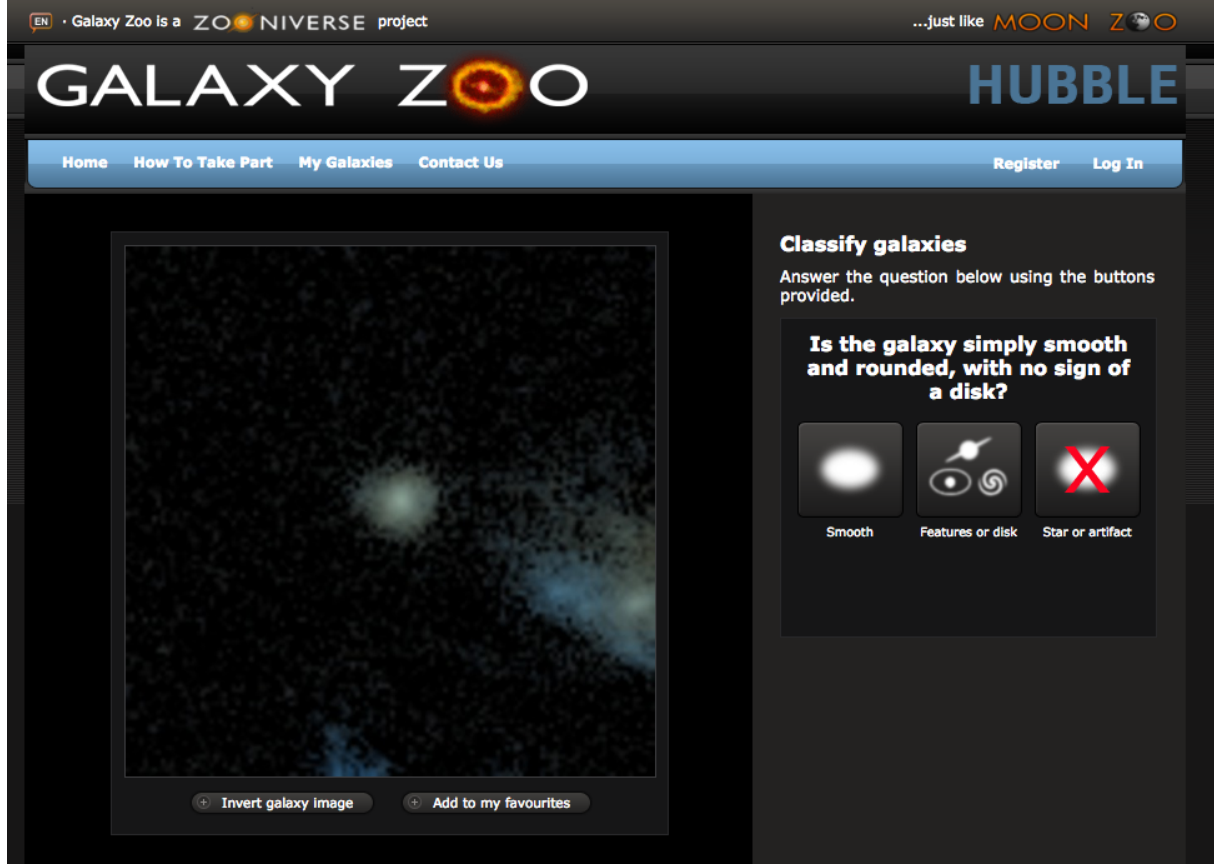
Figure 2. Screenshot of the Galaxy Zoo: Hubble interface showing an example COSMOS image at the first step in the decision tree.

Table 3. Summary of morphological categories selected for FERENGI sample.

