

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

23 March 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:

The morphology of galaxies encodes information on the orbital parameters and assembly history of the contents, including gas, dust, stars, and the central black hole. The morphology is also closely related to the local environment of the galaxy, as mutual interactions such as gravitational tides, shocks in cluster environments and direct mergers can all change the shape of the galaxy's potential. For M^* galaxies in the local Universe, this typically manifests at the most basic level as the difference between bulge-dominated, virialized systems resembling ellipticals (early-types) and disk-dominated, rotationally-supported disks (late-types) frequently exhibiting spiral arms. This dichotomy has been used to explore much of the astrophysics encoding galaxy formation and evolution, and has been shown to be closely linked with other galactic properties such as stellar mass, halo mass, luminosity, black hole activity, size, and the relative ages of the stellar populations.

The advent of larger telescopes in an increasing range of observing wavelengths has revealed that the distribution and properties of galaxy morphology have strongly evolved over the lifetime of the Universe. At redshift $z \simeq 1$ (roughly 6 Gyr after the Big Bang), many galaxies are still in the process of assembling the baryonic mass required to reproduce the structures seen in the present day. This occurs in a variety of pathways, including accretion of baryons from large-scale galactic filaments onto halos via streaming, mergers of individual dark matter halos along with their baryons, conversion of gas into stars via gravitational collapse and star formation, etc. The process can also be slowed or even reversed via feedback from stellar winds, supernovae, and active black holes. Each of these processes affects on the galaxy morphology, and so an accurate measurement of the demographics as a function of redshift provides an extremely powerful observational constraint on the physics taking place (for a recent review see Conselice 2014).

Theoretical predictions for the morphology of galaxies as a function of redshift are primarily computed within the Λ CDM cosmological framework. Full treatments of this model gravitational interactions between baryons and dark matter, hydrodynamics of the gas, and baryonic physics related to star formation and evolution. The most advanced simulations now include volumes up to ~ 100 Mpc³ while simultaneously resolving the smaller scales necessary to reproduce the baryonic physics (Vogelsberger et al. 2014; Schaye et al. 2015). Such simulations predict clustering of galaxies on large scales in a hierarchical assembly model (Silk & Mamon 2012). The structure of individual galaxies is affected by the merger history (Toomre & Toomre 1972; Hopkins et al. 2010), local environment (such as the morphology-density relation; Dressler 1980), initial dark halo mass, and many other factors. Morphologies of individual simulated high-mass galaxies at $z \sim 2 - 3$ commonly show kpc-scale “clumpy” structures and few galaxies that are either smooth or well-ordered spirals; asymmetric profiles with strong density constraints dominate down to at least $z \gtrsim$ (Genel et al. 2014).

Observational studies of galaxies at high-redshift also display a wide range of morphological types, many of which are rare or absent at $z \sim 0$. These include spheroids and disks (akin to the ellipticals/spirals seen in the local Universe), but also a significant population of massive, more irregular galaxies, including mergers, tadpoles, chains, double-clumps, and clump-clusters (Elmegreen et al. 2005, 2007; Cameron et al. 2011; Förster Schreiber et al. 2011; Kartaltepe et al. 2015). In contrast, while grand-design spirals have been observed as far back as $z = 2.18$ (Law et al. 2012a), their spatial density suggests that they are exceedingly rare, with a very low overall disk fraction (Mortlock et al. 2013). Current observational data thus strongly suggests that the classical Hubble sequence/tuning fork (Hubble 1936) is not a suitable framework for characterizing high-redshift morphology.

Space-based observatories, particularly the *Hubble*

Space Telescope, have been responsible for the bulk of imaging studies of high-redshift galaxies. Observations of fixed fields with very deep imaging (eg, Giavalisco et al. 2004; Davis et al. 2007; Grogin et al. 2011; Scoville et al. 2007; Beckwith et al. 2006; Williams et al. 1996) give the photometric sensitivity necessary to detect L^* galaxies at $z > 1$, while also providing the angular resolution to distinguish internal structure and characterize the morphology. While these measurements are helped by the fact that the angular diameter distance is relatively flat at $z > 1$ in a flat Λ CDM cosmology, the scales are only of the order $\sim 5 - 10$ kpc/'' (Wright 2006). *HST* can thus resolve much of the structure for a Milky Way-sized galaxy (at least for distinguishing a disk from a bulge), but will be limited for more compact structures. Since the size of galaxies evolves as roughly $r \propto (1+z)^{-1}$ (Law et al. 2012b), the more compact sizes of earlier galaxies make detailed morphologies a challenge even for *HST* Chevance et al. (2012). The public availability of more than 10^5 galaxies in archival imaging across various studies gives a data sample with the potential for high statistical significance.

One of the major difficulties in studying the morphologies of galaxies in large samples lies in the system of measurement. Visual classification by experts has been used for many years (eg, Hubble 1926; de Vaucouleurs 1959; Sandage 1961; van den Bergh 1976; Nair & Abraham 2010; Bailward et al. 2011; Kartaltepe et al. 2015). These methods have advantages in using the significant processing power of the human brain to identify patterns, but suffer from issues such as lack of scaling to large surveys and potential issues with replicability and calibration. Automated measurements, both parametric (Peng et al. 2002; Simard et al. 2011; Lackner & Gunn 2012) and non-parametric (Conselice 2003; Lotz et al. 2004; Scarlata et al. 2007; Freeman et al. 2013), scale well to arbitrary sample sizes, but do not always fully capture the relevant features, especially for asymmetric galaxies that become increasingly common at high redshifts. The Galaxy Zoo project (Lintott et al. 2008, 2011) utilizes crowdsourced visual classifications to measure galaxies in color-composite images. With $> 2 \times 10^5$ classifiers, this allows for multiple independent classifications of each image which are combined and calibrated to give a distribution of vote fractions proportional to the probability of a feature being visible. While the crowdsourced data require extensive calibration (Bamford et al. 2009; Willett et al. 2013), they have a proven reliability and have been used in dozens of papers (eg, Land et al. 2008; Bamford et al. 2009; Darg et al. 2010; Masters et al. 2011; Skibba et al. 2012; Simmons et al. 2013; Schawinski et al. 2014; Willett et al. 2015).

This paper presents the results from the Galaxy Zoo Hubble (GZH) project. GZH was the third phase of Galaxy Zoo, following its initial results classifying $\sim 900,000$ SDSS images into primarily early/late types (Lintott et al. 2011) and Galaxy Zoo 2, which covered $\sim 250,000$ SDSS images using a more detailed classification scheme that included bars, spiral arms, and galactic bulges (Willett et al. 2013). GZH used a similarly detailed classification scheme, but focused for the first time on images of high-redshift galaxies taken with the *Hubble Space Telescope*.

We describe the sample selection and creation of the images used for GZH in Section 2. Section 3 describes the GZH

interface and the collection of classifications. Section 4 outlines the process used to calibrate and “debias” the crowd-sourced votes. Section 5 gives the main results as a catalog, with several examples of how the data may be queried in Section 6. Section 7 gives a short overview of the observed morphological demographics and compares them to several other catalogs, with a summary in Section 8.

This paper uses the WMAP9 cosmology parameters of $(\Omega_m, \Omega_\Lambda, h) = (0.258, 0.718, 0.697)$ (Hinshaw et al. 2013).

2 SAMPLE AND DATA

The GZH project contains images drawn from a number of different dedicated surveys and sample selection criteria, which we describe below. The majority of the images (as implied by the project name) are taken directly from *Hubble Space Telescope* Legacy Surveys,

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 $\alpha = 14^{\text{h}}17^{\text{m}}, \delta = +52^\circ30'$. The strip was originally selected due to low extinction and Galactic/zodiacal emission, making it a prime target for multi-wavelength 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 ($\alpha = 12^{\text{h}}36^{\text{m}}, \delta = +62^\circ14'$) and the Chandra Deep Field-South ($\alpha = 03^{\text{h}}32^{\text{m}}, \delta = -27^\circ48'$). 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 5 σ 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 $\alpha = 10^{\text{h}}00^{\text{m}}28^{\text{s}}, \delta = +02^\circ12'21''$. 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% com-

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 $0.03''/\text{pixel}$. 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 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 compiled redshifts from a variety of sources to include in the GZH catalog. For each galaxy, the redshift selected is in the `Z_BEST` column of the data (see Table 7), its type (spectroscopic: `SPEC_Z`, photometric: `PHOTO_Z`, or grism: `GRISM_Z`) is listed in the column `Z_BEST_TYPE`, and the source catalog (**Griffith** (Griffith et al. 2012), **3DHST** (Momcheva et al. 2015), **MUSYC** (Cardamone et al. 2010), or **UltraVISTA** (Ilbert et al. 2013)) of the redshift is in column `Z_BEST_SOURCE`.

For galaxies which have redshifts from multiple sources, we use the following algorithm to select the `Z_BEST` redshift. We first prioritize spectroscopic redshifts; these are provided in the Griffith, 3DHST, and MUSYC catalogs. If a high quality spec-z exists in the Griffith catalog we use that, else 3DHST, else MUSYC. We show in Figure 2 that over 98% of the the spec-z's are consistent with each other, and therefore the priority order of selection makes no negligible difference. If no spectroscopic redshifts are available, we compare the $1-\sigma$ errors of the photometric (Griffith, 3DHST, MUSYC, UltraVISTA) and grism (UltraVISTA) redshifts, and use the redshift with the smallest error. Table 2 shows the results of this selection.

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

Table 2. GZH redshifts by sur

Survey	Griffith		3DHST		MUSYC		Ultra	
	spec-z	photo-z	spec-z	grism-z	photo-z	spec-z	photo-z	
AEGIS	3,656	2,941	12	515	249	0	0	
COSMOS	7,201	77,435	35	358	26	0	0	
GEMS	387	628	6	99	40	279	7,304	
GOODS-N	1,947	37	418	1,545	1,381	0	0	
GOODS-S	1,080	4	327	1,348	281	816	1,184	
SDSS	0	0	0	0	0	0	0	
Total	14,271	81,045	798	3,865	1,977	1,095	8,488	

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.

