Draft version April 27, 2016
Preprint typeset using LaTeX style AASTeX6 v. 1.0

GALAXY ZOO: MORPHOLOGICAL CLASSIFICATIONS FOR GALAXIES IN HST LEGACY IMAGING

Kyle W. Willett[1,2], Melanie A. Galloway[1], Chris J. Lintott[3], B.D. Simmons[3,4,5,6], Karen L. Masters[7,8], Tom Melvin[7], Steven P. Bamford[9], Edward M. Edmondson[7], Lucy F. Fortson[1], Kevin Schawinski[10], Edmond Cheung[11], Claudia Scarlata[1], Melanie Beck[1], Arfon M. Smith[3,12,13], Michael Parrish[12], Roger L. Griffith[14,15], Anna Han[16]

ABSTRACT

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

Keywords: galaxies — catalogs — surveys

1. INTRODUCTION

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 $M^{\star}$ 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 ΛCDM 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 Mpc^3 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 individial 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 constrasts 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;

[a] 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.

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, Williams et al. 1996; Giavalisco et al. 2004; Beckwith et al. 2006; Davis et al. 2007; Scoville et al. 2007; Grogin et al. 2011) give the photometric sensitivity necessary to detect $L^\star$ 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 ΛCDM 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; Baillard 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; Bamford et al. 2008; 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 crowdsourced 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 *HST* Legacy Surveys.

2.1. *Summary of Hubble Legacy Survey Imaging*

- *Hubble Space Telescope* 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 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

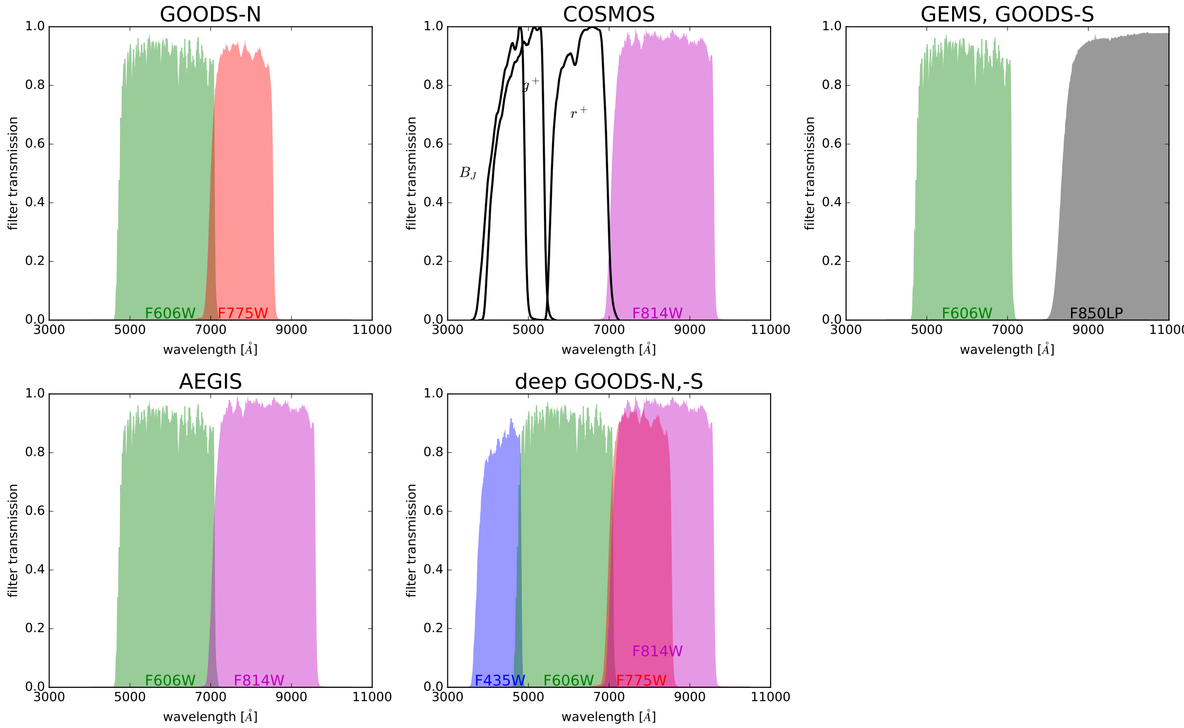


Figure 1. Transmission curves of the filters used by *HST* Advanced Camera for Surveys (ACS) in wide-field channel mode for the various surveys in Galaxy Zoo: Hubble. The unfilled black curves show the filters for the Suprime Camera on the *Subaru* telescope which were used to create color gradients in the composite images for COSMOS.

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 $\sim$ 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 $\sim$ 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 *HST* 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 $\sim$ 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. *Galaxy selection*

Table 1. Summary of Galaxy Zoo: Hubble imaging

Survey	$t_{\rm exp}$ [sec]	Filters	Resolution [$''$/pix]	Area [arcmin^2]	$N_{\rm galaxies}$
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

For images from the ACS-GC Griffith et al. (2012), galaxies are identified using a combination of SExtractor (Bertin & Arnouts 1996) and the galaxy-profile fitting framework Galapagos (Barden et al. 2012). We selected all galaxies with $m < 23.5$, where m is in the I_{814W}, I_{850LP}, or I_{775W} for the AEGIS + COSMOS, GEMS + GOODS-S, and GOODS-N surveys, respectively. This yielded a total of 113,166 images (Table 1).

Single-epoch images from SDSS Stripe 82 were selected using the same criteria from Willett et al. (2013), which required limits of `petroR90_r`$> 3''$ (where `petroR90_r` is the radius containing 90% of the r' Petrosian flux) and a magnitude brighter than $m_r < 17.77$. 21,522 galaxies in SDSS met these criteria. Coadded images from Stripe 82 were selected from the union of galaxies with coadded magnitudes brighter than 17.77 mag and the galaxies detected in the single-depth images and matched to a coadd source. This resulted in a larger set of 30,339 images. Of the images in the coadded sample, 5144 (17 percent) are dimmer than the initial magnitude cut of 17.77.

2.3. *Image creation*

The images used for classification in GZH are color-composite JPGs made from multi-band data. The exact process for creating the images depends on the bands and resolutions in the surveys.

For galaxies in the ACS General Catalogue (AEGIS, COSMOS, GEMS, 2-epoch GOODS; Griffith et al. 2012), the color composites are made using a fixed pixel intensity scaling with weights of $[2.4\times10^{-4}, 3\times10^{-4}, 3\times10^{-4}]$ in the red, green, and blue channels respectively. We also apply a nonlinear mapping to each pixel to emphasize the contrast in faint features (Lupton et al. 2004), taking the form of:

$$I_{\rm channel,new} = I_{\rm channel,old} \times \frac{\mathrm{arcsinh}(b\times r)}{(b\times r)} \qquad (1)$$

We adopt a value of $b = 3$ for the *HST* images in Griffith et al. (2012).

Since the majority of the data in Legacy surveys described in §2.1 had images in only 1 or 2 bands available when GZH was launched, making standard 3-channel RGB images in the same method used for the SDSS images in the original Galaxy Zoo was not possible. For surveys with images in two bands (AEGIS, GEMS, and the two-epoch GOODS-N and GOODS-S), the lower-wavelength band is mapped to the blue channel, the higher-wavelength band to the red channel, and the green channel created by taking the geometric mean of the two. Bands used in each of the surveys are listed in Table 1. Note that the 2-epoch GOODS-N and GOODS-S images use different filters — this was a deliberate choice made so that the GEMS images could be directly compared with the overlapping coverage of GOODS-S (Figure 1).

By the time that the 5-epoch sets of GOODS data were put into GZH, coverage in four separate *HST* bands had become publicly available. The deeper GOODS images were created using the arithmetic mean of B_{435W} and V_{606W} in the blue channel, I_{814W} in the red channel, and I_{850LP} in the green channel. These images also had the speckled noise pixels decolorized.

The COSMOS images only had the I_{814W} imaging available at the time of classification in GZH. For these galaxies, we created "pseudocolor" images by using the ACS I_{814W} data as the illumination map and ground-based imaging from the *Subaru* telescope in B_J, r^+, and i^+ filters to provide the color gradients (see Griffith et al. 2012, for further details). This means that the images have the angular resolution of *HST* ($\sim$ 0.05 $''$/pixel) for the overall intensity, but the color gradients are at ground-based resolution, with seeing between $0''.95$ and $1''.05$ (Taniguchi et al. 2007).

The Stripe 82 single-epoch images were taken directly from the color composites on the DR7 SDSS Skyserver, which combines g', r', and i' exposures into the RGB channels. The coadded Stripe 82 images were assembled from runs (106,206) in DR7 and generated into JPGs using the method of Lupton et al. (2004).

To address a number of images in which the sky noise was highly colored (which might have been a distraction to users), we apply a soft-edged object mask to the image that preserves the color balance for galaxies, but desaturates the speckled noise against regions of blank sky. This technique was applied only to the coadded Stripe 82 and the COSMOS images.

2.4. *Artificial AGN*

We also created a set of images designed to measure the effect of AGN on morphological classifications. Since galactic nuclei can have bright, unresolved optical emission, this has the potential to mimic the effect of a strong bulge. The presence of an AGN is simulated by modeling the PSF for *HST* and then inserting a bright source near the center of a real galaxy. For each image, the simulated AGN is assigned one of three colors – either blue, red, or flat (white) as seen in the color images – and a range of brightnesses such that $L_{\rm ratio} \equiv L_{\rm galaxy}/L_{\rm AGN}$ 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.

Two sets of simulated AGN were generated 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 in GOODS-S; this version used deeper imaging and improved PSFs from `TinyTim`.

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.

2.5. *Galaxy metadata*

Photometric data for the bulk of the GZH sample is largely drawn from the tables provided in Griffith et al. (2012). This includes photometric parameters from both SExtractor and GALFIT, such as the fluxes, magnitudes, radii, ellipticities, position angles, and positions. GALFIT also provides the parametric Sérsic index and effective half-light radius for the best-fit model. All parameters are measured in both bands of the ACS imaging, with the exception of the single-band COSMOS images.

Redshifts for the GZH catalog are compiled 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 (`ACSGC` (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 ACS-GC, 3DHST, and MUSYC catalogs. If a high quality spec-z exists in the ACS-GC. 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 spectroscipic redshifts are available, we compare the 1-σ errors of the photometric (ACS-GC, 3DHST, MUSYC, UltraVISTA) and grism (UltraVISTA) redshifts, and use the redshift with the smallest error. Table 2 shows the results of this selection.

Table 2. GZH redshifts by survey

	ACS-GC		3DHST			MUSYC		UltraVISTA	Total	
Survey	spec-z	photo-z	spec-z	grism-z	photo-z	spec-z	photo-z	photo-z	with redshift	in survey
AEGIS	3,656	2,941	12	515	249	0	0	0	7,373	8,507
COSMOS	7,201	77,435	35	358	26	0	0	2,665	85,020	92,808
GEMS	387	628	6	99	40	279	7,304	0	8,743	9,304
GOODS-N	1,947	37	418	1,545	1,381	0	0	0	5,328	12,030
GOODS-S	1,080	4	327	1,348	281	816	1,184	0	5,040	10,284
SDSS	0	0	0	0	0	0	0	0	37,545	51,861
Total	14,271	81,045	798	3,865	1,977	1,095	8,488	2,665	114,204	184,794

Photometric and spectroscopic data for the SDSS Stripe 82 galaxies are taken from the CasJobs DR7 tables. This includes *ugriz* Petrosian magnitudes and fluxes, as well as the relative de Vaucouleurs and exponential fits from the model magnitudes. Redshifts used for the Sloan galaxies are all spectroscopic. 82.6% of galaxies in the single-depth images and 65.1% of galaxies in the coadd images have a measured DR7 spectroscopic

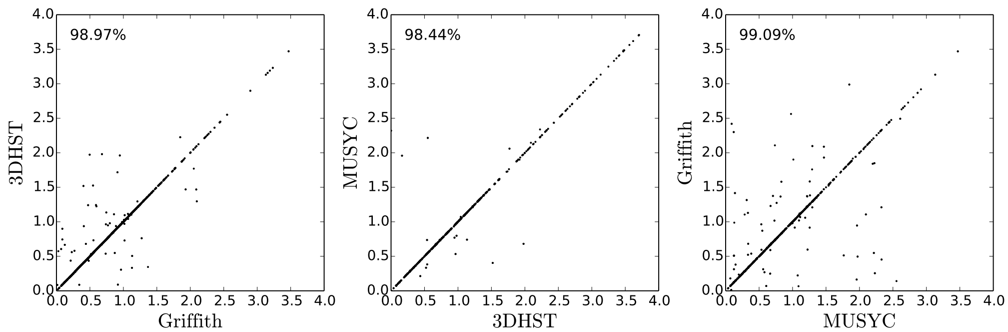


Figure 2. Spectroscopic redshifts from the ACS-GC, 3DHST, and MUSYC catalogs. The number in the upper left of each plot is the percentage of redshifts which agree within $\Delta z < 0.05$ between the two catalogs being compared in each panel. Within this range there is over 98% agreement in redshifts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

redshift.