Morphology	Label	Selection	$N_{\rm objects}$ $[N_z \times N_\mu]$
Features	Yes	$p_{\rm features} > 0.8$, $p_{\rm odd} < 0.1$	12
	Int.	$0.3 < p_{\rm smooth} < 0.6$, $p_{\rm odd} < 0.1$	12
	No	$p_{\rm smooth} > 0.8$, $p_{\rm odd} < 0.1$	12
Merger	No	$p_{\rm features} > 0.8$, $p_{\rm odd < 0.1}$, $p_{\rm merger} < 0.1$	12
	Int.	$p_{\rm odd} > 0.5$, $0.1 < p_{\rm merger} < 0.4$	12
	Yes	$p_{\rm odd} > 0.5$, $p_{\rm merger} > 0.4$	12
Edge-on	Yes	$p_{\rm edgeon} > 0.8$, $p_{\rm features} > 0.5$	12
	Int.	$0.4 < p_{\rm edgeon} < 0.8$, $p_{\rm features} > 0.5$	12
	No	$p_{\rm edgeon} < 0.2$, $p_{\rm features} > 0.5$	12
Bar	No	$p_{\rm bar} < 0.1$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.2$	24
	Int.	$0.2 < p_{\rm bar} < 0.4$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.2$	24
	Yes	$p_{\rm bar} > 0.8$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.2$	24
Visible spiral	No	$p_{\rm spiral} < 0.2$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.2$, $p_{\rm bar} < 0.1$	12
	Int.	$0.2 < p_{\rm spiral} < 0.8$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.2$, $p_{\rm bar} < 0.1$	12
	Yes	$p_{\rm spiral} > 0.8$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.2$, $p_{\rm bar} < 0.1$	12
Oblique bulge size	No	$p_{\rm nobulge > 0.6}$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.5$, $p_{\rm bar} < 0.2$	12
	Int.	$p_{\rm justnoticeable} > 0.6$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.5$, $p_{\rm bar} < 0.2$	12
	Yes	$p_{\rm obvious|dominant} > 0.5$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} < 0.5$, $p_{\rm bar} < 0.2$	12
Edge-on bulge shape	Round	$p_{\rm rounded} > 0.5$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} > 0.5$	12
	Boxy	$p_{\rm boxy} > 0.4$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} > 0.2$	12
	No bulge	$p_{\rm nobulge} > 0.5$, $p_{\rm features} > 0.5$, $p_{\rm edgeon} > 0.5$	12
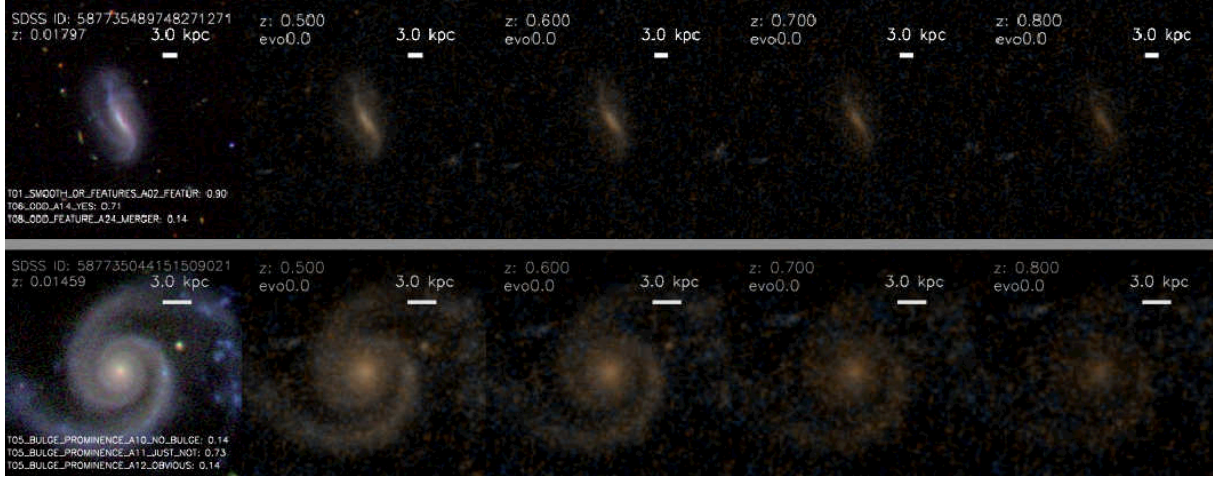
Figure 3. Examples of two galaxies which have been run through the FERENGI code to produce simulated HST images. The value of p_{features} for each panel is (1) Top row: $p_{\text{features}} = 0.9, 0.625, 0.35, 0.35, 0.225$ and (2) Bottom row: $p_{\text{features}} = 1.00, 0.875, 0.875, 0.625, 0.375$.

4.2 Correcting GZH morphologies for classification bias

The approach used in GZH for correcting the weighted classifications for user bias rests on the assumption that the *amount* of bias is a function of the apparent size and brightness of the image as seen on screen. This is controlled by two types of parameters: **intrinsic** properties of the galaxy itself, such as its physical diameter and luminosity, and **extrinsic** properties, such as the distance (redshift) of the galaxy and its relative orientation. There are likely other parameters that affect user accuracy, such as the proximity of close companions ("distraction bias"; see Johnson et al. 2015) or bias as a function of the individual user. The combination of all such parameters forms a high-dimensional space, and we have insufficient data to measure their individual effects. Instead, we use just two parameters that are intended to capture the bulk of the change in bias (based on GZ1/GZ2): a galaxy's r'-band surface brightness (μ_r; intrinsic) and redshift (z; extrinsic).

The change in bias as a function of μ_r and z is measured using the FERENGI images over all the evolutionary correction factors. We assume that the "true" (ie, debiased) vote fraction $f_{\mu,z}$ for a galaxy can be expressed as:

$$f_{\mu,z} = (f_{\mu,z=0.3}) \times e^{\frac{z-z_0}{\zeta}}, \tag{1}$$

where $f_{\mu,z=0.3}$ is the "calibrated" vote fraction at the lowest redshift in the FERENGI bins ($z = 0.3$) and ζ is a positive parameter that controls the rate at which f decreases with increasing redshift. This formula fits the data relatively well (with almost no exceptions, the vote fractions for featured galaxies decrease monotonically with increasing redshift), and the exponential function bounds the observed vote fractions between $f_{\mu,z=0.3}$ and zero. Figure 4 show examples of the change in vote fraction and their fits to Equation 1 for a random selection of galaxies in the FERENGI images.