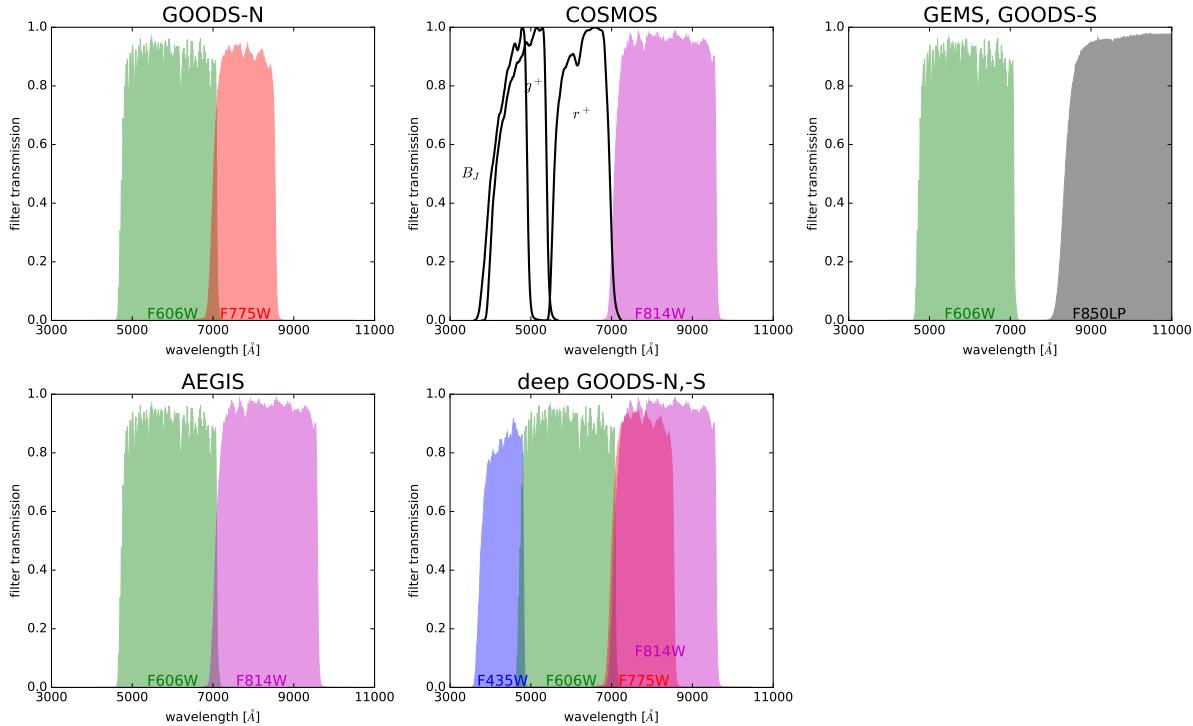
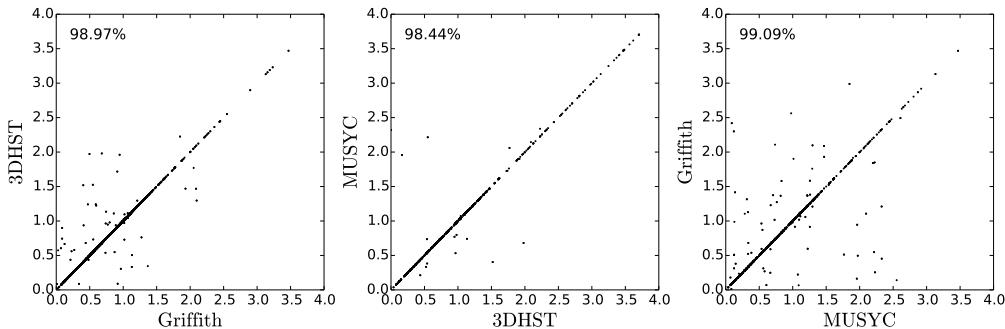
In order to test and correct for the effects of redshift, we generated a set of calibration images. These images consist of the same galaxy as it would appear over a variety of redshifts. The input images are from the SDSS (York et al. 2000; Strauss et al. 2002) and are processed using the FERENGI code (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)

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
<i>GOODS-N</i>	—	—	—	—	2551
<i>GOODS-S</i>	—	—	—	—	4785
total	—	—	—	8310	113166

**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.**Figure 2.** Spectroscopic redshifts from Griffith, 3DHST, and MUSYC catalogs. The number in the upper left of each plot is the percentage of redshifts which agree within $\Delta z < 0.05$ between the two catalogs being compared in each panel. Within this range there is over 98% agreement in redshifts between all three catalogs.

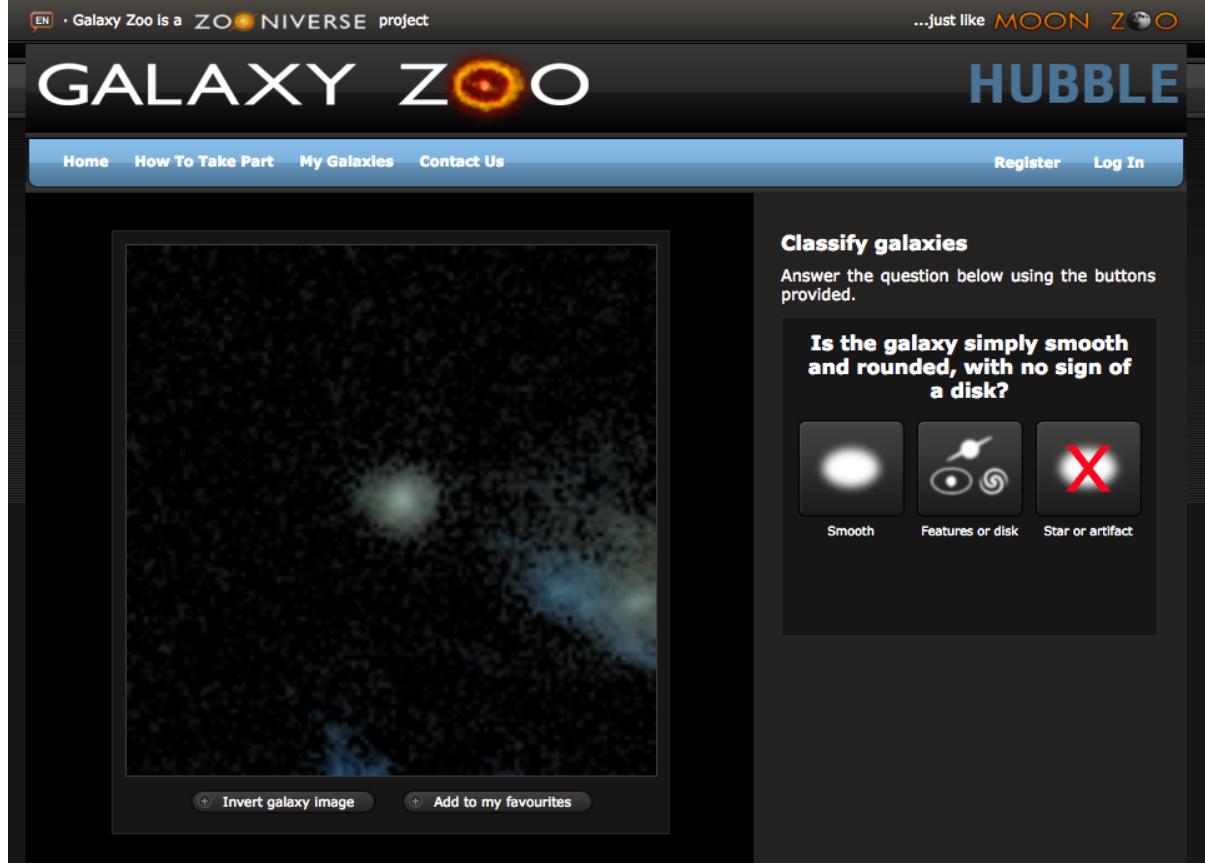


Figure 3. Screenshot of the Galaxy Zoo: Hubble interface showing an example COSMOS image at the first step in the decision tree.

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.

The selection criteria for the different morphological categories is summarised in Table 3. The surface brightness selection ($N_\mu = 3$) was (1) low: $\mu > 21.5 \text{ mag arcsec}^{-2}$; (2) mid: $20.5 < \mu < 21.5 \text{ mag arcsec}^{-2}$; and (3) high: $\mu < 20.5 \text{ mag arcsec}^{-2}$. 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 to $e = -3.5$ in increments of $\Delta e = 0.5$. Figure 4 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

Table 4. Summary of FERENGI artificial redshifting

z_{target}	N_{zbins}	$N_{\text{evolution}}$	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

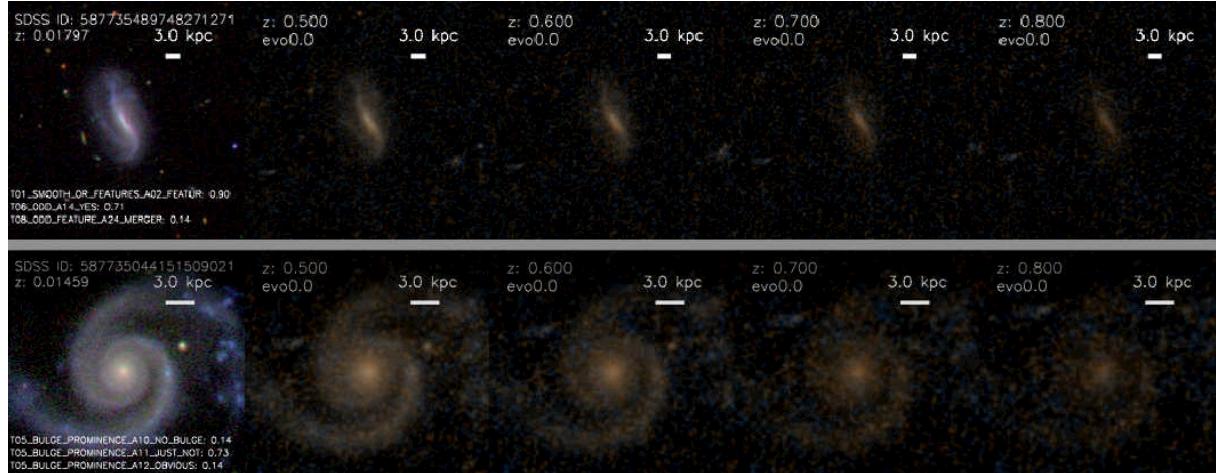
e values selected. Table 4 summarizes the total sample of redshifted images produced for GZH.