3. GALAXY ZOO INTERFACE AND CLASSIFICATIONS

3.1. *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 user weighting w is 1 for the top 95% of users as ranked by consistency. For the bottom 5% of users, w drops smoothly and is effectively zero for only the bottom 1% of classifiers in the CDF. The overall effect on the GZH dataset is relatively small, but this method is effective at filtering out contributions from random or delibrately malicious classifiers.

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.

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,

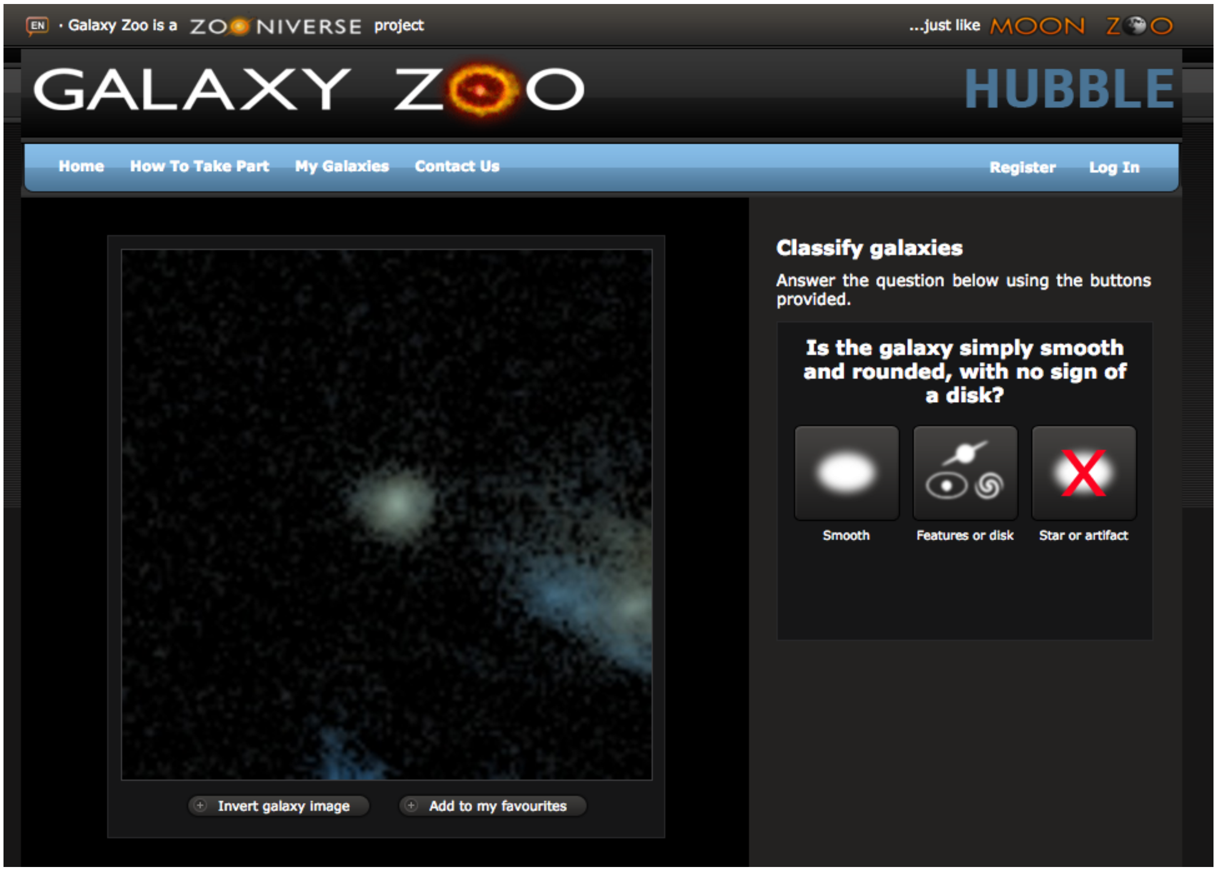


Figure 3. Screenshot of the Galaxy Zoo: Hubble interface showing an example COSMOS image at the first step in the decision tree.

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 5 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.

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:

$$f_{\mu,z} = (f_{\mu,z=0.3}) \times e^{\frac{z-z_0}{\zeta}}, \qquad (2)$$

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

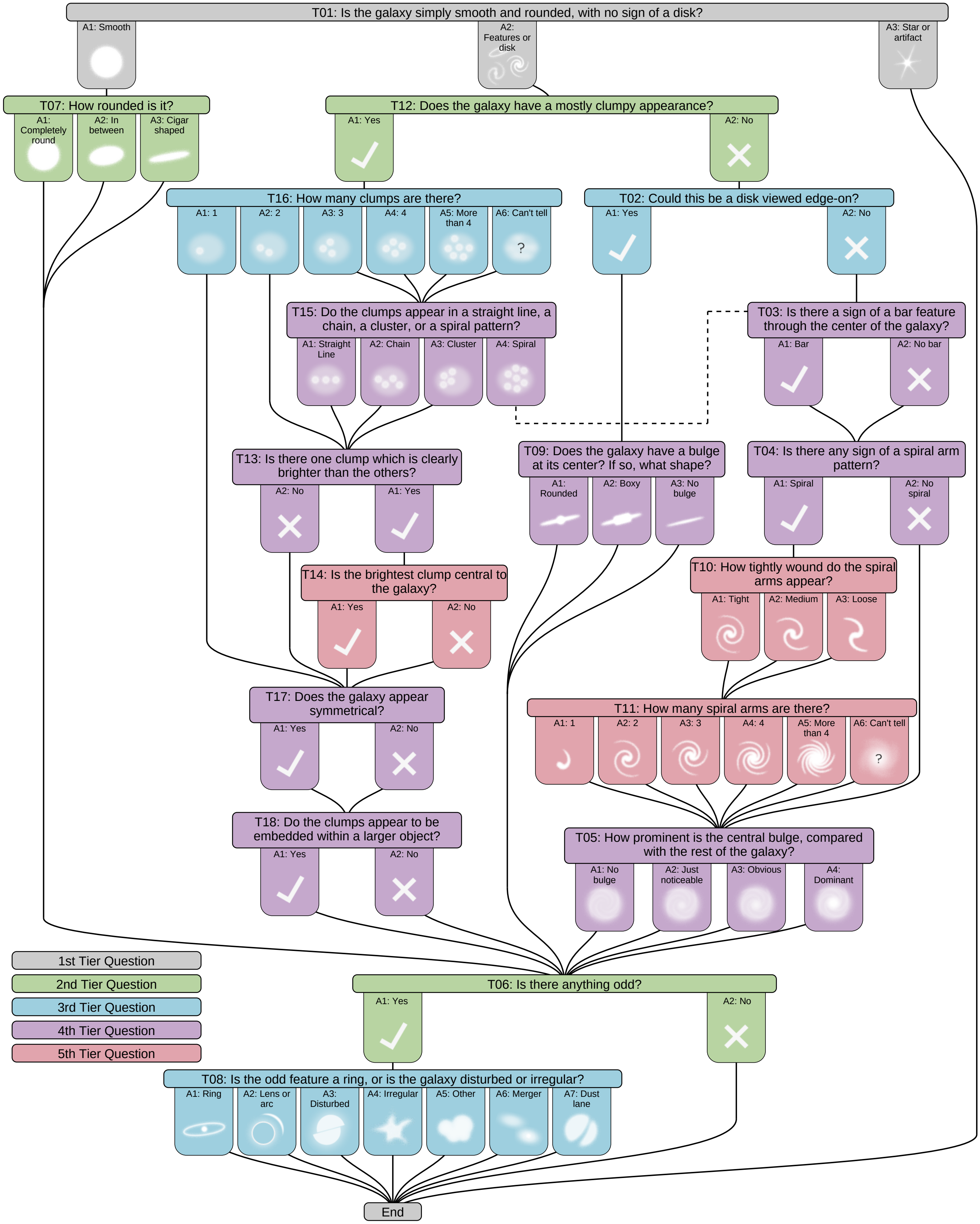


Figure 4. Flowchart of the questions presented to GZH users, labeled with the corresponding Task numbers. Tasks in the decision tree are color-coded by tier level: Gray-colored Tasks are 1st tier questions which are asked of all users. Tasks colored green, blue, purple, and pink (respectively) are one, two, three, or four steps below branching points 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

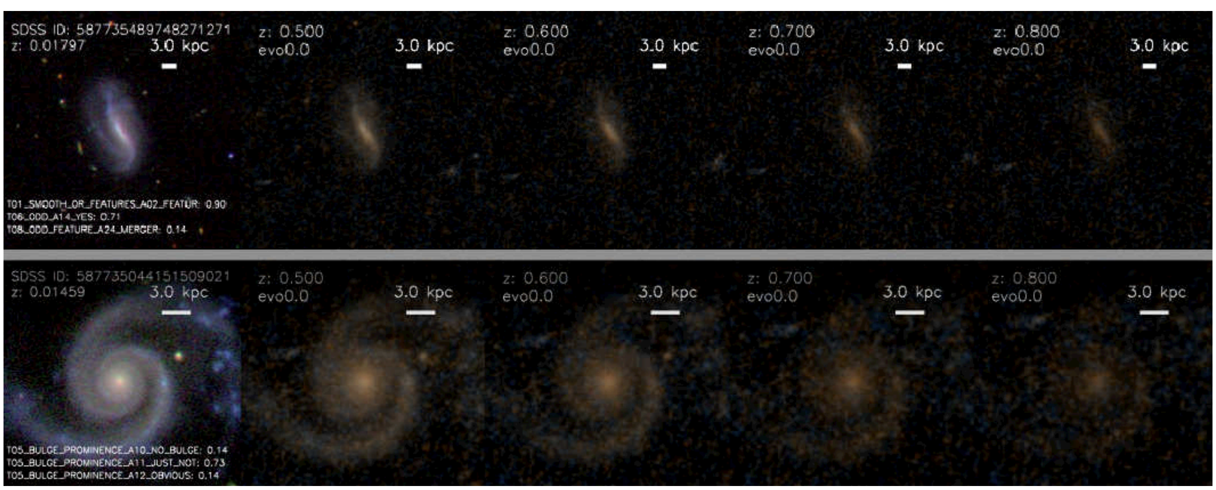


Figure 5. Examples of two galaxies which have been run through the FERENGI code to produce simulated *HST* images. The value of $p_{\rm features}$ for each panel is (1) Top row: $p_{\rm features}$ = 0.9, 0.625, 0.35, 0.35, 0.225 and (2) Bottom row: $p_{\rm features}$ = 1.00, 0.875, 0.875, 0.625, 0.375.

Table 4. Summary of FERENGI artificial redshifting

$z_{\rm target}$	$N_{z\rm bins}$	$N_{\rm evolution}$	$e_{\rm max}$	$N_{\rm galaxies}$	$N_{\rm images}$
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

zero. Figure 6 show examples of the change in vote fraction and their fits to Equation 2 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 $f_{\rm features}$ at $z = 0.3$, and with a reasonable fit to the exponential model ($\Delta\chi^2 > 3.0$).