We use the values of ζ for *all* sets of artifically redshifted galaxies to fit the overall distribution as a function of surface brightness, since we expect the correction being applied to vary as a function of the intrinsic galaxy properties. We restrict the galaxies that can be used to measure the calibration to those with data at the pivot redshift of $z = 0.3$, non-zero f_{features} at $z = 0.3$, and with a reasonable fit to the exponential model ($\Delta\chi^2 > 3.0$).

Figure 5 shows the results of fitting the FERENGI images with Equation 1; the correction is a weak function of galaxy surface brightness. Higher-surface brightness galaxies have stronger average corrections, likely because these galaxies are more likely to have larger f_{features} values at high redshifts. Low surface brightness galaxies are more likely to begin low and remain low; the bounded nature of the dropoff (and Poissonian-like variance among the individual voters) means that the average magnitude of ζ will be less.

We fit the data in Figure 5 with a linear function such that:

$$\log_{10}(\hat{\zeta}) = \zeta_0 + \zeta_1 \times \mu, \tag{2}$$

where $\hat{\zeta}$ is the correction factor applied to each galaxy as a function of surface brightness. The best-fit parameters to the linear fit (from least-squares optimization) are $\zeta_0 = 0.1$, $\zeta_1 = 1.4$. To make the final debiased correction, we modify the simple exponential form of Equation 1 to bound the debiased vote fractions between f and 1:

$$f_{\text{features,debiased}} = 1 - (1-f)e^{\frac{z-z_0}{\hat{\zeta}}}. \tag{3}$$

4.3 Results of ζ approach

In Figure 7 we examine the change in p_{features} for the FERENGI galaxies relative to their lowest simulated redshift. In this analysis, only galaxies whose lowest simulated redshift image was ($z_{\text{sim}} = 0.3$) were used (see Table 4), and only those which had detectable surface brightness measurements in SEXTRACTOR; this includes 3,950 of the total 6,466 images. For each simulated redshift value z, and at a fixed surface brightness μ, we plot $p_{\text{features,z}}$, the value measured at that simulated redshift, vs $p_{\text{features,z=0.3}}$, the value measured for the same galaxy imaged at $z = 0.3$.

Our objective is to use these data to predict, for a galaxy with a measured $p_{\text{features,z}}$ value, what its p_{features}
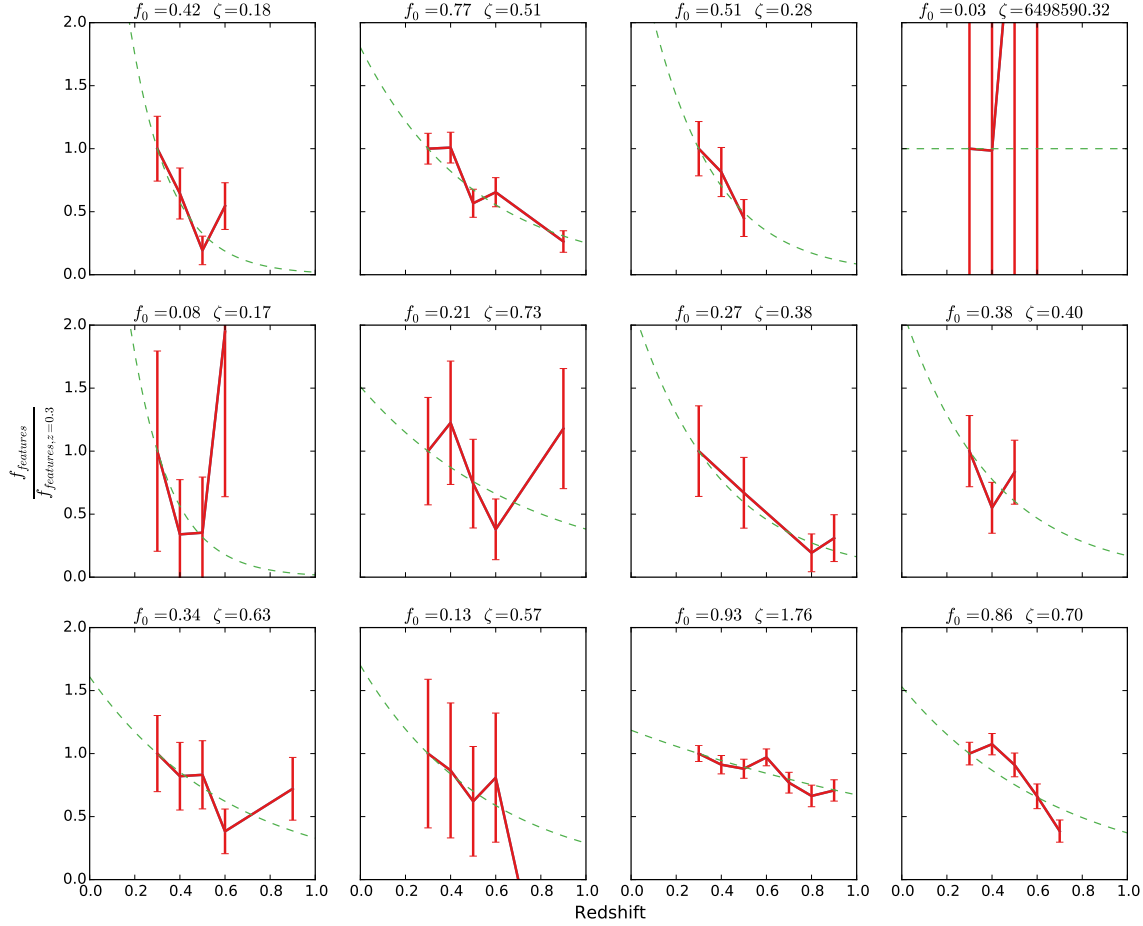
Figure 4. Behavior of the normalized, weighted vote fractions of features visible in a galaxy (f_{features}) as a function of redshift in the artifical FERENGI images. Galaxies are a random selection of images with $e = 0$ and at least three detectable images in redshift bins of $z \geq 0.3$. The measured vote fractions (red points) are fit with an exponential function (Equation 1); the best-fit parameters are given above each plot. Error bars are Poissonian, assuming a median of 40 votes per galaxy.