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.

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

Morphology	Label	Selection	N_{objects} [$N_z \times N_\mu$]
Features	Yes	$p_{\text{features}} > 0.8, p_{\text{odd}} < 0.1$	12
	Int.	$0.3 < p_{\text{smooth}} < 0.6, p_{\text{odd}} < 0.1$	12
	No	$p_{\text{smooth}} > 0.8, p_{\text{odd}} < 0.1$	12
Merger	No	$p_{\text{features}} > 0.8, p_{\text{odd}} < 0.1, p_{\text{merger}} < 0.1$	12
	Int.	$p_{\text{odd}} > 0.5, 0.1 < p_{\text{merger}} < 0.4$	12
	Yes	$p_{\text{odd}} > 0.5, p_{\text{merger}} > 0.4$	12
Edge-on	Yes	$p_{\text{edgeon}} > 0.8, p_{\text{features}} > 0.5$	12
	Int.	$0.4 < p_{\text{edgeon}} < 0.8, p_{\text{features}} > 0.5$	12
	No	$p_{\text{edgeon}} < 0.2, p_{\text{features}} > 0.5$	12
Bar	No	$p_{\text{bar}} < 0.1, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.2$	24
	Int.	$0.2 < p_{\text{bar}} < 0.4, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.2$	24
	Yes	$p_{\text{bar}} > 0.8, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.2$	24
Visible spiral	No	$p_{\text{spiral}} < 0.2, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.2, p_{\text{bar}} < 0.1$	12
	Int.	$0.2 < p_{\text{spiral}} < 0.8, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.2, p_{\text{bar}} < 0.1$	12
	Yes	$p_{\text{spiral}} > 0.8, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.2$	12
Oblique bulge size	No	$p_{\text{bulge}} > 0.6, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.5, p_{\text{bar}} < 0.2$	12
	Int.	$p_{\text{justnoticeable}} > 0.6, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.5, p_{\text{bar}} < 0.2$	12
	Yes	$p_{\text{obvious}} \text{dominant} > 0.5, p_{\text{features}} > 0.5, p_{\text{edgeon}} < 0.5, p_{\text{bar}} < 0.2$	12
Edge-on bulge shape	Round	$p_{\text{rounded}} > 0.5, p_{\text{features}} > 0.5, p_{\text{edgeon}} > 0.5$	12
	Boxy	$p_{\text{boxy}} > 0.4, p_{\text{features}} > 0.5, p_{\text{edgeon}} > 0.2$	12
	No bulge	$p_{\text{nobulge}} > 0.5, p_{\text{features}} > 0.5, p_{\text{edgeon}} > 0.5$	12

**Figure 4.** 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: $p_{\text{features}} = 1.00, 0.875, 0.875, 0.625, 0.375$.

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:

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

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 5 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 artificially 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

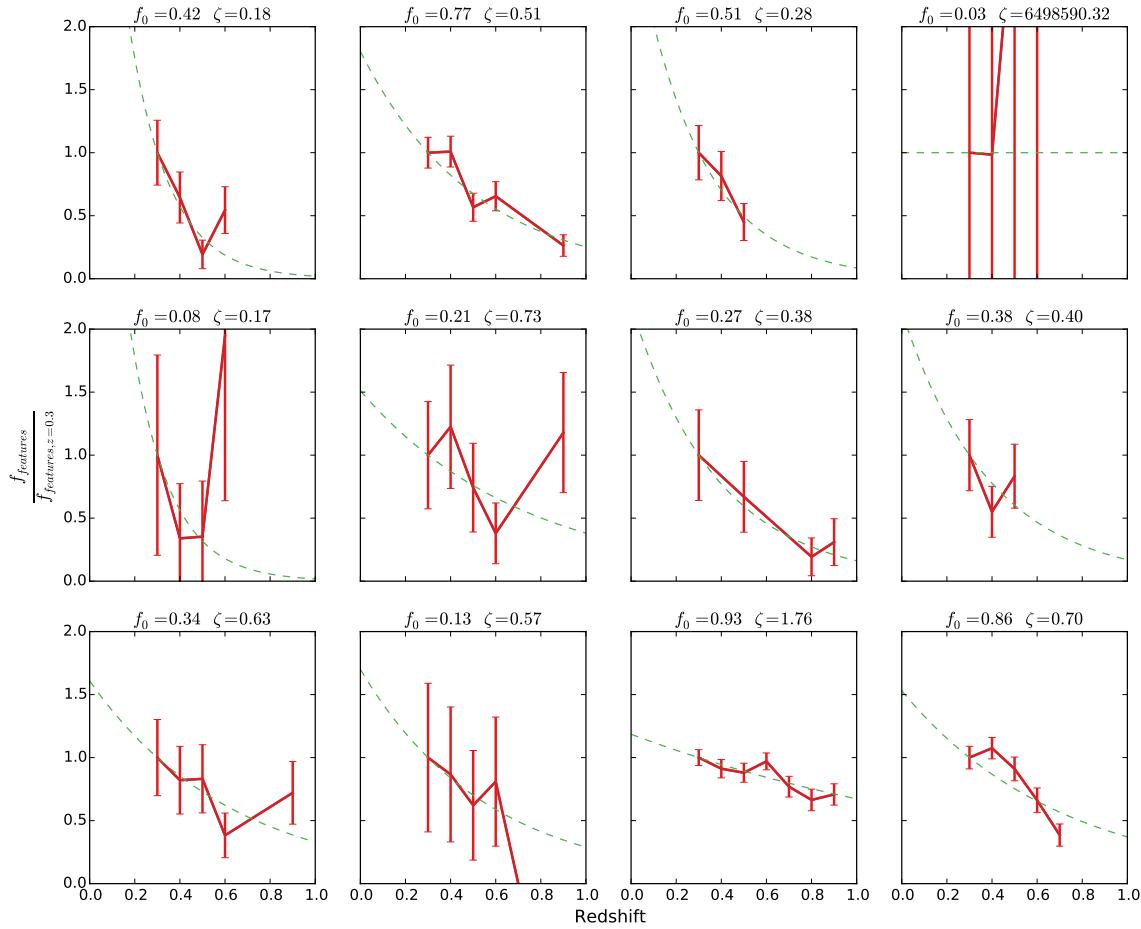


Figure 5. Behavior of the normalized, weighted vote fractions of features visible in a galaxy (f_{features}) as a function of redshift in the artificial FERENGI images. Galaxies are a random selection of images with $e = 0$ and at least three detectable images in redshift bins of $z \geq 0.3$. The measured 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.

$z = 0.3$, non-zero f_{features} at $z = 0.3$, and with a reasonable fit to the exponential model ($\Delta\chi^2 > 3.0$).

Figure 6 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 6 with a linear function such that:

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

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:

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

4.3 Results of ζ approach

In Figure 8 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

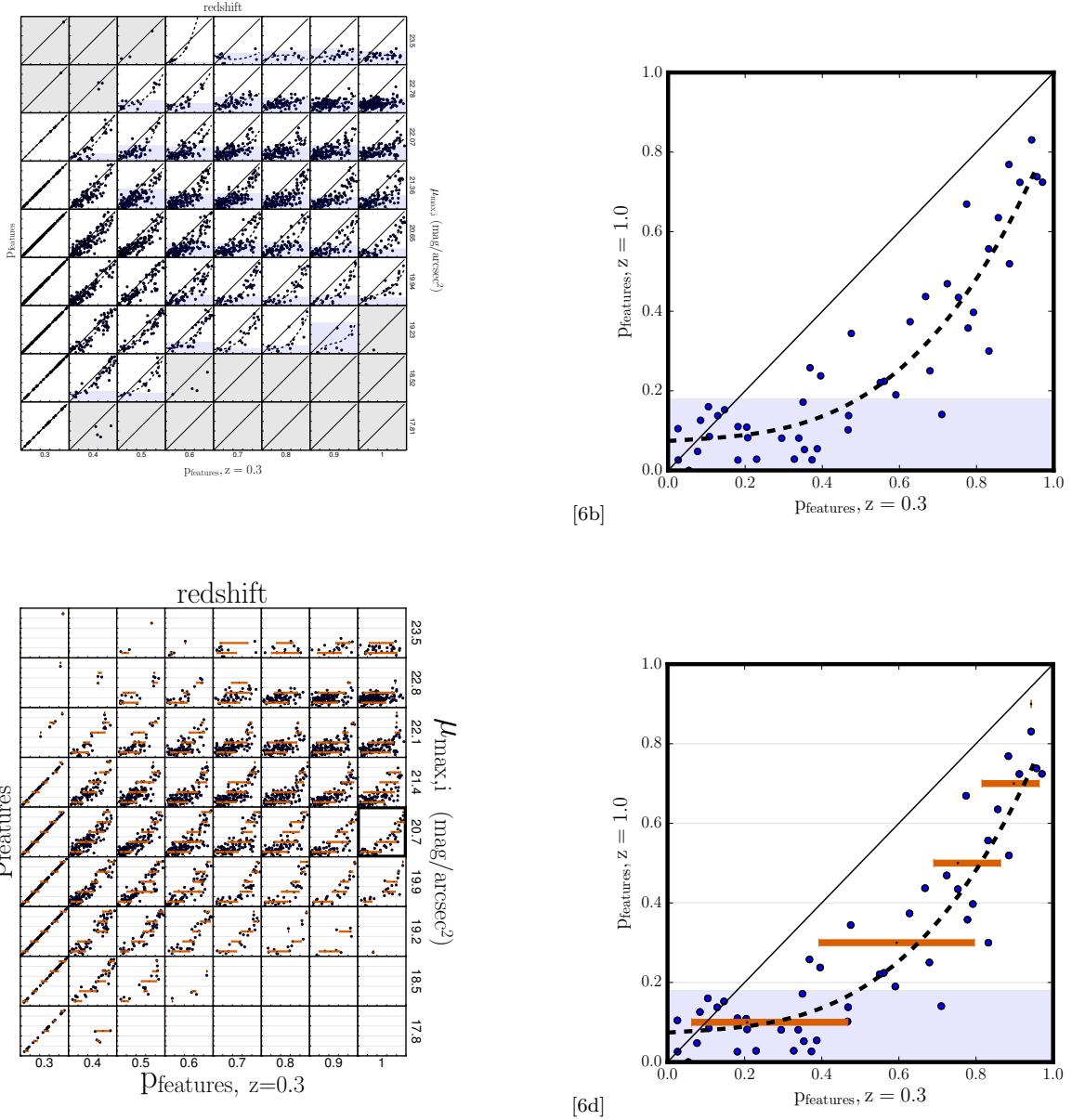


Figure 8. Effects of redshift bias in 3,950 images in the FERENGI sample. [6a]: Each point in a given redshift and surface brightness bin represents a unique galaxy. On the y-axis in each bin is the p_{features} value of the image of that galaxy redshifted to the value corresponding to that redshift 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 $20.3 < \mu < 21.0$ ($\text{mag}/\text{arcsec}^2$). [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.