Figure 7 shows the results of fitting the FERENGI images with Equation 2; 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 $f_{\rm features}$ 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 7 with a linear function such that:

$$\log_{10}(\hat{\zeta}) = \zeta_0 + (\zeta_1 \times \mu), \tag{3}$$

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 2 to bound the debiased vote fractions between f and 1:

$$f_{\rm features,debiased} = 1 - (1 - f)e^{\frac{z-z_0}{\hat{\zeta}}}. \tag{4}$$

4.3. *Results of ζ approach*

In Figure 9 we examine the change in $p_{\rm features}$ for the FERENGI galaxies relative to their lowest simulated redshift. In this analysis, only galaxies whose lowest simulated redshift image was ($z_{\rm sim} = 0.3$) 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 $p_{{\rm features},z}$, the value measured at that simulated redshift, vs $p_{{\rm features},z=0.3}$, 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 $p_{{\rm features},z}$ value, what its $p_{\rm features}$ value *would have been* if it had been viewed at $z = 0.3$. This predicted value is defined as the debiased vote fraction $p_{\rm features,debiased}$, and is calculated by applying a correction to the measured value of $p_{\rm features}$, determined by the ζ function described in the previous section. A reliable predicted value can be obtained so long as the relationship between $p_{{\rm features},z}$ and $p_{{\rm features},z=0.3}$ is single-valued; that is, for a given $p_{{\rm features},z}$, there is exactly one corresponding value of $p_{\rm features}$ at $z = 0.3$.

Figure 9 shows that the relationship between $p_{{\rm features},z}$ and $p_{{\rm features},z=0.3}$ is *not* always single valued; hence, it is not appropriate to correct galaxies that lie in certain regions of surface brightness/redshift/$p_{\rm features}$ space. These regions tend to have low $p_{\rm features}$ 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 $p_{\rm features}$ at both high and low redshift. Second are disks whose features become washed out at high redshift; hence their $p_{\rm features}$ 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 $p_{{\rm features},z}$ and $p_{{\rm features},z=0.3}$ 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 9(a). Any flat regions of the polynomial fits are areas in which there is not a clear single-valued relationship between $p_{{\rm features},z}$ and $p_{{\rm features},z=0.3}$; 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 9(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 $p_{{\rm features},z}$ and $p_{{\rm features},z=0.3}$ relationship, colored gray in Figure 9(a).

The unshaded regions in Figure 9(a) define discrete ranges of redshift, surface brightness, and $p_{\rm features}$ a

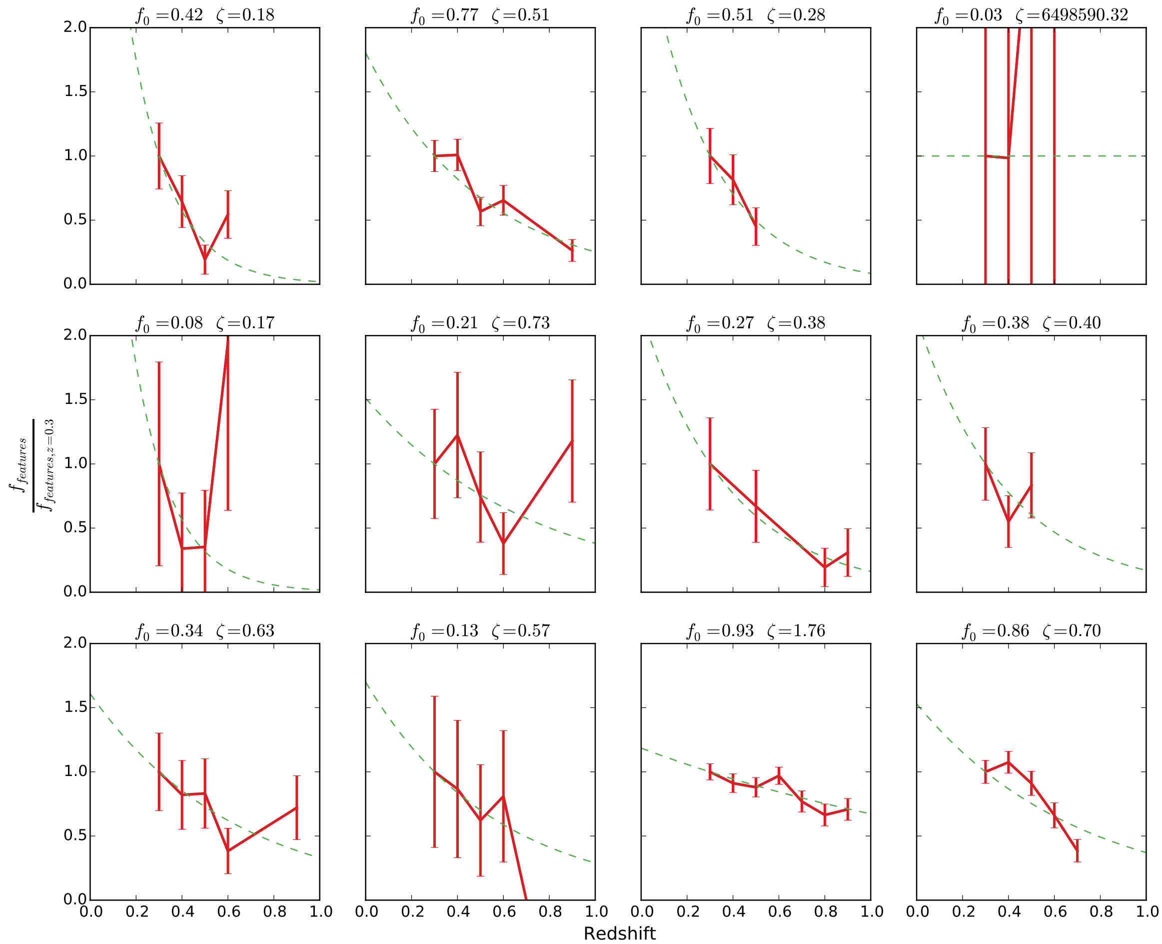


Figure 6. Behavior of the normalized, weighted vote fractions of features visible in a galaxy ($f_{\rm features}$) as a function of redshift in the artifical FERENGI images. Galaxies are a random selection of images with $e = 0$ and at least three detectable images in redshift bins of $z \geq 0.3$. The measured vote fractions (red points) are fit with an exponential function (Equation 2); the best-fit parameters are given above each plot. Error bars are Poissonian, assuming a median of 40 votes per galaxy.

galaxy must have in order for the ζ approach to be confidentely 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 $p_{\rm features}$. To define this smooth space, we generate a convex hull that encloses the correctable and uncorrectable FERENGI galaxies in z-μ-$p_{\rm features}$ space. 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 *HST* galaxies. The results of this method and final categorization of the *HST* 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 $p_{\rm features}$ value, we instead determine a likely *range* of debiased values, using a method visualized in Figure 9(c). Here we again use the FERENGI simulated data to analyze the range of intrinsic $p_{{\rm features},z=0.3}$ values for any given observed $p_{\rm features}$ value, again as a function of surface brightness and redshift. In each z,μ bin, we examine the spread of intrinsic values of $p_{{\rm features},z=0.3}$ for 4 ranges of observed $p_{\rm features}$. We quantify the range of intrinsic values as the inner 80% of the data; this range is repre-

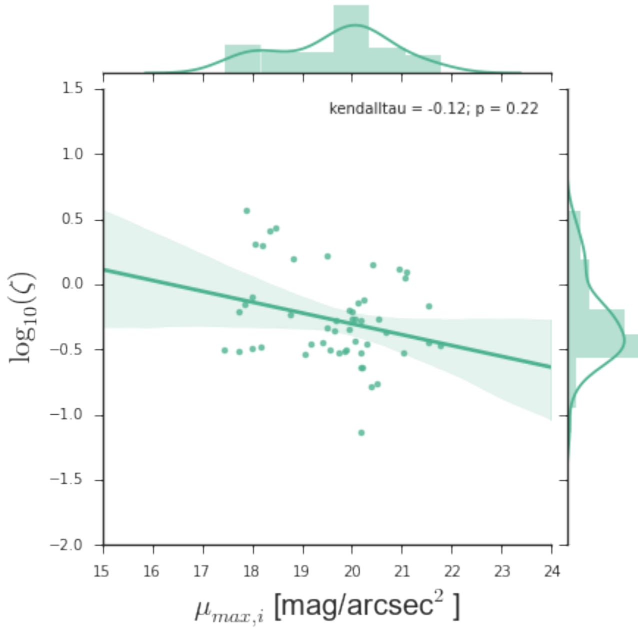


Figure 7. All fits for the vote fraction dropoff parameter ζ for $f_{\rm features}$ in the FERENGI galaxies as a function of surface brightness. This includes only the 37 galaxies with a reasonably bounded range on the dropoff ($-10 < \log(\zeta) < 10$) and sufficient points to fit the function.

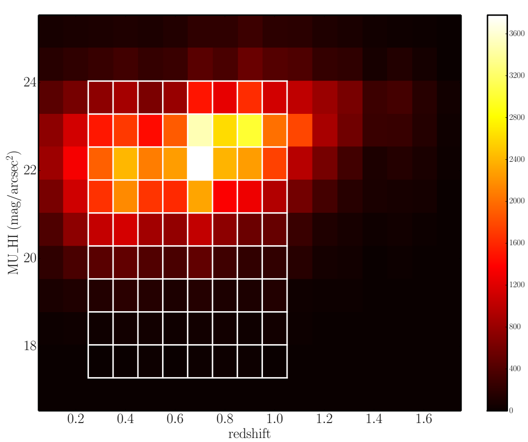


Figure 8. Surface brightness vs. redshift of 118,083 galaxies in the ACS sample. The white grid denotes the surface brightness and redshift range of the FERENGI images, subdivided in bins corresponding to fixed ranges used for analysis in Figure 9.

sented by the orange bars in Figure 9(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 $p_{{\rm features},z=0.3}$ to be for any observed value of $p_{\rm features}$.

Table 5. Distribution of FERENGI images analysed in Figure 9. Correctable images had a single-valued relationship between their measured $p_{\rm features}$ values at high and low redshifts (white regions in Figure 9). 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%

4.4. *Challenges of debiasing questions beyond "smooth or features"*

As with the HST images, each FERENGI subject has a varying number of users answering the various questions in the hierarchical 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 $p_{\rm smooth}$, $p_{\rm features}$, and $p_{\rm artifact}$ are all computed with the minimum statistical error for any question, with roughly 40 total answers (see Section 3). The number of users to answer any subsequent question, however, is always equal to or less than the number to answer the preceding question. The average number of responses per task for fourth- or fifth-tier questions such as spiral arm structure (Tasks 12–14) is only 4 ± 4 for the FERENGI sample; while this is distribution is strongly bimodal (reflecting the true morphologies of selected galaxies), the very low absolute numbers of votes introduce very high variance when attempting to calculate a statistical correction.

In the FERENGI data, these numbers severely limit the amount of information we can extract for the higher-tier questions. The debiasing technique used (Section 4.2) require that at least 5 users answer each question for a galaxy image at $z = 0.3$ *and* its image at higher redshift. This requirement placed on both images is not met by a significant number of galaxies for questions beyond Task 01. The FERENGI images classified do not have sufficient numbers of objects in the surface brightness/redshift bins to accurately measure if the relation between vote fractions and redshift is single-valued.

for this reason we only offer debiased vote fractions for Task 01 (smooth/features/artifact).

Perhaps compute number of galaxies that can be fit to zeta for each question, show a table? overkill?

In Section A.2 we show results of an attempt to measure ζ for $p_{\rm bar}$.