value *would have been* if it had been viewed at $z = 0.3$. This predicted value is defined as the debiased vote fraction $p_{\text{features,debiased}}$, and is calculated by applying a correction to the measured value of p_{features}, determined by the ζ function described in the previous section. A reliable predicted value can be obtained so long as the relationship between $p_{\text{features,z}}$ and $p_{\text{features,z=0.3}}$ is single-valued; that is, for a given $p_{\text{features,z}}$, there is exactly one corresponding value of p_{features} at $z = 0.3$.

Figure 7 shows that the relationship between $p_{\text{features,z}}$ and $p_{\text{features,z=0.3}}$ is *not* always single valued; hence, it is not appropriate to correct galaxies that lie in certain regions of surface brightness/redshift/p_{features} space. These regions tend to have low p_{features} values at high redshift, but a wide range of values at $z = 0.3$. These regions contain two morphological types of galaxies: First are genuine ellipticals, which have low values of p_{features} at both high and low redshift. Second are disks whose features become washed out

at high redshift; hence their p_{features} value at $z = 0.3$ may be quite high, while the value observed at high redshift is very low. This effect is strongest at high z and low μ, where features become nearly impossible to discern in the images.

Our criteria for determining whether a region of this space is single-valued, and therefore correctable, is as follows: In each surface brightness and redshift bin, we model the relationship between $p_{\text{features,z}}$ and $p_{\text{features,z=0.3}}$ by fitting the data with a polynomials of degrees 3, 2, and 1, and use the best fit out of the three. These fits are shown as the dashed black lines in Figure 7(a). Any flat regions of the polynomial fits are areas in which there is not a clear single-valued relationship between $p_{\text{features,z}}$ and $p_{\text{features,z=0.3}}$; we quantify this by setting a minimum slope cut of 0.4. Any data in which the polynomial fit has a slope less than this value is considered *not* one-to-one, and therefore "uncorrectable." These regions are highlighted in blue in figure 7(a). Uncolored (white) regions of the plot have suf-

[6a]

[6b]

[6c]