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} 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 8 shows that the relationship between $p_{\text{features},z}$

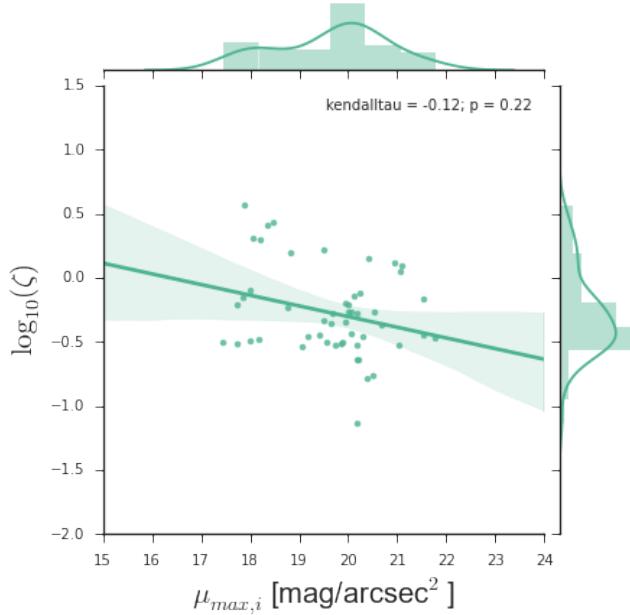


Figure 6. All fits for the vote fraction dropoff parameter ζ for f_{features} in the FERENGI galaxies as a function of surface brightness. This includes only the 37 galaxies with a reasonably bounded range on the dropoff ($-10 < \log(\zeta) < 10$) and sufficient points to fit the function.

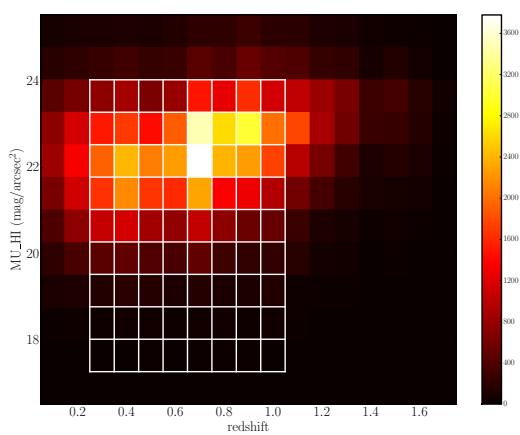


Figure 7. Surface brightness vs. redshift of 118,083 galaxies in the ACS sample. The white grid denotes the surface brightness and redshift range of the FERENGI images, subdivided in bins corresponding to fixed ranges used for analysis in Figure 8.

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 8(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 8(a). Uncolored (white) regions of the plot have sufficiently 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 8(a).

The unshaded regions in Figure 8(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 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 in 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 categorization 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 8(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 8(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 $p_{\text{features},z=0.3}$ to be for any observed value of p_{features} .

Table 5. Distribution of FERENGI images analysed in Figure 8. Correctable images had a single-valued relationship between their measured p_{features} values at high and low redshifts (white regions in Figure 8). 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 $z\text{-}\mu$ bins (gray regions).

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

4.4 Challenges of debiasing questions beyond “smooth or features”

Each FERENGI image does not have the same number of users answering each question, due to the structure of the decision tree. Every user answers the first question, “Is the galaxy smooth and rounded, with no sign of a disk?”, as such the vote fractions p_{smooth} , p_{features} , and p_{artifact} are all computed with the minimum statistical error for any question, with roughly 40 total answers (see Section ??). The number of users to answer any subsequent question, however, is always equal to or less than the number to answer the preceding question. For this reason, some galaxies may have very few (or even zero) answers to a question further down the tree (see Figure make-figure-of-count-distribution-for-each-question). To minimize statistical error in computing vote fractions, a cut on the number of answers to a given question is always implemented.

In the FERENGI data, we find that this places large limitations on the amount of information we can extract for the higher order questions. We require that at least 5 users answer each question for a galaxy image at $z = 0.3$ and its image at higher z . This requirement placed on both images is not met by a significant number of galaxies for questions beyond question 1. Without sufficient galaxies in each surface brightness/redshift bin, we cannot accurately measure a relationship between vote fractions and redshift; for this reason we only offer debiased vote fractions for question 1. **perhaps compute number of galaxies that can be fit to zeta for each question, show a table? overkill?** In Section A2 we show results of an attempt to measure ζ for p_{bar} .

- 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

4.4.1 TODO LIST

We need to:

- 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.

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

4.6 Duplicate images

4.7 Effect of changing depth for GOODS

4.8 SDSS Stripe 82 images

4.9 Fake AGN

We also classified a set of images designed to measure the effect of AGN on morphological classifications. This could occur since galactic nuclei can have bright, unresolved optical emission that could either mimic the effect of a strong bulge or otherwise obscure classification. The presence of an AGN is simulated by modeling the PSF for Hubble for a variety of images and then inserting a bright source near the center of the galaxy. For each galaxy, the simulated AGN can have one of three colors – either blue, red, or flat (white) as seen in the color images – and a range of brightnesses such that $L_{\text{ratio}} \equiv L_{\text{galaxy}}/L_{\text{AGN}}$ is in $(0.2, 1.0, 2.0, 5.0, 10.0, 50.0)$. Combining these parameters generates 15 images with different simulated AGN for each chosen host.

Images with simulated AGN were classified in the main interface in an identical manner and evenly distributed with unaltered images of the galaxies. Classifiers were not explicitly told that the images had been altered, as the goal was to measure the effect on normal classifications in as unbiased a manner as possible.

Two sets of simulated AGN were measured in GZH. The first set (version 1) was assembled from 95 galaxies from GEMS imaging and PSFs from `daophot`. The second set (version 2) was assembled from 96 galaxies with GOODS-S imaging; this version used deeper imaging and improved PSFs from `TinyTim`.

5 THE CATALOG

The data release for GZH includes morphological data for 181,101 images (generated from a total of 150,771 unique galaxies). The full table can be accessed at <http://data.galaxyzoo.org>. We also include a secondary metadata table, which contains data from a variety of sources explained in Section 2.

Each image is listed under a unique project ID (eg AHZ000001); the actual galaxy in the image is identified by the combination of the OBJNO and original survey. For each of the 55 responses 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 ($N_{\text{answer}}/N_{\text{votes}}$), and `weighted` is the weighted fraction, which takes into account user consistency (Section 2.4).

The GZH vote fractions can be largely dependent on

Table 6. Breakdown of what we can correct out of the GZH data, by sample. *updated from 3-8-16: Switching to full depth for all GOODS data. Shallow depth information in appendix.*

	Correction type	AEGIS	COSMOS	GEMS	GOODS-N	GOODS-S	SDSS	Total
Correctable	0	1,654	15,170	1,837	993	835	0	20,489
Uncorrectable	1	1,917	26,113	2,423	1,385	1,282	0	33,120
No Correction Needed ($z \leq 0.3$)	2	955	11,926	1,175	415	400	37,545	52,416
NEI	3	2,847	34,511	3,308	2,535	2,523	0	45,724
No Redshift Information	4	1,134	5,088	561	687	102	14,316	21,888
Total		8,507	92,808	9,304	6,015	5,142	51,861	173,637

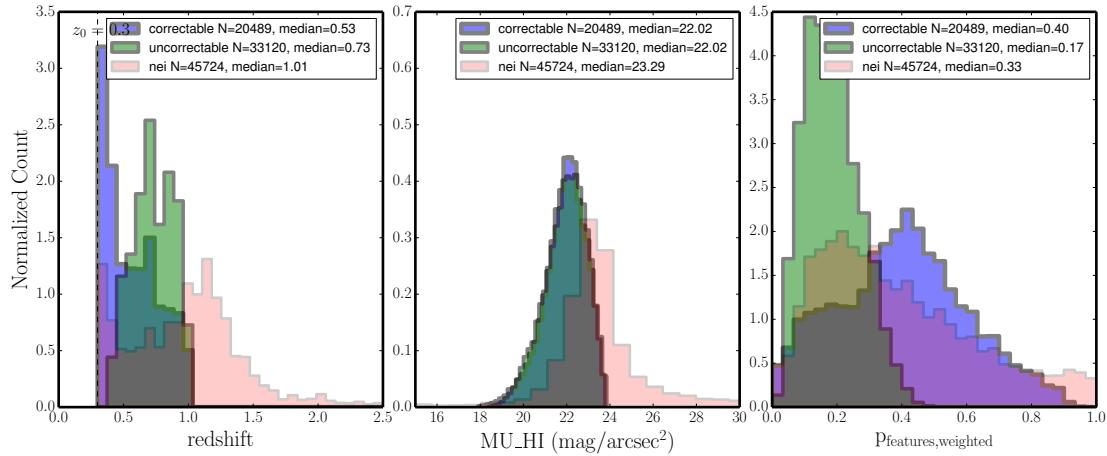


Figure 9. Distributions of redshift, surface brightness, and $p_{features}$ for correctable (purple), uncorrectable (green), and NEI (pink) galaxies in the full GZH sample. The uncorrectable galaxies tend towards higher redshift, slightly lower in surface brightness, and lower values of $p_{features}$ 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 $\mu > 24$.