- talk about where the *HST* sample falls in this space, reference Table 6
- justify $N > 5$ and spread < 0.2 (or find a better

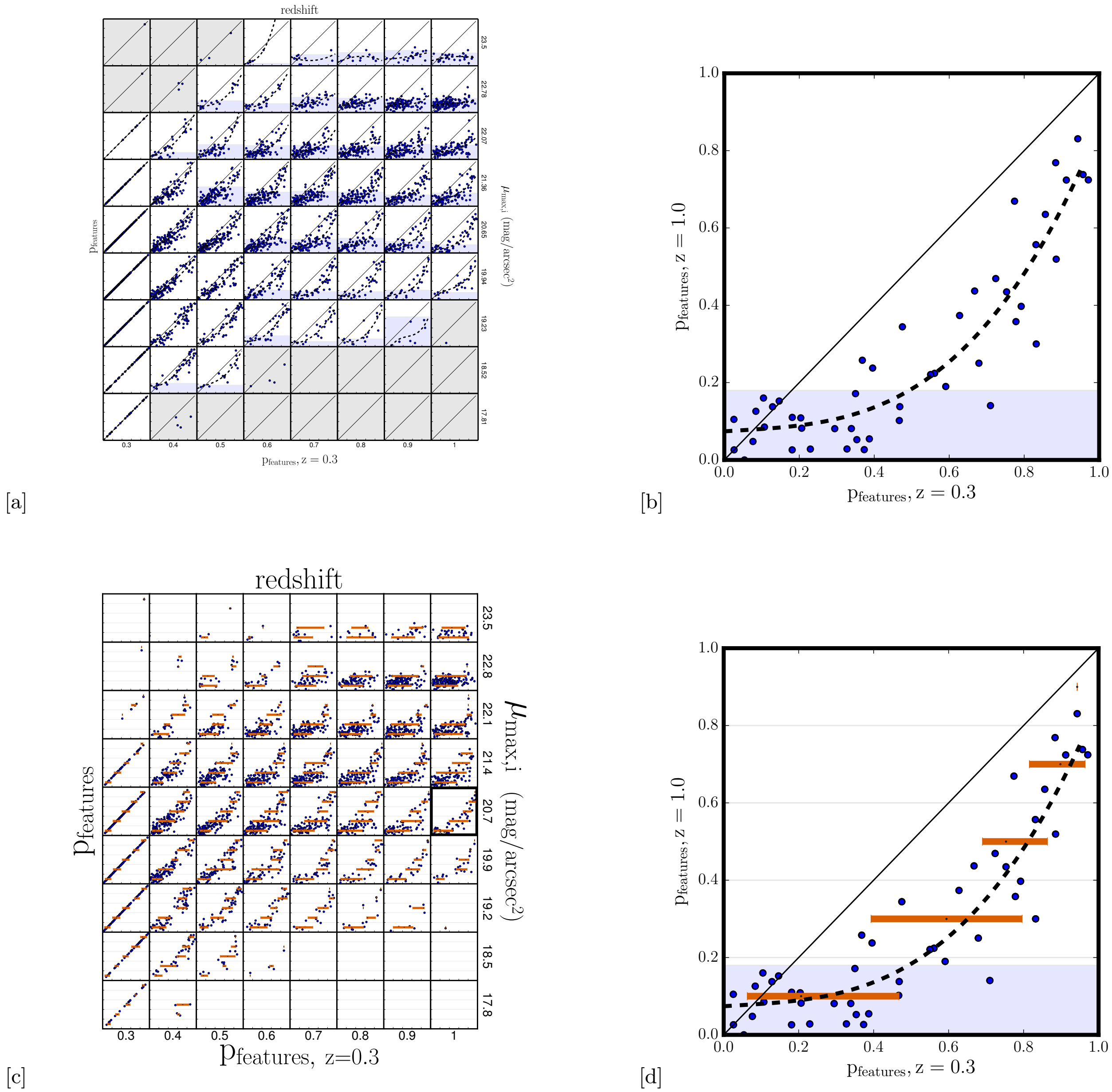


Figure 9. Effects of redshift bias in 3,950 images in the ferengi sample. [a]: Each point in a given redshift and surface brightness bin represents a unique galaxy. On the y-axis in each bin is the $p_{\rm features}$ value of the image of that galaxy redshifted to the value corresponding to that redshift bin. On the x-axis is the $p_{\rm features}$ 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 $p_{{\rm features},z}$=$p_{{\rm features},z=0.3}$. Regions in which there is a single-valued relationship between $p_{\rm features}$ 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. [b]: A larger version of the dark-outlined square in [a], 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). [c]: The same data as [a] is shown. Each z, μ bin is divided into 4 sub-bins to determine the range of intrinsic $p_{{\rm features},z=0.3}$ for a given range of observed $p_{{\rm features},z}$ 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. [d]: The same data as [b], but highlighting the upper and lower limit regions.

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*

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, $\mathtt{N_{votes}}$ is the number of users to answer that question. For each unique answer, `fraction` is the fraction of users to select that answer ($\mathrm{N_{answer}/N_{votes}}$, and `weighted` is the weighted fraction, which takes into account user consistency (Section 3.1).

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 and best vote fractions for $\mathtt{p_{smooth}}$ were calculated on the critera that vote fractions for all answers must sum to unity. Explicitely: $\mathtt{p_{smooth}} = \mathtt{1} - \mathtt{p_{features}} - \mathtt{p_{artifact}}$. In some rare cases, representing $< 1\%$ of the sample, this requirement resulted in negative vote fractions for $\mathrm{p_{smooth}}$; these were cases in which the $\mathrm{p_{features}}$ vote fraction was boosted to a high value relative to the $\mathrm{p_{artifact}}$ vote fraction. In these cases, we enforce the restriction that the vote fractions must sum to 1 by raising $\mathrm{p_{smooth}}$ to a vote fraction of 0.0 and decreasing $\mathrm{p_{features}}$ accordingly. This correction was typically very small, with a median decrease/increase of $\mathrm{p_{features}/p_{smooth}}$ of $\Delta p = 0.04$.

We split the data products for GZH by the type of image being classified. Table 7 contains the classifications for the *HST* 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 reprocessed 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).

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

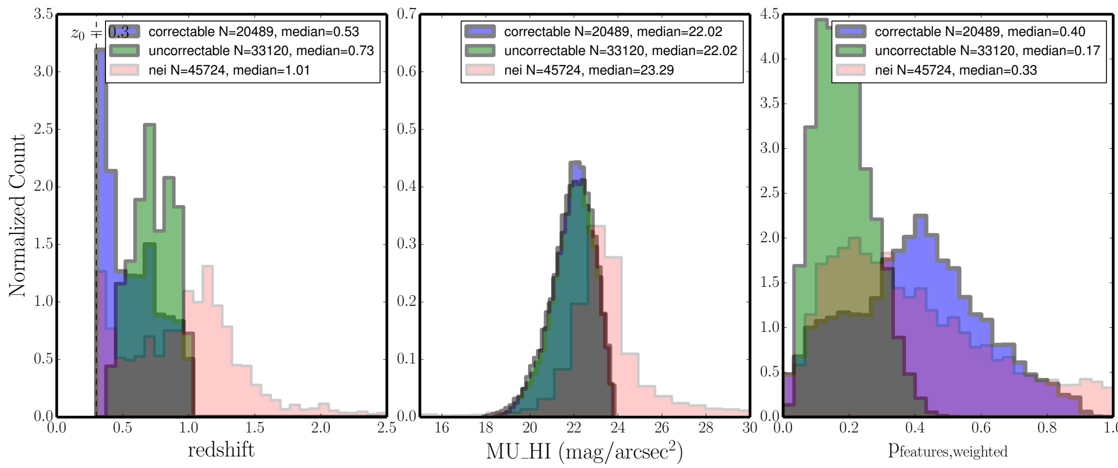


Figure 10. Distributions of redshift, surface brightness, and $p_{features}$ for correctable (purple), uncorrectable (green), and NEI (pink) galaxies in the full GZH sample. The uncorrectable galaxies tend towards higher redshift, slightly lower in surface brightness, and lower values of $p_{features}$ 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 $\mu > 24$.

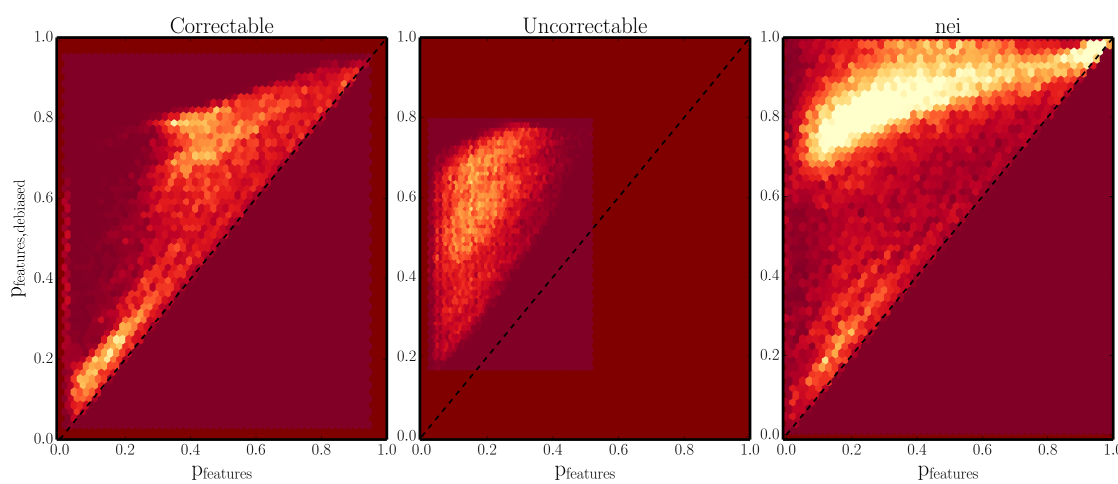


Figure 11. Debiased $p_{features}$ corrected to $z = 0.3$ vs weighted $p_{features}$ for the correctable (left), uncorrectable (middle), and NEI (right) galaxies in the GZH sample.

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

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_						...
Project ID	Hubble ID	Imaging	Correction[1]	$N_{\rm votes}$	fraction	weighted	debiased	best	lower limit	upper limit	
AHZ100002g	10010842	AEGIS	0	127	0.118	0.128	0.085	0.085	0.226	0.226	
AHZ100002h	10010870	AEGIS	4	127	0.567	0.592	0.927	0.592	–	–	
...											
AHZ20004kd	20014731	COSMOS	3	44	0.682	0.675	0.147	0.675	–	–	
AHZ20004ke	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	0.800	
AHZ4000416	90022735	GEMS	1	127	0.646	0.698	0.508	0.698	0.171	0.727	
...											
AGZ0007z47	10014	GOODS-N-FULLDEPTH	1	40	0.475	0.475	0.197	0.475	0.011	0.496	
AGZ0007z48	10017	GOODS-N-FULLDEPTH	3	40	0.675	0.675	0.048	0.675	0.168	0.669	
...											
AGZ00083jb	8869	GOODS-S-FULLDEPTH	1	40	0.425	0.425	0.109	0.425	0.070	0.548	
AGZ00083jc	8878	GOODS-S-FULLDEPTH	0	40	0.205	0.205	0.048	0.048	−0.005	0.287	
...											

[1] Flag indicating how the vote fractions for this galaxy were corrected through debiasing (§4.3), if possible. 0 = correctable, 1 = uncorrectable (p_{raw} — p_{adj} is r
single-valued), 2 = uncorrected ($z_{\rm gal} < 0.3$), 3 = uncorrectable (insufficient FERENGI galaxies in this z-μ 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 galax
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

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_						...
Project ID	Hubble ID	Imaging	Correction	$N_{\rm votes}$	fraction	weighted	debiased	best	lower limit	upper limit	
AHZF000001	20000002	COSMOS	1	48	0.708	0.755	0.228	0.754	0.325	0.829	
AHZF000003	20000004	COSMOS	3	49	0.367	0.379	0.100	0.379	0.198	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	0.316	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

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_						...
Project ID	Hubble ID	Imaging	Correction	$N_{\rm votes}$	fraction	weighted	debiased	best	lower limit	upper limit	
AHZC000001	20000002	COSMOS	1	168	0.615	0.664	0.160	0.663	0.271	0.775	
AHZC000003	20000004	COSMOS	0	235	0.333	0.364	0.002	0.002	0.063	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	0.249	0.796	
AHZC000010	20000104	COSMOS	2	158	0.843	0.882	0.936	0.881	–	–	

Table 9 continued

Table 9 *(continued)*

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_						...
Project ID	Hubble ID	Imaging	Correction	N_{votes}	fraction	weighted	debiased	best	lower limit	upper limit	
...											