[6d]
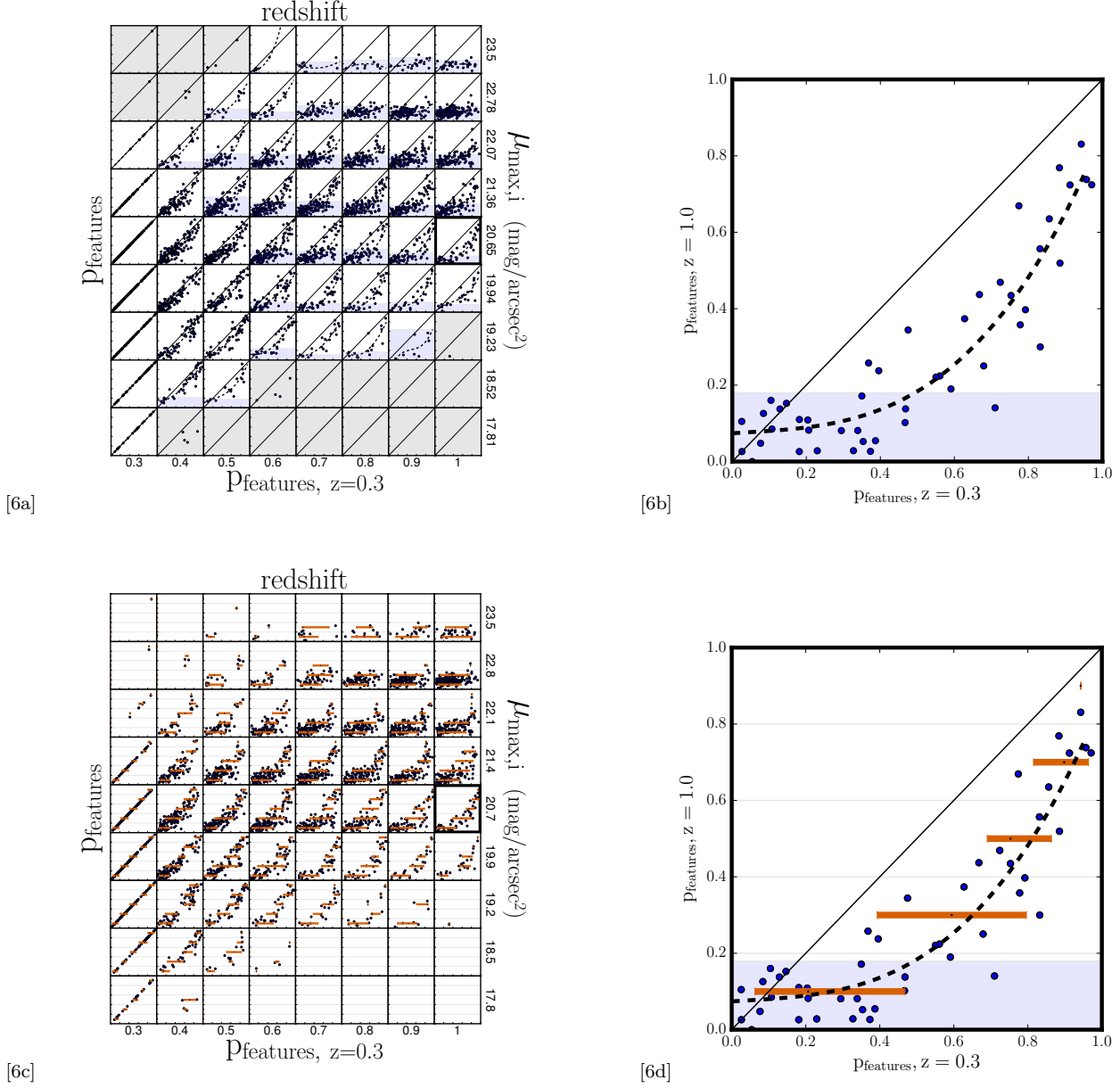
Figure 7. Effects of redshift bias in 3,950 images in the FERENGI sample. [6a]: Each point in a given redshift and surface brightness bin represents a unique galaxy. On the y-axis in each bin is the p_{features} value of the image of that galaxy redshifted to the value corresponding to that redshift bin. On the x-axis is the p_{features} value of the image of the same galaxy redshifted to $z = 0.3$. The dashed black lines represent the best-fit polynomials to the data in each square. The solid black line represents $p_{\text{features},z} = p_{\text{features},z=0.3}$. Regions in which there is a single-valued relationship between p_{features} at high redshift and at $z = 0.3$ are white; those in which there is not are blue, and those with not enough data ($N < 5$) are gray. [6b]: A larger version of the dark-outlined square in [6a], containing FERENGI galaxies that have been artificially redshifted to $z = 1.0$ and have surface brightnesses between $20.3 < \mu < 21.0$ (mag/arcsec^2). [6c]: The same data as [6a] is shown. Each z, μ bin is divided into 4 sub-bins to determine the range of intrinsic $p_{\text{features},z=0.3}$ for a given range of observed $p_{\text{features},z}$ values. In each sub-bin, the orange bars represent the inner 80th percentiles of the data, the boundaries of which are the lower and upper limits of the debiased values. [6d]: The same data as [6b], but highlighting the upper and lower limit regions.

ficiently high slopes for us to consider the relationship to be single-valued; galaxies in these regions are considered "correctable", and only these are used in measuring the parameters for the ζ function (Section 4.2). Only surface brightness/redshift bins with at least 5 galaxies were considered; regions with fewer than 5 galaxies we consider to have "not enough information" to determine the $p_{\text{features},z}$

and $p_{\text{features},z=0.3}$ relationship, these are colored gray in Figure 7(a).