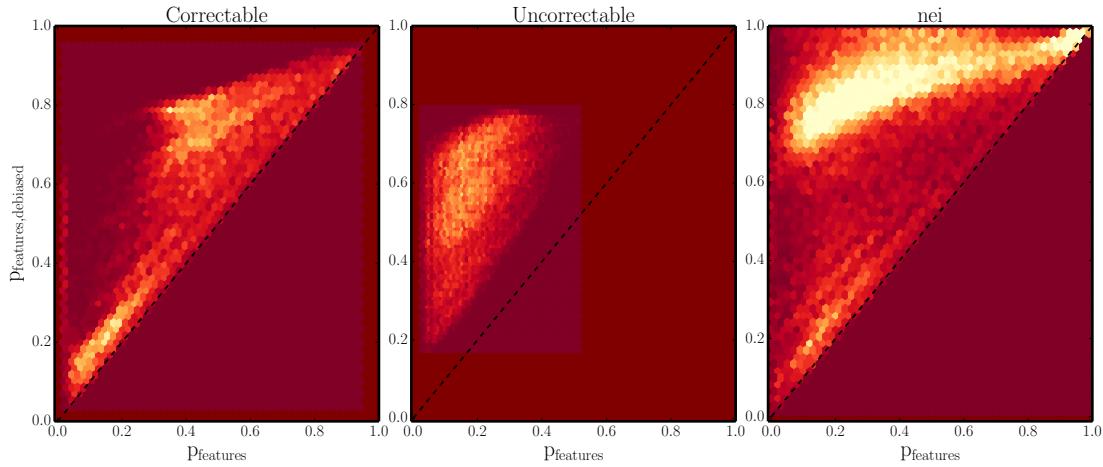


Figure 10. Debiased $p_{features}$ corrected to $z = 0.3$ vs weighted $p_{features}$ for the correctable (left), uncorrectable (middle), and NEI (right) galaxies in the GZH sample.

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 criteria that vote fractions for all answers must sum to unity. Explicitely:

$$p_{\text{smooth}} = 1 - p_{\text{features}} - p_{\text{artifact}}$$

We split the data products for GZH by the type of image being classified. Table 7 contains the classifications for the Hubble images from the AEGIS, COSMOS, and GEMS surveys, as well as 5-epoch deep imaging from the GOODS-N and GOODS-S surveys. This contains 118,425 galaxies and is the primary output from the GZH project. The next two tables have data for a small subset of 3,927 COSMOS images that were re-processed to study the effect of color balance on morphological classification. Table 8 has images that are desaturated to minimize the color contrast; Table 9 has images with the red and blue color channels inverted. Table 10 contains data for 6,144 galaxies with 2-epoch images from GOODS. These have been mostly supplanted in the main table with deeper 5-epoch GOODS imaging; however, there are 1,683 galaxies in the shallower imaging that were not classified in the deeper mosaics. This data can also be compared to the counterparts in Table 7 to study the effect of depth on morphological classification. Tables 11 and 12 contain data for the SDSS Stripe 82 single-depth and co-added images, respectively, that were classified using the GZH interface and decision tree. Finally, Table 13 contains classifications for images with artificial point sources intended to simulate the effect of a bright AGN, as used in Simmons et al. (2014).

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, à la GZ2.

6 USING THE CATALOG

Include cookbook for selecting morphologies.

Table 7. GZH morphological classifications for Hubble images from AEGIS, COSMOS, GEMS, and GOODS

Project ID	Hubble ID	Imaging	t01_smooth_or_features_		t01_smooth.or_features-all_smooth	best	lower limit	upper limit	...
			Correction ¹	N_votes					
AHZ100002g	10010842	AEGIS	0	127	0.118	0.128	0.085	0.226	0.226
AHZ100002h	10010870	AEGIS	4	127	0.567	0.592	0.592	—	—
...									
AHZ200041kd	20014731	COSMOS	3	44	0.682	0.675	0.147	0.675	—
AHZ200041ke	20014732	COSMOS	2	45	0.689	0.756	0.893	0.756	—
...									
AHZ400043g	90022729	GEMS	1	121	0.702	0.733	0.487	0.734	0.483
AHZ4000416	90022735	GEMS	1	127	0.646	0.698	0.508	0.698	0.171
...									
AGZ0007z47	10014	GOODS-N-FULLDEPTH	1	40	0.475	0.475	0.197	0.475	0.011
AGZ0007z48	10017	GOODS-N-FULLDEPTH	3	40	0.675	0.675	0.048	0.675	0.168
...									
AGZ00083jb	8869	GOODS-S-FULLDEPTH	1	40	0.425	0.425	0.109	0.425	0.070
AGZ00083jc	8878	GOODS-S-FULLDEPTH	0	40	0.205	0.205	0.048	0.048	-0.005
...									

¹Flag indicating how the vote fractions for this galaxy were corrected through debiasing (§4.3), if possible. 0 = correctable, 1 = uncorrectable ($p_{\text{row}} - p_{\text{adj}}$ is not single-valued), 2 = uncorrected ($z_{\text{gal}} < 0.3$), 3 = uncorrectable (insufficient FERENGI galaxies in this $z-\mu$ bin), 4 = uncorrectable (no galaxy redshift available).

Note. — The full version of this table is available in electronic form, as well as at <http://data.galaxyzoo.org>. The complete version includes data for 118,425 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 8. GZH morphological classifications for color-faded Hubble images

Project ID	Hubble ID	Imaging	t01_smooth_or_features_			t01_smooth_or_features_a01_smooth_			...
			Correction	N_votes	fraction	weighted	debiased	best	
AHZF000001	20000002	COSMOS	1	48	0.708	0.755	0.228	0.754	0.325
AHZF000003	20000004	COSMOS	3	49	0.367	0.379	0.100	0.379	0.198
AHZF000004	20000006	COSMOS	3	49	0.265	0.271	0.010	0.270	—
AHZF00000z	20000102	COSMOS	1	44	0.727	0.78	0.233	0.780	0.820
AHZF000010	20000104	COSMOS	2	53	0.811	0.849	0.904	0.848	—
...									

Note. — The full version of this table is available in electronic form, as well as at <http://data.galaxyzoo.org>. The complete version includes data for 3,927 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 9. GZH morphological classifications for color-inverted Hubble images

Project ID	Hubble ID	Imaging	t01_smooth_or_features_			t01_smooth_or_features_a01_smooth_			...
			Correction	N_votes	fraction	weighted	debiased	best	
AHZC000001	20000002	COSMOS	1	168	0.615	0.664	0.160	0.663	0.271
AHZC000003	20000004	COSMOS	0	235	0.333	0.364	0.002	0.002	0.063
AHZC000004	20000006	COSMOS	3	316	0.235	0.252	-0.011	0.252	—
AHZC00000z	20000102	COSMOS	1	207	0.755	0.757	0.272	0.756	0.796
AHZC000010	20000104	COSMOS	2	158	0.843	0.882	0.936	0.881	—
...									

Note. — The full version of this table is available in electronic form, as well as at <http://data.galaxyzoo.org>. The complete version includes data for 3,927 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 10. GZH morphological classifications for GOODS 2-epoch images

Project ID	Hubble ID	Imaging	<u>t01_smooth_or_features_</u>			<u>t01_smooth_or_features_a01_smooth_</u>			upper limit	...
			Correction	N_votes	fraction	weighted	debiased	best		
AHZ3000001	50000000	GOODS-N	0	123	0.390	0.415	0.090	0.090	—	
AHZ3000002	50000001	GOODS-N	2	126	0.341	0.355	0.356	0.356	0.220	0.279
AHZ3000003	50000005	GOODS-N	1	129	0.760	0.826	0.633	0.825	0.596	0.834
AHZ3000004	50000008	GOODS-N	1	120	0.758	0.787	0.639	0.787	0.658	0.834
AHZ3000005	50000010	GOODS-N	1	123	0.854	0.890	0.611	0.889	0.597	0.914
...										

Note. — The full version of this table is available in electronic form, as well as at <http://data.galaxyzoo.org>. The complete version includes data for 6,144 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 11. GZH morphological classifications for SDSS Stripe 82 single-epoch images

Project ID	SDSS DR7 ObjID	Imaging	<u>t01_smooth_or_features_</u>			<u>t01_smooth_or_features_a01_smooth_</u>			best	...
			Correction	N_votes	fraction	weighted	debiased	best		
AHZ5000001	587730845812064684	SDSS	2	41	0.585	0.595	0.759	0.759	0.594	
AHZ5000002	587730845812065247	SDSS	2	46	0.609	0.651	0.897	0.897	0.651	
AHZ5000003	587730845812196092	SDSS	2	51	0.039	0.044	0.067	0.067	0.043	
AHZ5000004	587730845812196825	SDSS	2	35	0.514	0.605	0.928	0.928	0.605	
AHZ5000005	587730845812524122	SDSS	2	47	0.766	0.812	1.038	1.038	0.810	
AHZ5000006	587730845812654984	SDSS	2	42	0.5	0.542	0.680	0.680	0.541	
AHZ5000007	587730845812655451	SDSS	2	41	0.488	0.526	0.697	0.697	0.525	
AHZ5000008	587730845812720365	SDSS	2	53	0.792	0.84	1.050	1.050	0.839	
AHZ5000009	587730845812720640	SDSS	4	43	0.0	0.0	0.0	0.0	0.0	
AHZ500000a	587730845812720699	SDSS	2	40	0.425	0.478	0.588	0.588	0.477	
...										