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

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_						...
Project ID	Hubble ID	Imaging	Correction	N_{votes}	fraction	weighted	debiased	best	lower limit	upper limit	
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

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_				...
Project ID	SDSS DR7 ObjID	Imaging	Correction	N_{votes}	fraction	weighted	debiased	best	
AHZ5000001	587730845812064684	SDSS	2	41	0.585	0.595	0.759	0.594	
AHZ5000002	587730845812065247	SDSS	2	46	0.609	0.651	0.897	0.651	
AHZ5000003	587730845812196092	SDSS	2	51	0.039	0.044	0.067	0.043	
AHZ5000004	587730845812196825	SDSS	2	35	0.514	0.605	0.928	0.605	
AHZ5000005	587730845812524122	SDSS	2	47	0.766	0.812	1.038	0.810	
AHZ5000006	587730845812654984	SDSS	2	42	0.5	0.542	0.680	0.541	
AHZ5000007	587730845812655541	SDSS	2	41	0.488	0.526	0.697	0.525	
AHZ5000008	587730845812720365	SDSS	2	53	0.792	0.84	1.050	0.839	
AHZ5000009	587730845812720640	SDSS	4	43	0.0	0.0	0.0	0.0	
AHZ500000a	587730845812720699	SDSS	2	40	0.425	0.478	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 21,522 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

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_				...
Project ID	SDSS DR7 ObjID	Imaging	Correction	N_{votes}	fraction	weighted	debiased	best	
AHZ6000001	8647474690312306978	SDSS	4	40	0.275	0.289	0.762	0.289	
AHZ6000002	8647474690312307154	SDSS	2	43	0.605	0.634	0.858	0.635	
AHZ6000003	8647474690312307877	SDSS	2	51	0.608	0.627	0.906	0.627	
AHZ6000004	8647474690312308301	SDSS	4	52	0.038	0.038	0.723	0.038	
AHZ6000005	8647474690312308318	SDSS	2	44	0.614	0.632	0.776	0.631	
AHZ6000006	8647474690312308880	SDSS	2	36	0.667	0.683	0.901	0.683	
AHZ6000007	8647474690312372644	SDSS	4	48	0.646	0.674	1.145	0.674	
AHZ6000008	8647474690312372789	SDSS	4	45	0.489	0.571	0.964	0.570	
AHZ6000009	8647474690312372931	SDSS	4	47	0.553	0.587	0.926	0.587	
AHZ600000a	8647474690312373190	SDSS	4	47	0.574	0.559	1.008	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

										t01_smooth_or_features_a01_smooth_		
Project ID	SDSS DR7 ObjID	Imaging	Correction	Version	L_{ratio}	AGN color[1]	N_{votes}	fraction	weighted	debiased	best	lower limit
AHZ7000001	90024700	GEMS	1	1	0.2	1	42	0.238	0.239	−0.110	0.238	−0.113
AHZ7000002	90024700	GEMS	1	1	1.0	1	51	0.255	0.265	−0.107	0.264	−0.128
AHZ7000003	90024700	GEMS	0	1	5.0	1	47	0.170	0.167	−0.018	−0.018	−0.049
AHZ7000004	90024700	GEMS	0	1	10.0	1	41	0.195	0.195	0.045	0.045	0.044
AHZ7000005	90024700	GEMS	0	1	50.0	1	47	0.170	0.178	0.067	0.067	0.146
...												
AHZ700013m	90024700	GEMS	0	2	0.0	0	35	0.171	0.136	0.011	0.011	0.029
AHZ700013n	90024700	GEMS	1	2	0.2	1	20	0.150	0.158	−0.278	0.049	−0.351
AHZ700013o	90024700	GEMS	1	2	1.0	1	32	0.281	0.300	−0.086	0.281	−0.119
AHZ700013p	90024700	GEMS	0	2	5.0	1	29	0.103	0.115	−0.098	−0.098	−0.152
AHZ700013q	90024700	GEMS	0	2	10.0	1	35	0.171	0.181	0.027	0.027	0.023
AHZ700013r	90024700	GEMS	0	2	50.0	1	34	0.206	0.206	−0.005	−0.005	−0.056
...												

[1] Flag 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 galaxie information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

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

6. USING THE CATALOG

The intended purpose of the GZH catalog is to provide a simple, yet flexible, means of identifying samples of galaxies with a desired morphological type. Here we provide instructions for creating such samples using the vote fractions corresponding to the tasks shown in Figure 4. We stress that the selection process will vary based on the particular science case. More conservative cuts can be applied to create pure, but not necessarily complete, samples. These are useful for selecting galaxies exhibiting unique morphologies for individual case studies or observing proposals. Looser cuts can be applied to obtain samples with a higher level of completeness, which will increase the sample size but can potentially decrease purity. Population studies would make use of such large samples, in which statistical significance is a crucial factor in evaluating results. Masters et al. (2011); Melvin et al. (2014); Cheung et al. (2015), and Galloway et al. (2015) are examples of GZ papers which select morphologies in this way.

In GZH, volunteers answer questions about a galaxy's morphology in a decision tree format. With this structure, questions shown to a user are dependent on their answers to the previous questions. For example, a user is only asked Task 11, "How many spiral arms?" if they had answered "yes" to Task 10, which asks whether any spiral arms are present. This structure is used to reduce the workload of the users by only asking questions relevant to the particular galaxy; in this example, it would not be useful, nor would it make physical sense, to ask a user how many spiral arms in a galaxy which they had declared had no spiral arms to begin with. Figure 4 shows the possible paths available for answering the questions offered in GZH. The colors represent the tier level of Tasks, which indicate the number of previous Tasks the given Task is dependent upon. The arm-number question, Task 11, is a 5th tier Task, meaning that whether this question is seen by a user is dependent on four Tasks preceding it. In this case, it is only shown to volunteers who voted the galaxy was featured/disk-shaped in Task 01, not clumpy in Task 12, not edge-on in Task 02, and had spiral structure in Task 10.

To select galaxies of a morphological type identified with a particular Task, a cut is placed on the vote fraction for that Task ($p_{\rm task}$), *as well as* the vote fractions for the Tasks preceding it, because of the dependency induced by the decision-tree structure. For example, to select barred galaxies, a cut may be placed on $p_{\rm bar}$ such that only galaxies where a high fraction of votes for this task voted for the bar, yes answer. This is not the only necessary cut, however, since not all users answer this question; only those who have previously selected "features" in Task 01, "not clumpy" in Task 12, and "not edge-on" in Task 02 will have the opportunity to vote on the bar question, Task 03. To ensure that $p_{\rm bar}$ is well-sampled, cuts on all previous tasks must be applied.

The flexibility of this catalog allows the user to set their own selection criteria for vote fraction thresholds to create a morphologically-pure sample. In Table 14 we offer baseline cuts for selecting galaxies of a variety of morphologies. We detemined these thresholds by a visual inspection of subsamples. For each Task, we first required that at least 20 users voted on the given question. We then applied a cut on the vote fraction for the previous task and analyzed a subsample of 50 galaxies meeting these criteria, as well as a control sample of galaxies which had 20 users vote on the task, but did not meet the threshold cut set for the previous task. The threshold cut was adjusted and new subsamples were inspected until both the original and control samples achieved $> 80\%$ purity.

As an example of how to use Table 14 to create a sample of 3-armed spirals galaxies, one should select objects with $N_{arm\ number} \geq 20$, $p_{\rm features} > 0.23, p_{\rm clumpy,no} > 0.30, p_{edgeon,no} > 0.25$, and $p_{spiral,yes} > 0.25$. These cuts define a sample of galaxies of "arm number candidates"; ie galaxies for which answering the arm number question makes physical sense and the vote fraction $p_{\rm arm\ number}$ is well-sampled. In this example, such galaxies are featured, non-clumpy, non-edge on, spiral galaxies. At this point a cut can be made on $p_{\rm arm\ number=3}$ to select spirals with three arms.

Tasks 03, 04 and 05 have an additional possible pathway; as shown in Figure 4, a user might also be shown this question if they select "featured/disk" in Task 01, "clumpy" in Task 12, two or more clumps in Task 16, and "spiral arrangement" in Task 15. By applying the appropriate thresholds for this path, we find that $< 0.5\%$ of the galaxies which have ≥ 20 answers to these questions used this pathway to arrive at these Tasks. Further, of these subjects, none were found to exhibit disk structure, although the clumps within were arranged in a spiral pattern. We therefore do not find this pathway necessary in examining responses to Tasks 03, 04, and 05.

In this section we have described how to use the previous Task thresholds in Table 14 to select well-sampled galaxies which may or may not contain the feature associated with a unique Task. We will now offer two examples of how to use the vote fractions for the final Task to obtain a sample of galaxies with a certain morpholgical type, by creating a sample of bars and samples of clumpy galaxies grouped by clump mulitiplicity.

6.1. *Example 1: Selecting bars*

We created a sample of barred galaxies by applying cuts on the previous tasks as listed in Table 14. 11,049 "bar candidates", galaxies for which asking the bar question is meaningful, were selected by applying the cuts $N_{bar} \geq 20$, $p_{\rm features} > 0.23, p_{\rm clumpy,no} > 0.30$, and $p_{edgeon,no} > 0.25$. These galaxies are featured, non-clumpy, non-edge on galaxies. Of these, a pure sample of 730 barred disks was identified by applying a cut of $p_{bar} > 0.7$. A subsample of 50 galaxies were visually inspected and 94% were found to contain strong bars. A complete sample of strong and weak bars was created by applying a cut of $p_{bar} > 0.3$. This sample contained 3,218 galaxies, 86% of which were found to contain weak or strong bars through visual inspection in a subsample of 50.

We use the complete bar sample to estimate the redshift evolution of bar fraction and find a steady decrease of $f_{bar} \sim 0.32$ at $z = 0.4$ to $f_{bar} \sim 0.24$ at $z = 1.0$. The decrease in bar fraction agrees with **??**, however they report a lower bar fraction overall, finding a decrease from $f_{bar} = 0.22$ at $z = 0.4$ to $f_{bar} = 0.11$ at $z = 1.0$. The difference in total bar fraction is expected, as our analysis used a looser cut on $p_{\rm bar}$, we did not apply a luminosity cut, and we use debiased values for $p_{\rm features}$ which increases the total amount of disks in the sample. For a more thorough analysis of the bar fraction, a cut on $p_{\rm bar}$ that evolves with redshift may yield more accurate results, since there are no explicit debiased values of $p_{\rm bar}$ that would take redshift induced bias into account.

6.2. *Example 2: Identifying clump multiplicity*

Clumps are known to be a characteristic feature of galaxies outside the local universe, and there is evidence they play a crutial role in the evolution of modern spirals, particularly in the formation of central bulges (Elmegreen et al. 2005; Elmegreen & Elmegreen 2014; Guo et al. 2015; Behrendt et al. 2016). Simulations show clumps migrate from the outer disk to the galactic center on relatively short (few orbital periods) timescales (Mandelker et al. 2015) and observations show increasing bulge to clump mass and density ratios as the universe evolves since $z \sim 1.5$ (Elmegreen et al. 2009), suggesting that clumps coalesce over time to form the modern bulges of disk galaxies. GZH added a "clumpy" path to the decision tree for the purpose of both identifying clumps and investigating their evolution with redshift. Here we provide an example of how to select galaxies with particular clump multiplicities, for investigations related to the transformation of clumps as galaxies continue to form.