The unshaded regions in Figure 7(a) define discrete ranges of redshift, surface brightness, and p_{features} a galaxy must have in order for the ζ approach to be confidently applied to a galaxy in the GZH sample. While the appropriate correctable regions were defined discretely, we assume
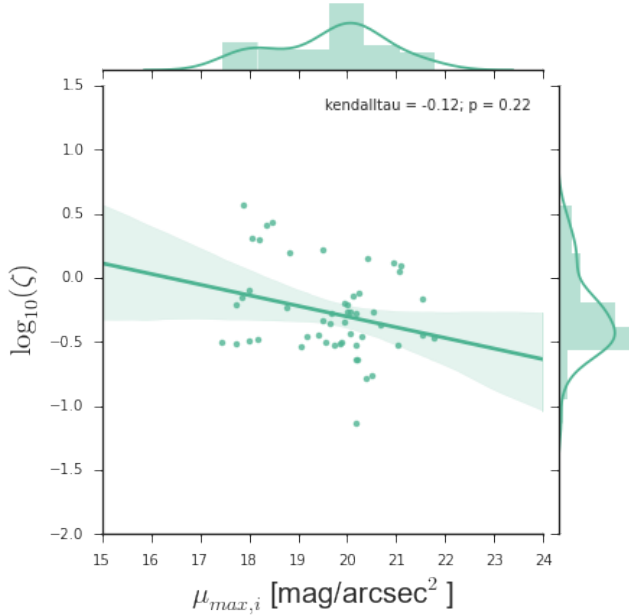
Figure 5. All fits for the vote fraction dropoff parameter ζ for f_{features} in the FERENGI galaxies as a function of surface brightness. This includes only the 37 galaxies with a reasonably bounded range on the dropoff $(-10 < \log(\zeta) < 10)$ and sufficient points to fit the function.
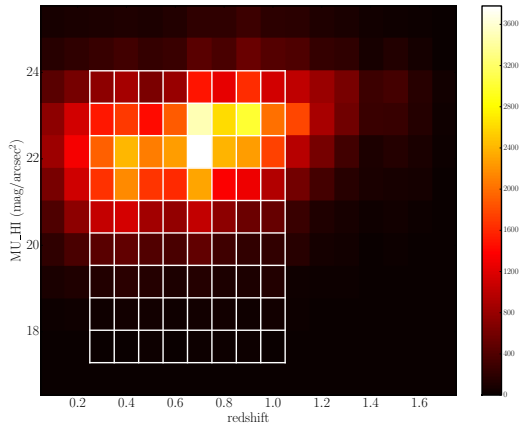
Figure 6. Surface brightness vs. redshift of 118,083 galaxies in the ACS sample. The white grid denotes the surface brightness and redshift range of the FERENGI images, subdivided in bins corresponding to fixed ranges used for analysis in Figure 7.

the true correctable region is a smooth function of z, μ, and p_{features}. To define this smooth space, we use a convex hull method to enclose the correctable and uncorrectable FERENGI galaxies is z-μ-p_{features} space. Due to scatter, the boundaries of the resulting hulls overlap. The boundaries are then adjusted until the contamination from both groups is minimized. We use the resulting hulls to define the correctable and uncorrectable regions for categorizing the Hubble galaxies. The results of this method and final categoriza-

Table 5. Distribution of FERENGI images analysed in Figure 7. Correctable images had a single-valued relationship between their measured p_{features} values at high and low redshifts (white regions in Figure 7). Uncorrectable images had a non single-valued relationship (blue regions). NEI images had undetermined relationships due to a lack of data $(N < 5)$ in their corresponding z-μ bins (gray regions).

	N	%
Correctable	1,884	48%
Uncorrectable	1,986	50%
NEI	80	2%
Total	3,950	100%

tion of the Hubble sample is displayed in Table 6. We find that of the galaxies at redshift higher than $z = z_0 = 0.3$, 17% of these are able to be debiased using the ζ method, 27% cannot be debiased, and 56% cannot be determined, due to a lack of redshift or information or due to a lack of FERENGI data corresponding to those galaxies' redshift/surface brightness values.

For the "uncorrectable" galaxies, those for which we cannot confidently assign a single debiased p_{features} value, we instead determine a likely *range* of debiased values, using a method visualized in Figure 7(c). Here we again use the FERENGI simulated data to analyze the range of intrinsic $p_{\text{features},z=0.3}$ values for any given observed p_{features} value, again as a function of surface brightness and redshift. In each z,μ bin, we examine the spread of intrinsic values of $p_{\text{features},z=0.3}$ for 4 ranges of observed p_{features}. We quantify the range of intrinsic values as the inner 80% of the data; this range is represented by the orange bars in Figure 7(c). For any galaxy which can't be directly debiased by the ζ method, then, we use these ranges to denote the upper and lower limits on what we expect $p_{features,z=0.3}$ to be for any observed value of p_{features}.

Stuff to do in this section:

• talk about where the Hubble sample falls in this space, reference table 6
• justify $N > 5$ and spread < 0.2 (or find a better way to choose criteria)
• check out corrections for correctable and NEI, show some sample images of corrected galaxies
• show some data for p_{bar}, determine or justify why we won't debias them

4.3.1   TODO LIST

We need to do:

• Calculate the magnitudes, surface brightnesses and sizes of the galaxies in the FERENGI images....
• Plot of magnitude distribution of galaxies in each of the four GZH subsamples with the magnitudes of our fake galaxies over plotted.
• Instructions of how to link the $z = 0$ p_X values for galaxies with a given size, magnitude (surface brightness) in the GZH images.

Table 6. Breakdown of what we can correct out of the GZH data, by sample.