Note. — The full version of this table is available in electronic form, as well as at <http://data.galaxyzoo.org>. The complete version includes data for 21,572 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 12. GZH morphological classifications for SDSS Stripe 82 coadded images

Project ID	SDSS DR7 ObjID	Imaging	<u>t01_smooth_or_features_N_votes</u>		<u>t01_smooth_or_features_a01_smooth</u>		...
			Correction	fraction	weighted	debiased	
AHZ60000001	8647474690312306978	SDSS	4	40	0.275	0.289	0.702
AHZ60000002	8647474690312307154	SDSS	2	43	0.605	0.634	0.858
AHZ60000003	8647474690312307877	SDSS	2	51	0.608	0.627	0.635
AHZ60000004	8647474690312308301	SDSS	4	52	0.038	0.038	0.038
AHZ60000005	8647474690312308318	SDSS	2	44	0.614	0.632	0.723
AHZ60000006	8647474690312308880	SDSS	2	36	0.667	0.683	0.776
AHZ60000007	86474746903123072644	SDSS	4	48	0.646	0.674	0.674
AHZ60000008	86474746903123072789	SDSS	4	45	0.489	0.571	0.964
AHZ60000009	86474746903123072931	SDSS	4	47	0.553	0.587	0.926
AHZ600000a	86474746903123073190	SDSS	4	47	0.574	0.559	0.559
...							

Note. — The full version of this table is available in electronic form, as well as at <http://data.galaxyzoo.org>. The complete version includes data for 30,339 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 13. GZH morphological classifications for HST images with simulated AGN

Project ID	SDSS DR7 ObjID	Imaging	Correction	Version	L_{ratio}	AGN color ¹	N_votes	fraction	<u>t01_smooth_or_features_a01_smooth</u>			upper limit
									weighted	debiased	best	
AHZ70000001	90024700	GEMS	1	1	0.2	1	42	0.238	0.239	-0.110	0.238	-0.113
AHZ70000002	90024700	GEMS	1	1	1.0	1	51	0.255	0.265	-0.107	0.264	-0.128
AHZ70000003	90024700	GEMS	0	1	5.0	1	47	0.170	0.167	-0.018	-0.049	0.033
AHZ70000004	90024700	GEMS	0	1	10.0	1	41	0.195	0.195	0.045	0.045	0.044
AHZ70000005	90024700	GEMS	0	1	50.0	1	47	0.170	0.178	0.067	0.067	0.127
...												0.167
AHZ7000013m	90024700	GEMS	0	2	0.0	0	35	0.171	0.136	0.011	0.029	0.112
AHZ7000013n	90024700	GEMS	1	2	0.2	1	20	0.150	0.158	-0.278	0.049	-0.351
AHZ7000013o	90024700	GEMS	1	2	1.0	1	32	0.281	0.300	-0.086	0.281	-0.119
AHZ7000013p	90024700	GEMS	0	2	5.0	1	29	0.103	0.115	-0.098	-0.098	-0.152
AHZ7000013q	90024700	GEMS	0	2	10.0	1	35	0.171	0.181	0.027	0.027	0.023
AHZ7000013r	90024700	GEMS	0	2	50.0	1	34	0.206	0.206	-0.005	-0.005	-0.056
...												0.026

¹Flag indicating the color of the PSF in the simulated AGN. 0 = no simulated AGN, 1 = blue, 2 = flat, 3 = red.Note. — The full version of this table is available in electronic form, as well as at <http://data.galaxyzoo.org>. The complete version includes data for 2,961 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

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.

ACKNOWLEDGEMENTS. 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://www.galaxyzoo.org/volunteers>.

We thank Meg Schwamb and the ASIAA for hosting the “Citizen Science in Astronomy” workshop, 3–7 Mar 2014 in Taipei, Taiwan, at which some of this analysis was done.

This project made heavy use of the Astropy packages in Python (Astropy Collaboration et al. 2013), the `seaborn` plotting package (Waskom et al. 2015), and astroML (Vanderplas et al. 2012).

HST acknowledgements.

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

- Astropy Collaboration et al., 2013, *A&A*, 558, A33
 Baillard A. et al., 2011, *A&A*, 532, A74
 Bamford S. P. et al., 2009, *MNRAS*, 393, 1324
 Barden M., Jahnke K., Häußler B., 2008, *ApJS*, 175, 105
 Beckwith S. V. W. et al., 2006, *AJ*, 132, 1729
 Caldwell J. A. R. et al., 2008, *ApJS*, 174, 136
 Cameron E., Carollo C. M., Oesch P. A., Bouwens R. J., Illingworth G. D., Trenti M., Labb   I., Magee D., 2011, *ApJ*, 743, 146
 Cardamone C. N. et al., 2010, *ApJS*, 189, 270
 Chevance M., Weijmans A.-M., Damjanov I., Abraham R. G., Simard L., van den Bergh S., Caris E., Glazebrook K., 2012, *ApJ*, 754, L24
 Conselice C. J., 2003, *ApJS*, 147, 1
 Conselice C. J., 2014, ArXiv e-prints, 1403.2783
 Darg D. W. et al., 2010, *MNRAS*, 401, 1552
 Davis M. et al., 2007, *ApJ*, 660, L1
 de Vaucouleurs G., 1959, *Handbuch der Physik*, 53, 275
 Dressler A., 1980, *ApJ*, 236, 351
 Elmegreen D. M., Elmegreen B. G., Ferguson T., Mullan B., 2007, *ApJ*, 663, 734
 Elmegreen D. M., Elmegreen B. G., Rubin D. S., Schaffer M. A., 2005, *ApJ*, 631, 85
 F  rster Schreiber N. M., Shapley A. E., Erb D. K., Genzel R., Steidel C. C., Bouch   N., Cresci G., Davies R., 2011, *ApJ*, 731, 65
 Freeman P. E., Izbicki R., Lee A. B., Newman J. A., Conselice C. J., Koekemoer A. M., Lotz J. M., Mozena M., 2013, *MNRAS*, 434, 282
 Genel S. et al., 2014, *MNRAS*, 445, 175
 Giavalisco M. et al., 2004, *ApJ*, 600, L93
 Griffith R. L. et al., 2012, *ApJS*, 200, 9
 Grogin N. A. et al., 2011, *ApJS*, 197, 35
 Hinshaw G. et al., 2013, *ApJS*, 208, 19
 Hopkins P. F. et al., 2010, *ApJ*, 715, 202
 Hubble E. P., 1926, *ApJ*, 64, 321
 Hubble E. P., 1936, *Realm of the Nebulae*. Yale University Press
 Ilbert O. et al., 2013, *A&A*, 556, A55
 Johnson L. C. et al., 2015, *ApJ*, 802, 127
 Kartaltepe J. S. et al., 2015, *ApJS*, 221, 11
 Lackner C. N., Gunn J. E., 2012, *MNRAS*, 421, 2277
 Land K. et al., 2008, *MNRAS*, 388, 1686
 Law D. R., Shapley A. E., Steidel C. C., Reddy N. A., Christensen C. R., Erb D. K., 2012a, *Nature*, 487, 338
 Law D. R., Steidel C. C., Shapley A. E., Nagy S. R., Reddy N. A., Erb D. K., 2012b, *ApJ*, 745, 85
 Lintott C. et al., 2011, *MNRAS*, 410, 166
 Lintott C. J. et al., 2008, *MNRAS*, 389, 1179
 Lotz J. M., Primack J., Madau P., 2004, *AJ*, 128, 163
 Masters K. L. et al., 2011, *MNRAS*, 411, 2026

- Momcheva I. G. et al., 2015, ArXiv e-prints, 1510.02106
 Mortlock A. et al., 2013, MNRAS, 433, 1185
 Nair P. B., Abraham R. G., 2010, ApJS, 186, 427
 Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
 Rix H.-W. et al., 2004, ApJS, 152, 163
 Sandage A., 1961, The Hubble atlas of galaxies. Carnegie Institute of Washington
 Scarlata C. et al., 2007, ApJS, 172, 406
 Schawinski K. et al., 2014, MNRAS, 440, 889
 Schaye J. et al., 2015, MNRAS, 446, 521
 Scoville N. et al., 2007, ApJS, 172, 1
 Silk J., Mamon G. A., 2012, Research in Astronomy and Astrophysics, 12, 917
 Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11
 Simmons B. D. et al., 2013, MNRAS, 429, 2199
 Simmons B. D. et al., 2014, MNRAS, 445, 3466
 Skibba R. A. et al., 2012, MNRAS, 423, 1485
 Strauss M. A. et al., 2002, AJ, 124, 1810
 Toomre A., Toomre J., 1972, ApJ, 178, 623
 van den Bergh S., 1976, ApJ, 206, 883
 Vanderplas J., Connolly A., Ivezić Ž., Gray A., 2012, in Conference on Intelligent Data Understanding (CIDU), pp. 47–54
 Vogelsberger M. et al., 2014, MNRAS, 444, 1518
 Waskom M. et al., 2015, seaborn: v0.6.0 (june 2015)
 Willett K. W. et al., 2013, MNRAS, 435, 2835
 Willett K. W. et al., 2015, MNRAS, 449, 820
 Williams R. E. et al., 1996, AJ, 112, 1335
 Wright E. L., 2006, PASP, 118, 1711
 York D. G. et al., 2000, AJ, 120, 1579

Table A1. Breakdown of what we can correct out of the GOODS shallow depth data.

	GOODS-N	GOODS-S	Total
Correctable	748	514	1,262
Uncorrectable	526	1,143	1,669
No Correction Needed ($z \leq 0.3$)	267	267	534
NEI	851	2,670	3,521
No Redshift Information	159	319	478
Total	2,551	4,913	7,464

APPENDIX A: GOODS SHALLOW DEPTH DATA

GZH used both 5-epoch and 2-epoch sets of data to construct the GOODS set of images. The 11,157 full depth 5-epoch images are used in the main catalog; the classifications for the 7,464 shallow depth 2-epoch images are offered as a supplementary table. Here we briefly analyze the effect of image depth on the ability of the GZ users to identify features or disk structure in the images.