For galaxies identified as "clumpy" in GZH, the number of clumps can be determined using Task 16. We identify samples of galaxies having one, two, three, four, and more than four clumps. Using Table 14, we identify 8,444 clumpy galaxies using $p_{\rm features} > 0.23$ and $p_{\rm clumpy,yes} > 0.80$ to ensure the vote fractions for Task 16 are well-sampled. We were able to identify with reasonable confidence the clump number of 1,112 of the clumpy galaxies; in the remainder, the unique clumps were less distinguishable from each other and the exact number of clumps could not be deduced without careful visual inspection. In the 1,112 which did have distinguishable clumps, we identified 61 one-clump galaxies using $p_{\rm 1\ clump} > 0.50$, 442 two-clump galaxies using $p_{\rm 2\ clumps} > 0.80$, 275 three-clump galaxies using $p_{\rm 3\ clumps} > 0.75$, 71 four-clump galaxies using $p_{\rm 4\ clumps} > 0.70$, and 263 galaxies with more than four clumps using $p_{\rm >4\ clumps} > 0.70$. Alternatively, these data may be used to create more general samples of clumpy galaxies with few clumps and many clumps. We define a sample of 989 "few clumps" galaxies using $(p_{\rm 1\ clump} + p_{\rm 2\ clumps}) > 0.5$ and 2,910 "many clumps" galaxies using $(p_{\rm 3\ clumps} + p_{\rm 4\ clumps} + p_{\rm >4\ clumps}) > 0.5$.

Table 14. Suggested thresholds for selecting morphological samples from Galaxy Zoo: Hubble

No.	Task	Previous task(s)	Vote fraction threshold $N_{\rm task} \geq 20$
01	smooth or features	–	–
02	edge on	01, 12	$p_{\rm clumpy,no} > 0.30$
03	bar	01, 12, 02	$p_{\rm edgeon,no} > 0.25$
		01, 12, 16, 15	$p_{\rm clumpy\ spiral} > 0.65$
04	spiral arms	01, 12, 02	$p_{\rm edgeon,no} > 0.25$
		01, 12, 16, 15	$p_{\rm clumpy\ spiral} > 0.65$
05	bulge prominence	01, 12, 02	$p_{\rm edgeon,no} > 0.25$
		01, 12, 16, 15	$p_{\rm clumpy\ spiral} > 0.65$
06	odd yes/no	–	–
07	rounded	01	$p_{\rm smooth} > 0.70$
08	odd feature	06	$p_{\rm odd,yes} > 0.50$
09	bulge shape	01, 12, 02	$p_{\rm edgeon,yes} > 0.40$
10	arms winding	01, 12, 02, 04	$p_{\rm spiral,yes} > 0.25$
11	arms number	01, 12, 02, 04	$p_{\rm spiral,yes} > 0.25$
12	clumpy	01	$p_{\rm features} > 0.23$
13	bright clump	01, 12, 16	$p_{\rm one\ clump} < 0.40$
14	bright central clump	01, 12, 16, 13	$p_{\rm bright\ clump,yes} > 0.50$
15	clump arrangement	01, 12, 16	$p_{\rm multiple\ clumps} > 0.45$
16	clump count	01, 12	$p_{\rm clumpy,yes} > 0.80$
17	clumps symmetrical	01, 12	$p_{\rm clumpy,yes} > 0.80$
18	clumps embedded	01, 12	$p_{\rm clumpy,yes} > 0.80$

- ~~explain decision tree format; vote fractions are dependent on previous task vote fractions~~
- ~~provide example of how to select bars~~
- ~~discuss difference between conservative/clean sample and loose/complete sample, emphasize goals differ depending on science case~~
- ~~science with bar example - estimate bar fraction vs redshift~~
- ~~science with clump multiplicity~~

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~~ (Häußler et al. 2007)
- ~~AEGIS morphologies~~ (Lotz et al. 2008)
- ~~GOODS N/S morphologies~~ (Bundy et al. 2005)

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.

7.2.1. *GEMS*

Morphologies for galaxies in GEMS imaging have been measured by Häußler et al. (2007), who perform a careful comparison of single-component Sérsic fits to the F850LP imaging using both the GALFIT and GIM2D codes. We compare the GZH morphological vote fractions to the GALFIT parameters, which are evaluated by Häußler et al. (2007) as more reliable due to GALFIT's ability to fit multiple galaxies and handle crowded fields.

A single Sérsic fit is often a poor fit to the light profile of a resolved galaxy, since large numbers of classical spirals have both a de Vaucouleurs bulge at the center and an exponential disc at larger radii. If a single parameter is used, however, a common division in the literature is to split on the Sérsic index n into disc-dominated galaxies ($n < 2.5$) and bulge-dominated galaxies ($n > 2.5$), where n is the exponent in the radial surface brightness profile such that $\Sigma(r) = \Sigma_e \exp\left(-\kappa[(r/r_e)^{1/n} - 1]\right)$.

A $1''$ positional match for the GEMS galaxies in Table 9 in Häußler et al. (2007) to the GZH sample gives 8,846 galaxies found in both samples. We analyze galaxies for which the GALFIT parameters are flagged as being constrained within plausible physical limits, indicating a moderately robust fit. There is a strong correlation between n and $p_{\mathrm{features,best}}$; the average value for disk-dominated galaxies ($n < 2.5$) is $\langle p_{\mathrm{features,best}}\rangle = 0.44 \pm 0.26$, compared to $\langle p_{\mathrm{features,best}}\rangle = 0.30 \pm 0.25$ for bulge-dominated galaxies ($n > 2.5$). The correlation is stronger for bulge-dominated galaxies; for $n > 2.5$, most galaxies have $p_{\mathrm{features,best}} < 0.2$. In contrast, while $n < 2.5$ do have higher $p_{\mathrm{features,best}}$ values, the spread is larger and there are significant numbers of galaxies even at the highest $p_{\mathrm{features,best}}$ values.

Inverting the comparison, we split the GZH galaxies into three samples: "clean" disks ($p_{\mathrm{features,best}} \geq 0.8$), "clean" ellipticals ($p_{\mathrm{features,best}} \leq 0.2$), and "intermediate" ($0.2 < p_{\mathrm{features,best}} < 0.8$) galaxies (eg, Schawinski et al. 2014). Splitting by $n < 2.5/n > 2.5$, clean disks are 85%/17%, intermediates 84%/16%, and clean ellipticals 59%/41%. There is a strong dependence for these ratios on the brightness of the galaxy, which affects both automatic fits and visual accuracy. Limiting the sample to galaxies with $m_{850LP} < 21$ gives fractions of: clean disks are 77%/23%, intermediates 52%/48%, and clean ellipticals 38%/62%. While the clean disk fraction is largely unchanged, intermediates are more evenly balanced and there is a larger population of smooth galaxies with high n, as expected.

We visually inspect galaxies for which the GALFIT and GZH measurements indicate strong disagreement. Of the 84 galaxies with $n > 2.5$ and $p_{\mathrm{features,best}} \geq 0.80$, almost all are spirals with obvious disks and clearly marked spiral arms. These appear to be cases where GALFIT fits the bulge at the center, but misses extended emission from the disk. We also examine the 35 galaxies with $n < 2.5$ and $p_{\mathrm{features,best}} \leq 0.20$; these galaxies generally have smoother light profiles with no spiral arms, although several have bars and/or the sharper truncation of light near the edges that indicates an S0 galaxy. These galaxies are also significantly bluer ($\langle m_{606} - m_{850}\rangle = 0.27 \pm 3.3$) compared to the majority of galaxies classified as "smooth" in GZH ($\langle m_{606} - m_{850}\rangle = 1.2 \pm 1.7$). We interpret this as evidence for the fact that single-component radial fits are often inadequate to accurately characterize galaxy morphology, and that a simple cut risks significant amounts of misclassified objects.

Clumpy galaxies in GZH have GALFIT parameters more consistent with disky, $n < 2.5$ galaxies than spheroids (Figure 12). Using the thresholds in Table 14, GZH clumpy galaxies have average values of $n = 1.6 \pm 1.2$ and $R_e = 32 \pm 16$ pix, with only 11% of clumpy galaxies falling in the bulge-dominated region of $n > 2.5$.

7.2.2. *AEGIS*

Galaxies in AEGIS were morphologically classified using non-parametric measurements by Lotz et al. (2008). They use the Gini coefficient (G), which measures the relative inequality in pixel brightness, and M_{20}, the second-order moment of the brightest 20% of the light (Lotz et al. 2004). A linear combination of G and M_{20} to delineate three broad categories of galaxy morphology: E/S0/Sa (bulge-dominated/spheroids), Sb/Sc/Ir (disk-dominated/irregulars), and mergers.

We compare the GZH vote fractions to the morphological categories defined by their G-M_{20} values, which

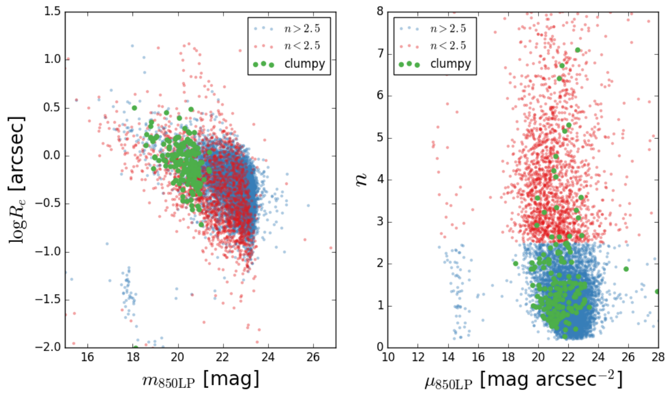


Figure 12. Parameter distributions for GEMS galaxies in Häußler et al. (2007) and GZH. Red and blue points indicate galaxies split by their Sérsic index (at $n = 2.5$) into disk- and bulge-dominated galaxies, respectively. Galaxies identified as "clumpy" in GZH are labeled in green.

are measured in both V- and I-band. A $1''$ positional match on AEGIS from Lotz et al. (2008) and GZH yields 7,508 galaxies; of those, we analyze a subsample of 4,031 with reliably-measured morphologies and $S/N > 3$ in both bands.

The GZH morphologies are reasonably correlated with the G-M_{20} morphological categories. For galaxies labeled as Sb/Sc/Ir by Lotz et al. (2008), 58% of the galaxies in I-band (61% in V-band) have $p_{\mathrm{features,best}}$ above a nominal cut of 0.5. E/S0/Sa galaxies have a tighter relation, with 78% of galaxies in I-band (79% in V-band) having $p_{\mathrm{smooth}} > 0.5$. Galaxies with high $p_{\mathrm{features,best}}$ are largely concentrated in the E/S0/Sa region of G-M_{20} space; the small overlaps into mergers and E/S0/Sa are close to what we expect for loosely defined boundaries. Galaxies with low $p_{\mathrm{features,best}}$ have a much broader range of G-M_{20} values, with nearly half lying well within the Sb/Sc/Ir region.

Of the 47 galaxies identified by GZH as mergers in the cross-matched AEGIS sample (selected as $p_{\mathrm{merger}} > 0.5$ and $N_{\mathrm{merger}} \geq 20$), roughly two-thirds fall in the merger region of G-M_{20} space defined in Lotz et al. (2008); almost all the remainder would be classified as Sb/Sc/Ir. Galaxies identified as mergers by G-M_{20}, in contrast, typically have lower p_{merger} values; 66%/50% in V-band and I-band, respectively. While the total number of mergers identified in GZH is roughly $\times 10$ fewer than the volume-limited G-M_{20} sample, we do confirm the result of a non-evolving merger fraction between $0.2 < z < 1.2$.

7.2.3. *GOODS*

Galaxies in the GOODS-N and GOODS-S fields down to a limit of $z_{AB} = 22.5$ have been visually classified by a single expert (R.S. Ellis), inspecting both z-band and composite Viz color images (Bundy et al. 2005). These morphologies are assigned a numerical value based on categories in Brinchmann et al. (1998); these are combined into broader categories of "E/S0" (classes 0,1,2), "Spirals" (classes 3,4,5), and "Peculiar/Irregular" (classes 6,7,8). Bundy et al. (2005) use these morphologies to show that there are an increasing proportion of early-type galaxies observed at later times.