	AEGIS	COSMOS	GEMS	GOODS-N	GOODS-S	GOODS FULL	SDSS	Total
Correctable	1,628	15,120	1,811	739	502	1,850	0	21,650
Uncorrectable	1,869	26,184	2,407	524	1,115	2,670	0	34,769
No Correction Needed ($z \leq 0.3$)	944	11,781	1,193	258	258	1,083	37,545	53,062
NEI	2,753	34,432	2,271	861	2,547	4,766	0	47,630
No Redshift Information	1,313	5,291	1,622	169	491	788	14,316	23,990
Total	8,507	92,808	9,304	2,551	4,913	11,157	51,861	181,101
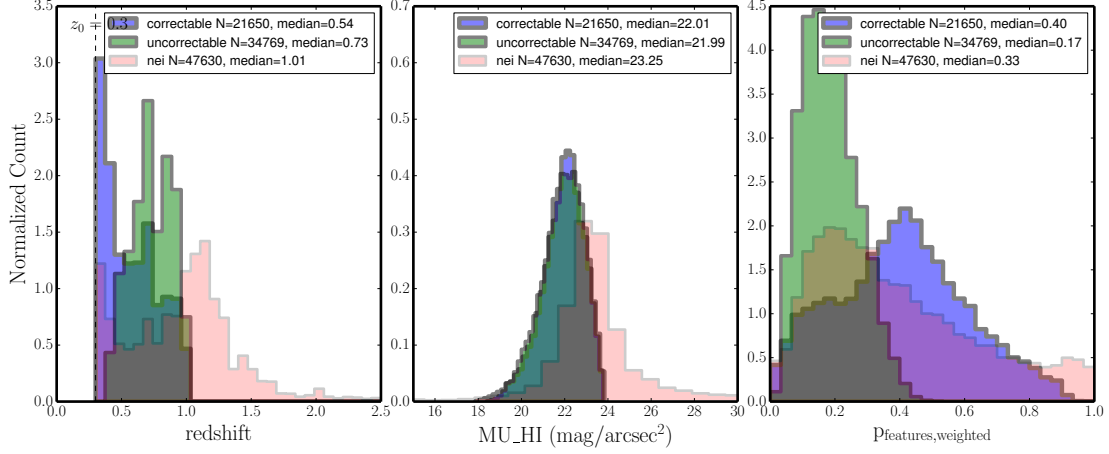
Figure 8. Distributions of redshift, surface brightness, and $p_{features}$ for correctable (purple), uncorrectable (green), and NEI (pink) galaxies in the full GZH sample. The uncorrectable galaxies tend towards higher redshift, slightly lower in surface brightness, and lower values of $p_{features}$ than the correctable galaxies. The long tail of NEI galaxies in redshift and surface brightness demonstrates the limits of the FERENGI sample, for which there is no data at $z > 1$ or $\mu > 24$.
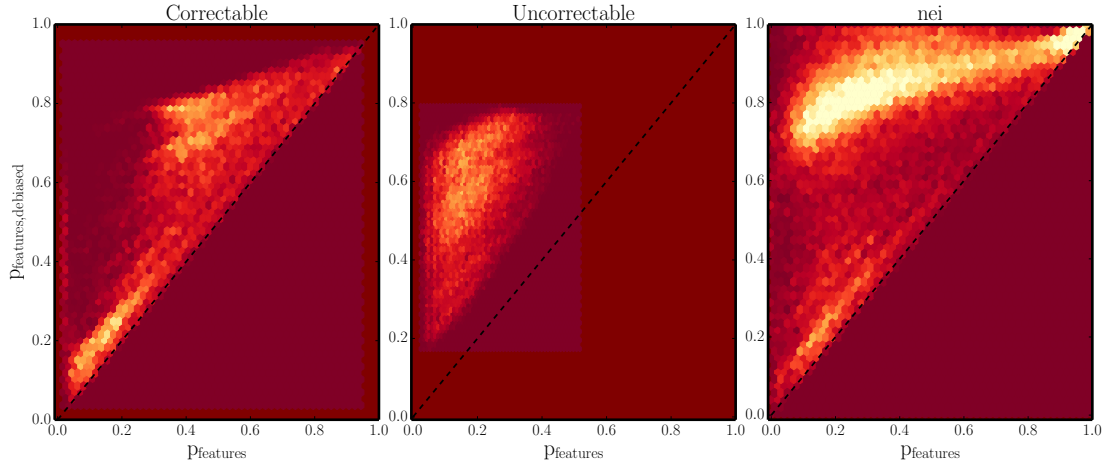
Figure 9. Debiased $p_{features}$ corrected to $z = 0.3$ vs weighted $p_{features}$ for the correctable (left), uncorrectable (middle), and NEI (right) galaxies in the GZH sample.

4.4 Morphological measurements in GZH beyond Task 1 - effects of debiasing?

4.5 Duplicate images

4.6 Effect of changing depth for GOODS

4.7 SDSS Stripe 82 images

4.8 Fake AGN

5 THE CATALOG

The data release for GZH includes morphological data for 181,101 galaxies. The full table can be accessed at some website. We also include a secondary metadata table, which contains data from a variety of sources explained in Section 2.

For each galaxy we list its unique objid, as well as the source's RA, DEC, and survey (AEGIS, COSMOS, GEMS, GOODS North (full and shallow depth), GOODS South (full and shallow depth), SDSS. For each of the 40 (?) questions in the GZH decision tree, the following classification data is provided: For each question, N_{votes} is the number of users to answer that question. For each unique answer, `fraction` is the fraction of users to select that answer (N_{answer}/N_{votes}, and `weighted` is the weighted fraction, which takes into account user consistency (Section 2.4).