A1 Comparing shallow and full depth morphologies

Of the 11,157 galaxies in the GOODS-N and GOODS-S full depth sample, 4,461 of these are in the shallow-depth sample. In Figure A1 we find a strong correlation between p_{features} for both sets of images. The mean change in p_{features} from the shallow to full depth images $p_{\text{features,full}} - p_{\text{features,shallow}} \equiv \Delta p = 0.00$, with a standard deviation of $\sigma = 0.17$. While there is some variance in Δp in the whole sample, the change is usually small and not often significant enough to change a morphological classification. Defining a clean sample of disk galaxies as those with $p_{\text{features,best}} > 0.8$, elliptical galaxies as those with $p_{\text{smooth,best}} < 0.2$, and intermediate as those in between, we find that 75% of the sample would not change morphology. Of the remaining 25% that would change morphology, only 0.3% (representing 10 galaxies total) drastically change morphology from smooth to featured or visa versa, while the rest would transition to or from the “intermediate” morphology. Details can be seen in Table A2 and examples of images representing the 9 possible changes (or lack of) in morphology are shown in Figures A2, A3, and A4.

A2 FERENGI analysis of p_{bar}

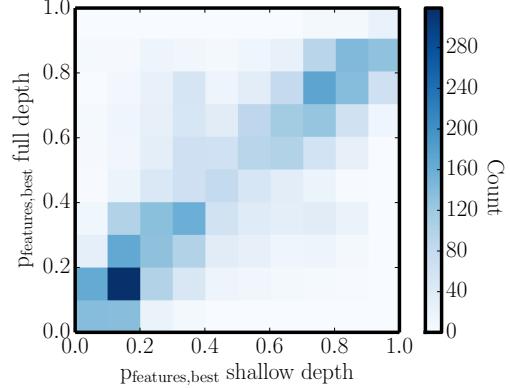


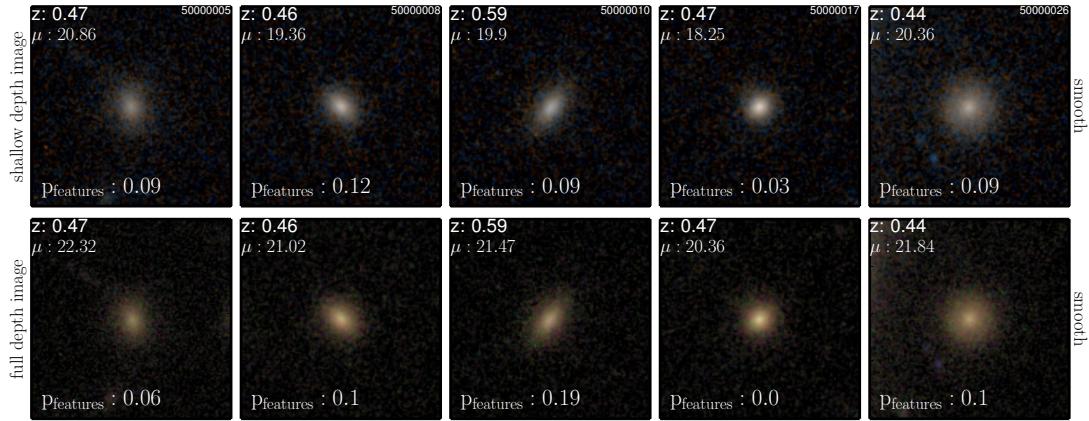
Figure A1. shallowfull

Table A2. Properties of galaxies whose morphologies changed or stayed the same in the shallow vs full images. Featured here is defined as $p_{\text{features,best}} > 0.8$, intermediate = $0.2 < p_{\text{features,best}} < 0.8$, smooth = $p_{\text{smooth,best}} < 0.2$.

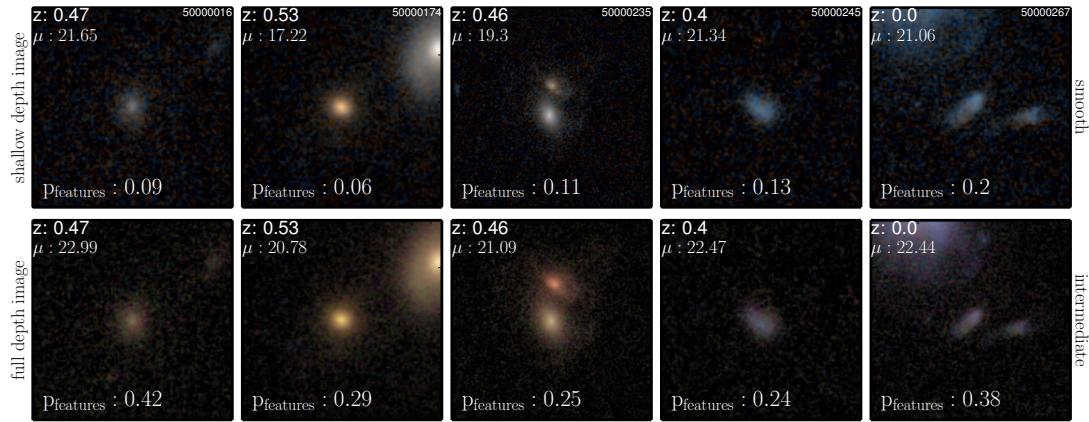
shallow to full morphology	N	%	$< \Delta p >$	$< z >$
smooth to smooth	758	17.0	-0.00	0.69
smooth to intermediate	367	8.2	0.18	0.69
smooth to featured	7	0.2	0.76	0.57
intermediate to smooth	214	4.8	-0.18	0.65
intermediate to intermediate	2,303	51.6	0.01	0.78
intermediate to featured	168	3.8	0.19	0.83
featured to smooth	3	0.1	-0.74	0.71
featured to intermediate	337	7.6	-0.18	0.68
featured to featured	301	6.8	-0.05	0.71
Total	4,461	100		

Table A3. Distribution of FERENGI images analysed in Figure A5. Correctable images had a single-valued relationship between their measured p_{bar} values at high and low redshifts (white regions in Figure A5). 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 $z-\mu$ bins (gray regions). Only 17% (maximum) of FERENGI galaxies in the sample were considered “correctable”, which is not sufficient to compute a ζ function applicable to the Hubble data.

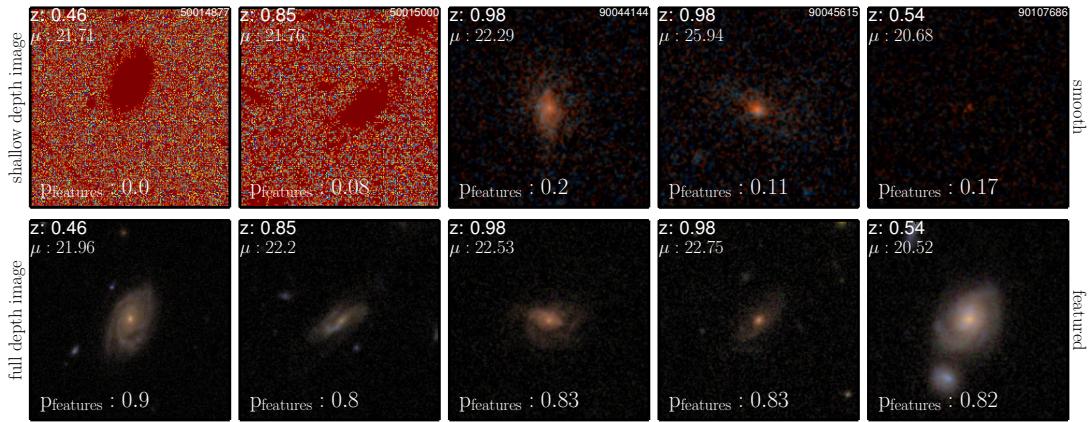
	N	%
Correctable	664	17%
Uncorrectable	483	12%
NEI	2,803	71%
Total	3,950	100%



[a]

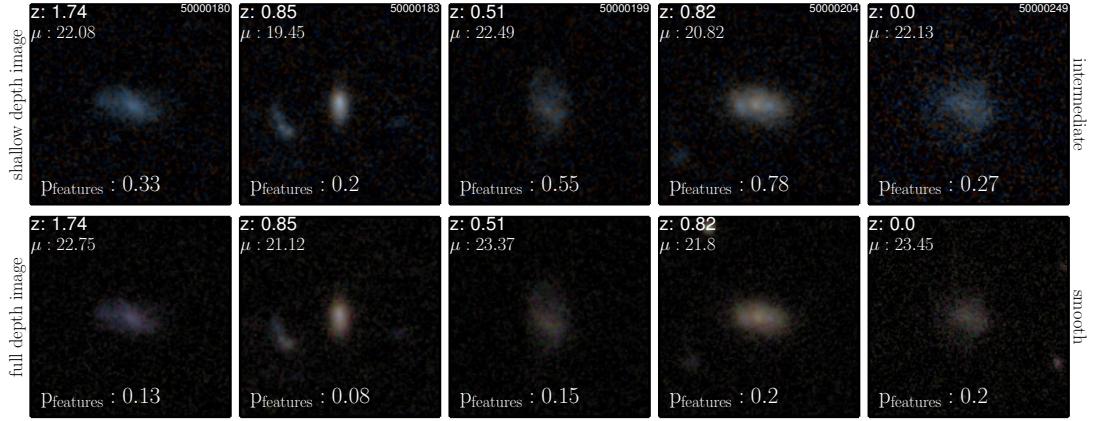


[b]

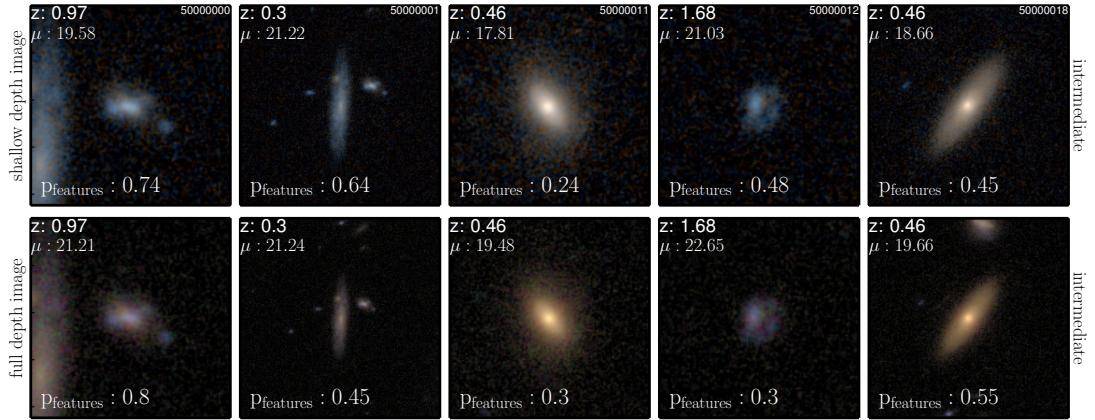


[c]