Matching the Bundy et al. (2005) catalog to GZH with a $0.5''$ radius yields 2435 galaxies (1300 in GOODS-N, 1135 in GOODS-S). We compare their broad morphological categories to $p_{\mathrm{features,best}}$ in GZH; demographics of the separate GOODS-N and GOODS-S samples are highly consistent. "E/S0" galaxies are dominated by high smooth fractions in GZH, with an average value of $\langle p_{\mathrm{smooth,best}} \rangle = 0.54 \pm 0.20$. The effect is stronger when examining sub-divisions; pure E galaxies have the highest $p_{\mathrm{smooth,best}}$ values, while S0 have the lowest. Visual inspection of galaxies designated as S0 show numerous examples of both edge-on galaxies and visible substructure in the disk, including spiral arms. This is consistent with the design of the GZH decision tree, and likely reflects that the expert categories are using different criteria.

"Spiral" galaxies in Bundy et al. (2005) have robustly high GZH vote fractions of features and/or disks, with $\langle p_{\mathrm{features,best}} \rangle = 0.59 \pm 0.06$. The vote fraction increases with Hubble type, with Scd galaxies having the highest average $p_{\mathrm{features,best}}$ values. "Peculiar/Irregular" galaxies have similar ranges of $p_{\mathrm{features,best}}$ to the "Spiral" galaxies, but can be sub-selected with cutoffs on p_{odd} and $p_{\mathrm{irregular}}$ or p_{merger}. We note that mergers are not strongly correlated in GZH and the expert visual classifications; between 20%–30% have GZH merger fractions of zero, while even clear tidal tails can be as low as $p_{\mathrm{merger}} \sim 0.1$. The total number of major mergers is small enough that expert visual inspection can be practically done on all candidates; a combination of methods may be the most accurate way of identifying these rarer objects, especially when the signs of interaction are relatively subtle.

8. SUMMARY

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

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. We also thank Jennifer Lotz for sharing her G-M_{20} measurements for the AEGIS sample.

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). Modified code from Nick Wherry and David Schlegel was used to create the JPG images.

Hubble Space Telescope 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
Bamford S. P., Rojas A. L., Nichol R. C., Miller C. J., Wasserman L., Genovese C. R., Freeman P. E., 2008, MNRAS, 391, 607
Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, MNRAS, 422, 449
Barden M., Jahnke K., Häußler B., 2008, ApJS, 175, 105
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Behrendt, M., Burkert, A., & Schartmann, M. 2016, ApJL, 819, L2
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Brinchmann J. et al., 1998, ApJ, 499, 112
Bundy K., Ellis R. S., Conselice C. J., 2005, ApJ, 625, 621
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., Labbé 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, ApJL, 754, L24
Cheung, E., Trump, J. R., Athanassoula, E., et al. 2015, MNRAS, 447, 506
Conselice C. J., 2003, ApJS, 147, 1
Conselice C. J., 2014, ARA&A, 52, 291
Darg D. W. et al., 2010, MNRAS, 401, 1552
Davis M. et al., 2007, ApJL, 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
Elmegreen, D. M., Elmegreen, B. G., Marcus, M. T., et al. 2009, ApJ, 701, 306
Elmegreen, D. M., & Elmegreen, B. G. 2014, ApJ, 781, 11
Förster Schreiber N. M., Shapley A. E., Erb D. K., Genzel R., Steidel C. C., Bouché 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
Galloway, M. A., Willett, K. W., Fortson, L. F., et al. 2015, MNRAS, 448, 3442
Genel S. et al., 2014, MNRAS, 445, 175
Giavalisco M. et al., 2004, ApJL, 600, L93
Griffith R. L. et al., 2012, ApJS, 200, 9
Grogin N. A. et al., 2011, ApJS, 197, 35
Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, ApJ, 800, 39
Häußler B. et al., 2007, ApJS, 172, 615
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. et al., 2008, ApJ, 672, 177
Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163
Lupton R., Blanton M. R., Fekete G., Hogg D. W., O'Mullane W., Szalay A., Wherry N., 2004, PASP, 116, 133
Masters K. L. et al., 2011, MNRAS, 411, 2026
Mandelker, N., Dekel, A., Ceverino, D., et al. 2015, arXiv:1512.08791
Melvin, T., Masters, K., Lintott, C., et al. 2014, MNRAS, 438, 2882
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

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

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
Taniguchi Y. et al., 2007, ApJS, 172, 9
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

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.

A.1. *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 $\mathrm{p_{features}}$ for both sets of images. The mean change in $\mathrm{p_{features}}$ from the shallow to full depth images $\mathrm{p_{features,full}} - \mathrm{p_{features,shallow}} \equiv \Delta\mathrm{p} = 0.00$, 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 $p_{features,best} > 0.8$, elliptical galaxies as those with $p_{smooth,best} < 0.2$, 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.

A.2. *FERENGI analysis of pbar*

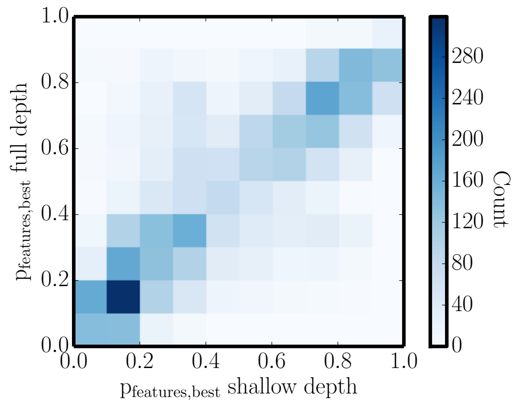


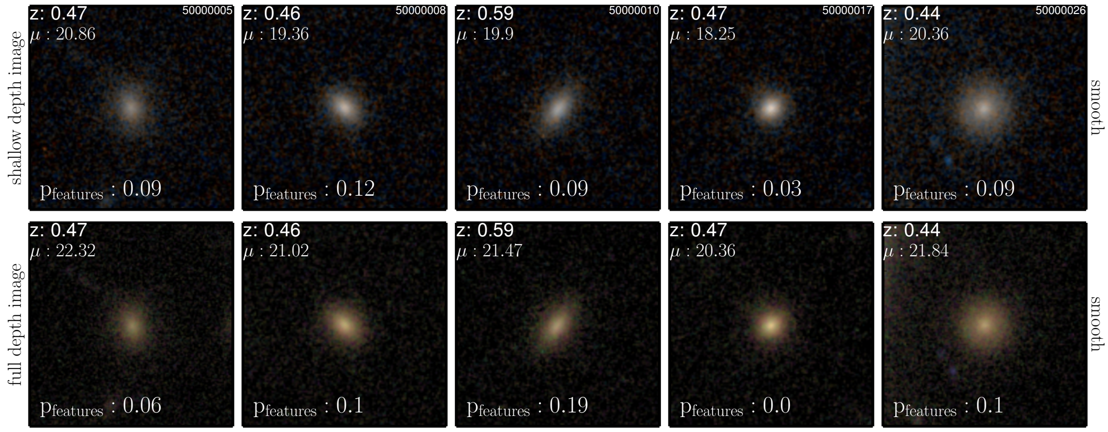
Figure A1. shallowfull

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

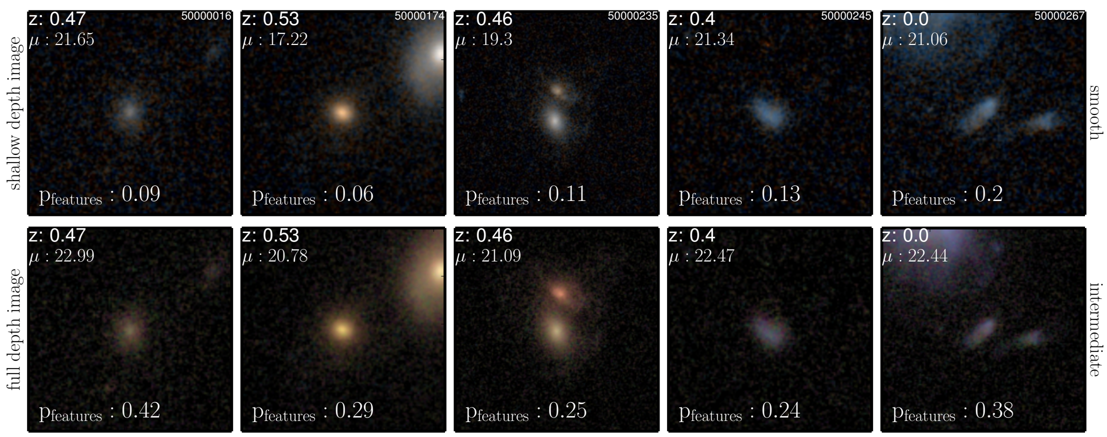
shallow to full morphology	N	%	$< \Delta p >$	$< z >$
smooth to smooth	758	17.0	-0.00	0.69
smooth to intermediate	367	8.2	0.18	0.69
smooth to featured	7	0.2	0.76	0.57
intermediate to smooth	214	4.8	-0.18	0.65
intermediate to intermediate	2,303	51.6	0.01	0.78
intermediate to featured	168	3.8	0.19	0.83
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-μ 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.

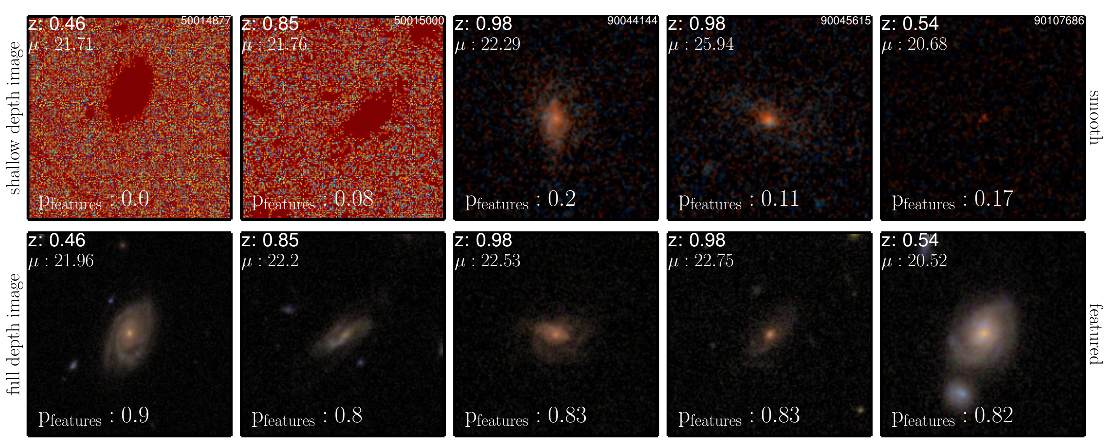
	N	%
Correctable	664	17%
Uncorrectable	483	12%
NEI	2,803	71%
Total	3,950	100%



[a]

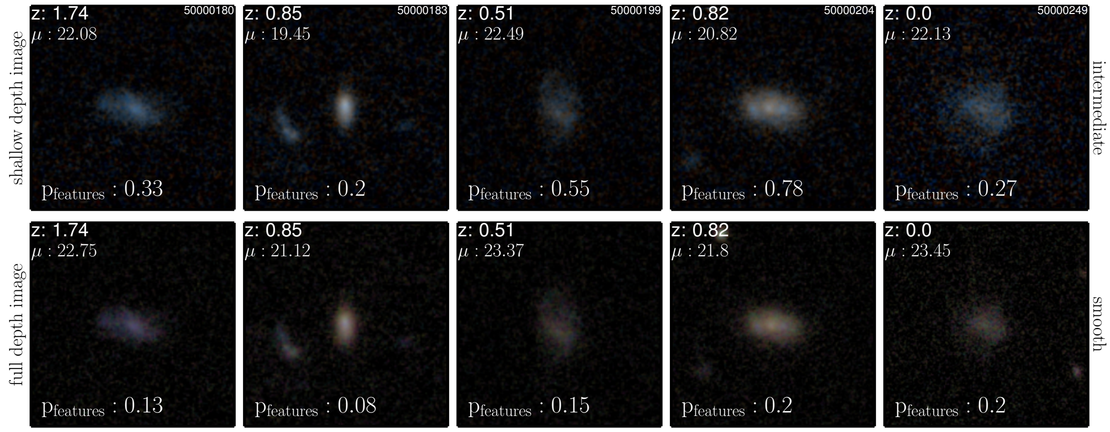


[b]