The GZH vote fractions can be largely dependent on the resolution of the image. Two otherwise morphologically identical galaxies which differ significantly in redshift, brightness, or size may result in very different vote fractions for any given question, given that many features of a galaxy are difficult to discern in less-resolved images (bars, spiral arms, disk structure, etc). For this reason, it is necessary to take caution utilizing vote fractions as cut-offs to determine morphological structure; we offer guidelines for careful classification in Section 6.

We corrected for the biases described for the first question of the GZH decision tree, which asks "Is the galaxy smooth and round, with no sign of a disk?" The method is described in Section 4. For this question, we provide the additional parameters `debiased`, `lower limit`, `upper limit`, and `best` vote fractions. The `best` fraction for $p_{features}$ is chosen based on the categorization of the galaxy: if it is "correctable", `best = debiased`, if "uncorrectable", `best = lower limit`, and if neither, `best = weighted`. The debiased vote fractions for p_{smooth} were calculated on the critera that vote fractions for all answers must sum to unity. Explicitly: $p_{smooth} = 1 - p_{features} - p_{artifact}$.

To include: detailed description of each column in the machine-readable tables.

Needs some test cases for extracting a given set of objects (eg, clump galaxies in a particular redshift range) and evaluation of the results. Possibly include suggested thresholds, á la GZ2.

6 USING THE CATALOG

Include cookbook for selecting morphologies.

Table 7. GZH morphological classifications

OBJNO	RA	DEC	Imaging	t01_smooth_or_features_ Correctable Category	N_{votes}	fraction	weighted	t01_smooth_or_features_a01_smooth_ debiased	lower limit	upper limit	best	...
10000034	188.93	62.23	AEGIS	correctable	72	10	10.0	0.625	0.625	0.629	0	
10000162	160.99	11.70	AEGIS	uncorrectable	84	17	17.0	0.810	0.810	0.810	1	
5001356	210.80	54.15	GOODS-N-Shallow	correctable	88	18	18.0	0.783	0.783	0.722	0	
9003947835	176.54	8.74	GOODS-S-Shallow	nei	48	16	16.0	1.000	1.000	1.000	1	
GDS S 16203	186.44	47.88	GOODS-FULL	nei	89	17	17.0	0.739	0.739	0.739	0	
86474690312438284	188.93	62.64	SDSS	zlt3	91	0	0.0	0.000	0.000	0.000	0	
20061690	160.99	11.70	COSMOS	nei	65	4	3.4	0.333	0.295	0.295	0	
20061692	176.54	8.45	COSMOS	correctable	75	14	14.0	0.609	0.629	0.666	0	
90022947	230.24	84.20	GEMS	zlt3	129	12	12.0	0.462	0.462	0.492	0	
20063829	203.23	93.20	COSMOS	correctable	69	18	17.0	0.900	0.895	0.840	1	

Note. — Full version is online at some website, here are 10 rows out of 181,101 in the full version. *These numbers are super fake right now - put real values in for final version!*

7 ANALYSIS

7.1 Demographics of morphology

Summarize the broad trends that are seen regarding the fraction of galaxies with various morphologies, how that relates to color, size, etc. Briefly discuss results as compared with literature and theory.

7.2 Comparison to other catalogs

Compare GZH data to:

- Scarlata et al. (ZEST; 2007) (COSMOS)
- Tasca (COSMOS)
- Cassata (COSMOS)
- Zajmoski (COSMOS)
- GEMS morphologies?
- AEGIS morphologies?
- GOODS N/S morphologies?
- expert visual inspection?

Address trends seen in broad morphological classes, possible reasons for difference. Also should attempt to map between the GZH vote fractions and whatever classification systems are used in the above systems.

8 SUMMARY

Now people go and do science with these awesome GZH classifications.

REFERENCES

Astropy Collaboration et al., 2013, A&A, 558, A33
Bamford S. P. et al., 2009, MNRAS, 393, 1324
Barden M., Jahnke K., Häußler B., 2008, ApJS, 175, 105
Caldwell J. A. R. et al., 2008, ApJS, 174, 136
Davis M. et al., 2007, ApJ, 660, L1
Giavalisco M. et al., 2004, ApJ, 600, L93
Griffith R. L. et al., 2012, ApJS, 200, 9
Ilbert O. et al., 2013, A&A, 556, A55
Johnson L. C. et al., 2015, ApJ, 802, 127
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Momcheva I. G. et al., 2015, ArXiv e-prints, 1510.02106
Rix H.-W. et al., 2004, ApJS, 152, 163
Scarlata C. et al., 2007, ApJS, 172, 406
Scoville N. et al., 2007, ApJS, 172, 1
Strauss M. A. et al., 2002, AJ, 124, 1810
Vanderplas J., Connolly A., Ivezić Ž., Gray A., 2012, in Conference on Intelligent Data Understanding (CIDU), pp. 47 –54
Waskom M. et al., 2015, seaborn: v0.6.0 (june 2015)
Willett K. W. et al., 2013, MNRAS, 435, 2835
York D. G. et al., 2000, AJ, 120, 1579