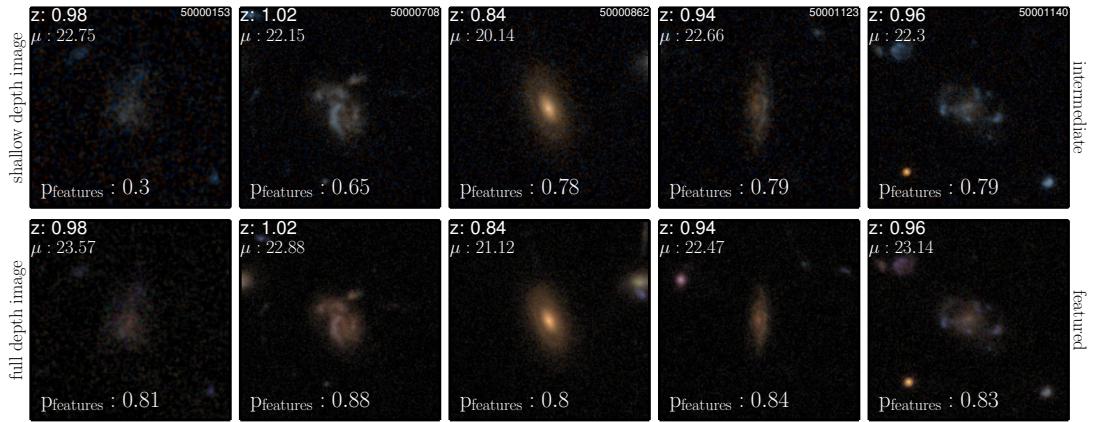
Figure A2. Galaxies whose shallow images were classified as smooth and full depth images were classified as smooth, intermediate, or featured.



[b]



[b]



[b]

Figure A3. Galaxies whose shallow images were classified as intermediate and full depth images were classified as smooth, intermediate, or featured.

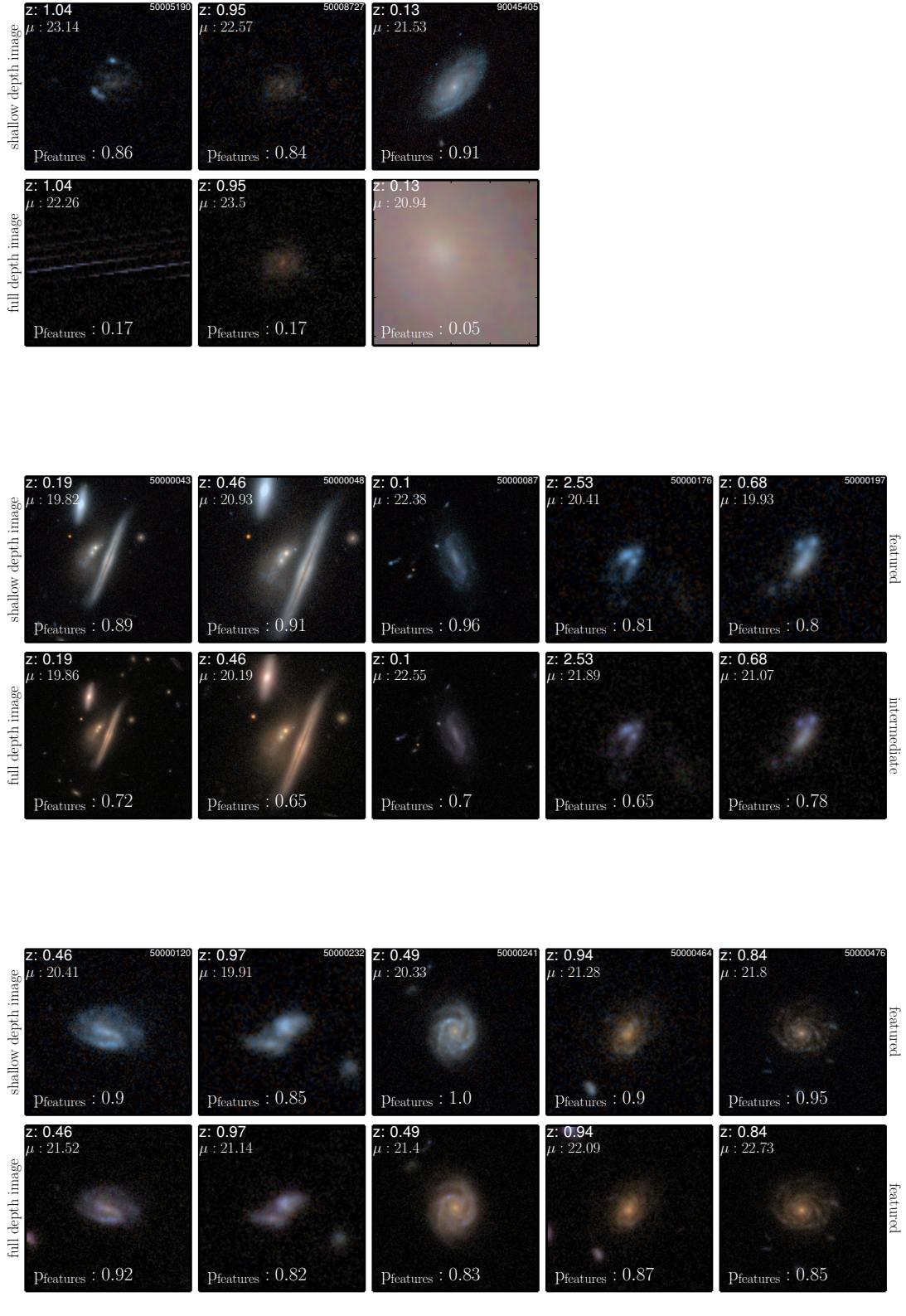


Figure A4. Galaxies whose shallow images were classified as featured and full depth images were classified as smooth, intermediate, or featured.

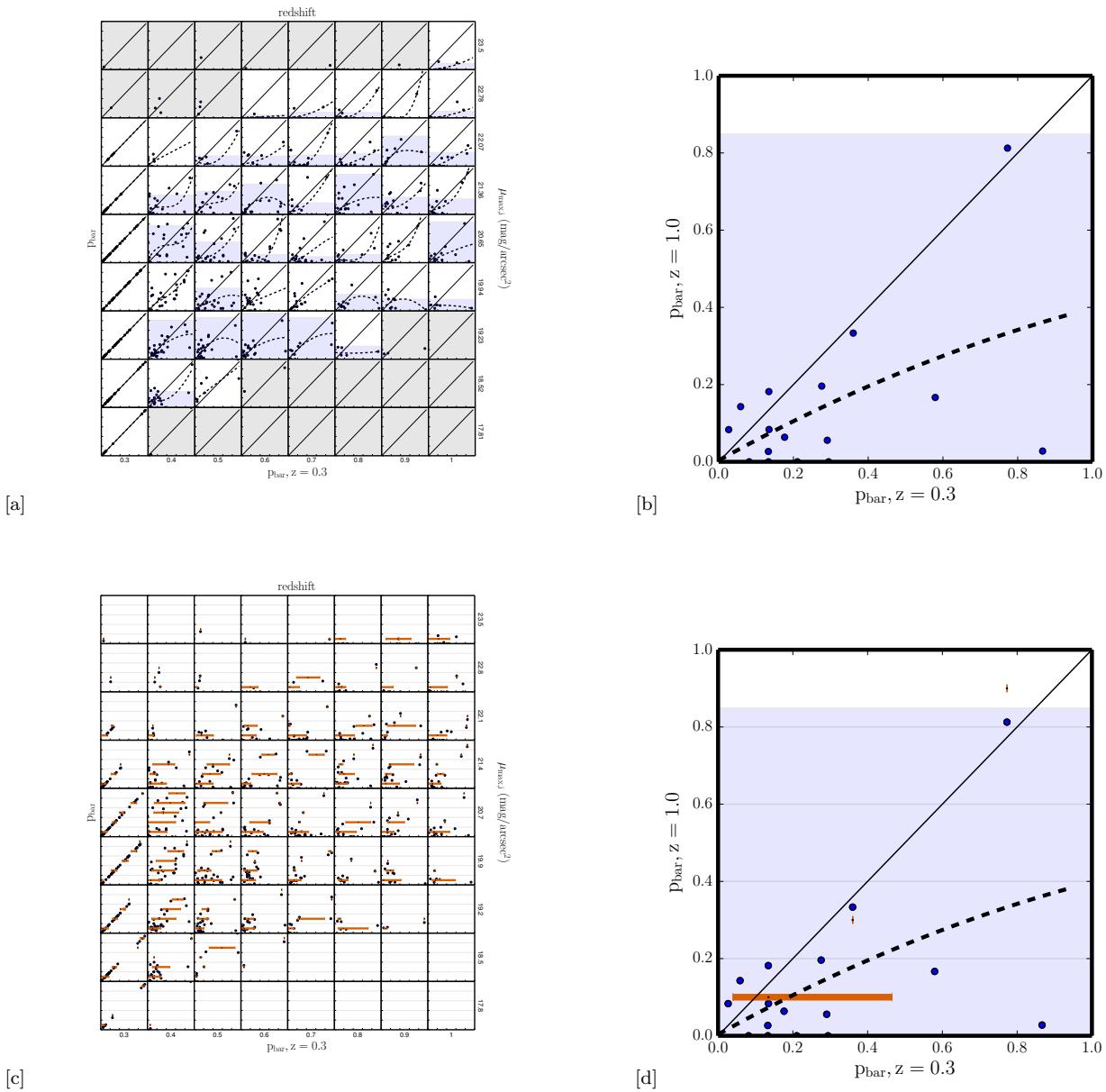


Figure A5. Same as Figure 8, but with the bar question.