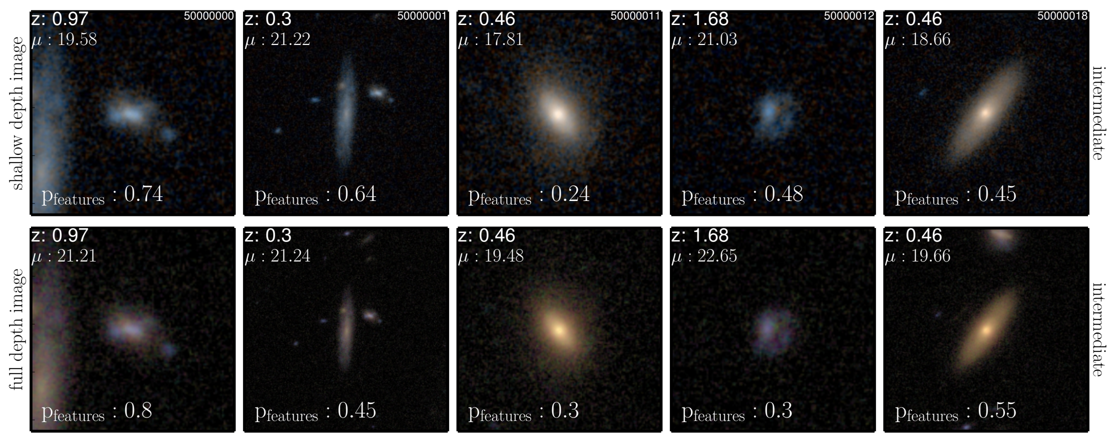


[c]

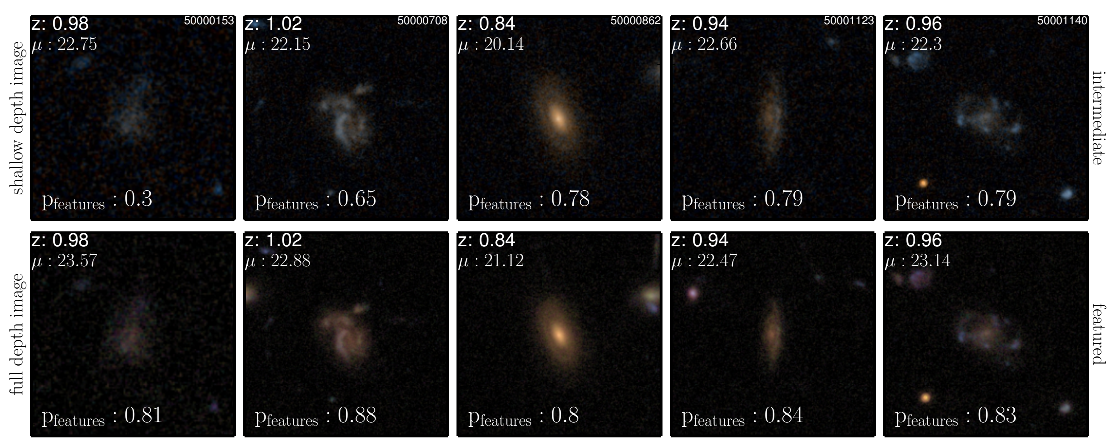
Figure A2. Galaxies whose shallow images were classified as smooth and full depth images were classified as smooth, intermediate, or featured.



[b]

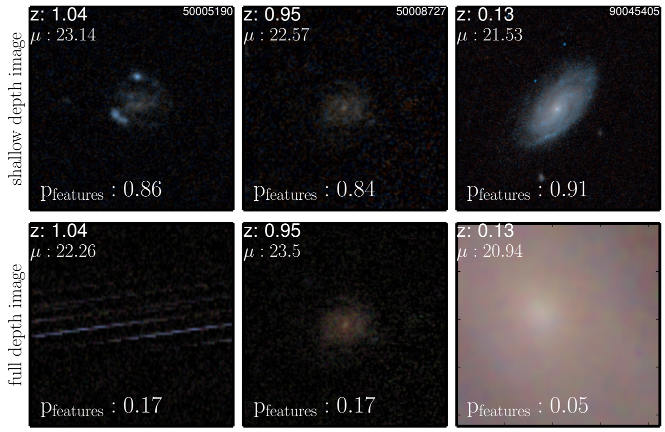


[b]

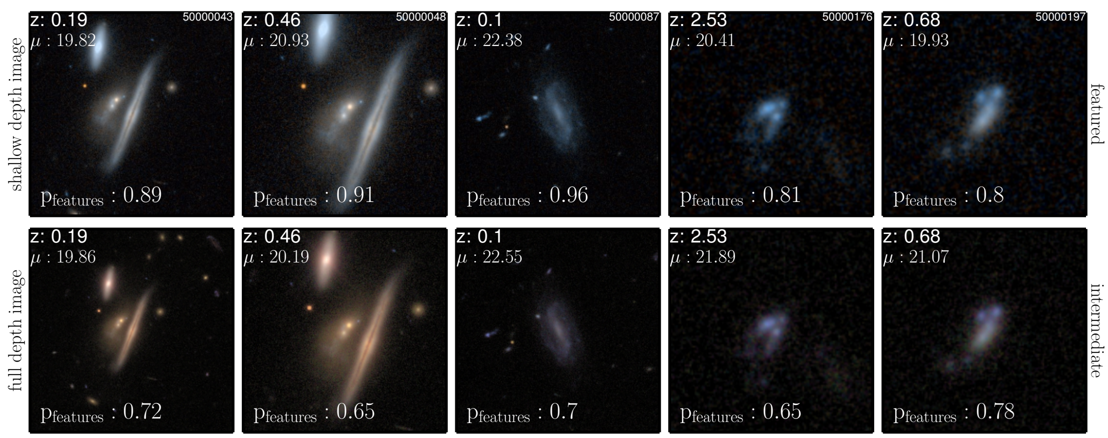


[b]

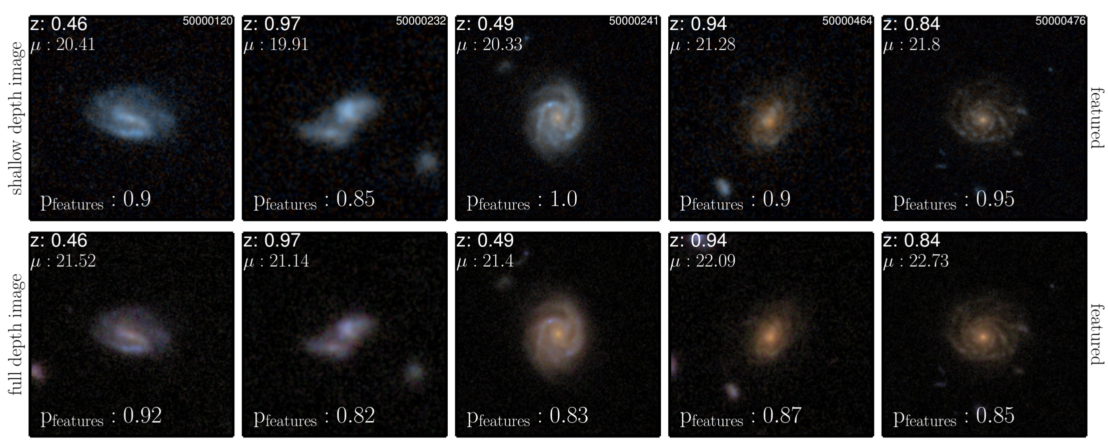
Figure A3. Galaxies whose shallow images were classified as intermediate and full depth images were classified as smooth, intermediate, or featured.



[b]



[b]



[b]

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

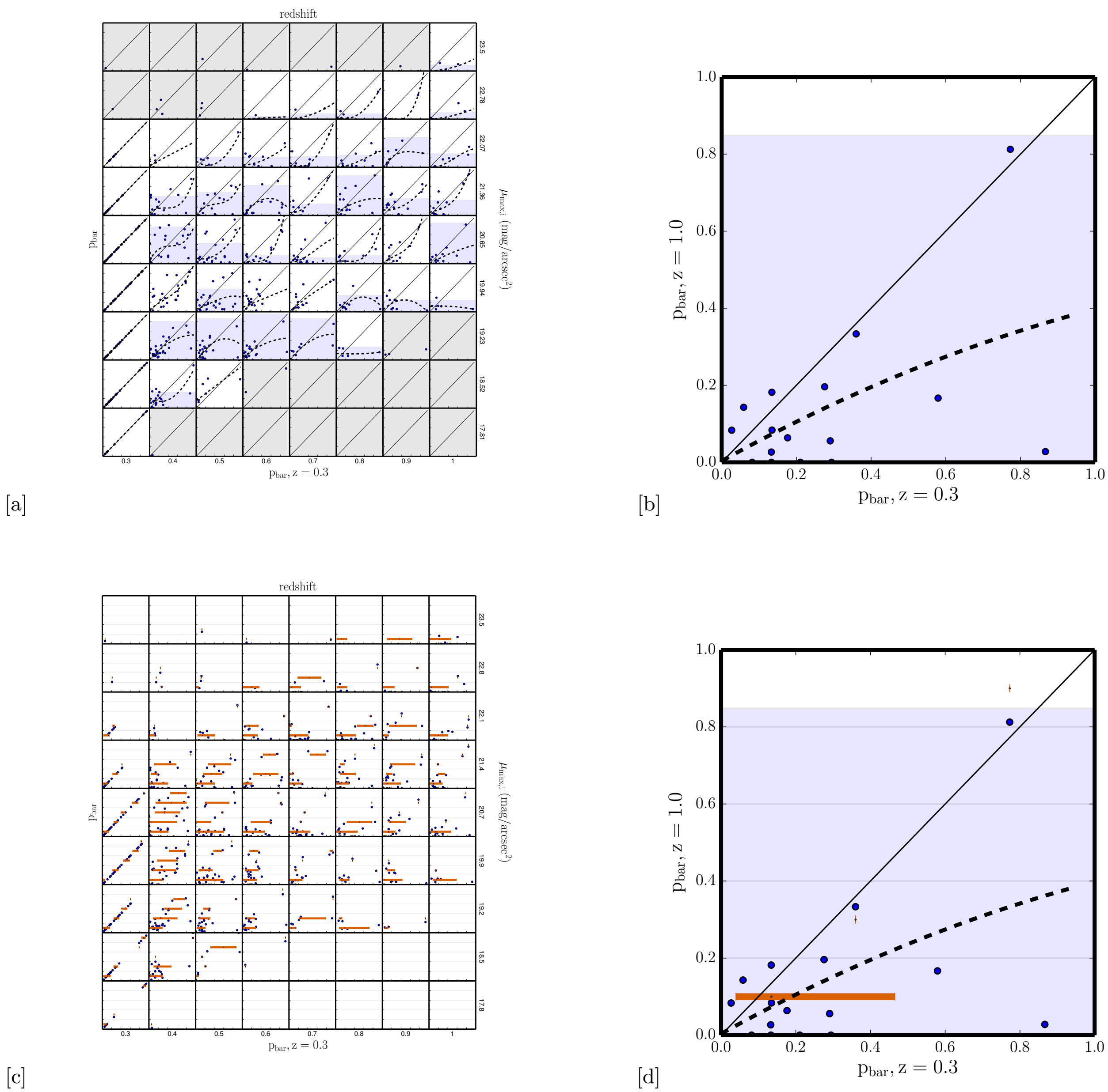


Figure A5. Same as Figure 9, but with the bar question.