Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **464,** 4176–4203 (2017)
Advance Access publication 2016 October 8

doi:10.1093/mnras/stw2568

Galaxy Zoo: morphological classifications for 120 000 galaxies in *HST* legacy imaging★

Kyle W. Willett,[1,2]† Melanie A. Galloway,[1] Steven P. Bamford,[3] Chris J. Lintott,[4] Karen L. Masters,[5,6] Claudia Scarlata,[1] B. D. Simmons,[4,7]‡ Melanie Beck,[1] Carolin N. Cardamone,[8] Edmond Cheung,[9] Edward M. Edmondson,[5] Lucy F. Fortson,[1] Roger L. Griffith,[10,11] Boris Häußler,[4,12,13] Anna Han,[14] Ross Hart,[3] Thomas Melvin,[5] Michael Parrish,[15] Kevin Schawinski,[16] R. J. Smethurst[4] and Arfon M. Smith[4,15,17]

Affiliations are listed at the end of the paper

Accepted 2016 October 5. Received 2016 October 3; in original form 2016 June 1; Editorial Decision 2016 October 4

ABSTRACT

We present the data release paper for the Galaxy Zoo: Hubble (GZH) project. This is the third phase in a large effort to measure reliable, detailed morphologies of galaxies by using crowdsourced visual classifications of colour-composite images. Images in GZH were selected from various publicly released *Hubble Space Telescope* legacy programmes conducted with the Advanced Camera for Surveys, with filters that probe the rest-frame optical emission from galaxies out to $z \sim 1$. The bulk of the sample is selected to have $m_{I814W} < 23.5$, but goes as faint as $m_{I814W} < 26.8$ for deep images combined over five epochs. The median redshift of the combined samples is $\langle z \rangle = 0.9 \pm 0.6$, with a tail extending out to $z \simeq 4$. The GZH morphological data include measurements of both bulge- and disc-dominated galaxies, details on spiral disc structure that relate to the Hubble type, bar identification, and numerous measurements of clump identification and geometry. This paper also describes a new method for calibrating morphologies for galaxies of different luminosities and at different redshifts by using artificially redshifted galaxy images as a baseline. The GZH catalogue contains both raw and calibrated morphological vote fractions for 119 849 galaxies, providing the largest data set to date suitable for large-scale studies of galaxy evolution out to $z \sim 1$.

Key words: methods: data analysis – catalogues – galaxies: evolution – galaxies: high-redshift – galaxies: structure.

1 INTRODUCTION

The morphology of galaxies encodes information on the orbital parameters and assembly history of their 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 tides, shocks in cluster environments, and direct mergers can all change the shape of the galaxy's gravitational potential. For $M^\star$ galaxies in the local Universe, this interplay between the physical development of a galaxy and its external appearance typically manifests at the most basic level as the difference between bulge-dominated systems with no/little spiral structure (early-types) and disc-dominated, rotationally supported galaxies (late-types) frequently exhibiting spiral arms. This dichotomy has been used to explore much of the astrophysics governing galaxy formation and evolution, and has been shown to be closely linked with other galactic properties such as stellar mass, halo mass, bolometric luminosity, black hole activity, effective radius, and the relative ages of the stellar populations.

The advent of larger telescopes sensitive to a full 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 mature, coherent structures seen in the present day. This growth occurs in a variety of ways, including accretion of baryons from large-scale galactic filaments on to haloes via streaming, mergers of individual dark matter haloes along with their baryons, conversion of gas into stars through gravitational

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

† E-mail: willett@physics.umn.edu
‡ Einstein Fellow.

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 the galaxy morphology in different ways, and so an accurate measurement of the demographics as a function of redshift provides an extremely powerful observational constraint on the physics involved (for recent reviews, see Buta 2013; Conselice 2014).

Theoretical predictions for the morphology of galaxies as a function of redshift are primarily computed within the Λ cold dark matter (ΛCDM) cosmological framework. Full treatments 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 span volumes up to $\sim$100 Mpc^3 while simultaneously resolving the smaller ($<$1 kpc) scales necessary to reproduce the influence of 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 their merger history (Toomre & Toomre 1972; Steinmetz & Navarro 2002; Hopkins et al. 2010; Kaviraj 2014a,b), local environment (e.g. the morphology–density relation; Dressler 1980), initial dark halo mass, secular evolution rate, and many other factors. Morphologies of individual simulated high-mass galaxies at $z \sim$ 2–3 commonly show kpc-scale 'clumpy' structures, with few galaxies that are either smooth or well-ordered spirals; asymmetric galaxies with strong density contrasts dominate simulated populations in the early Universe until at least $z \sim 1$ (Bell et al. 2012; 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 discs (akin to the ellipticals and spirals seen in the local Universe), but also a significant population of massive, more irregular galaxies, including mergers, tadpoles, chains, double-clumps, and clump-clusters (Elmegreen et al. 2005, 2007; Cameron et al. 2011; Förster Schreiber et al. 2011; Kartaltepe et al. 2015). In contrast, while grand-design spirals have been observed as far back as $z = 2.18$ (Law et al. 2012a), their spatial density suggests that they are exceedingly rare at these high redshifts, with a very low overall disc fraction (Mortlock et al. 2013). Current observational data thus strongly suggest 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* (*HST*), have been responsible for the bulk of imaging studies of high-redshift galaxies. Observations of fields with very deep imaging (e.g. Williams et al. 1996; Giavalisco et al. 2004; Beckwith et al. 2006; Davis et al. 2007; Scoville et al. 2007; Grogin et al. 2011; Koekemoer 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 beyond $z > 1$ in a flat ΛCDM cosmology, the relevant angular scales are only of the order of $\sim$5–10 kpc arcsec^{-1} (Wright 2006). *HST* can thus resolve much of the structure for a Milky Way-sized galaxy out to moderately high redshifts (at least distinguishing a disc from a bulge), but is limited for more compact structures. Since the size of galaxies evolves as roughly $r \propto (1+z)^{-1}$ (Mao, Mo & White 1998; Law et al. 2012b), the compact sizes of high-redshift galaxies make detailed morphologies a challenge even for *HST* (Chevance et al. 2012). However, the public availability of more than 10^5 galaxies in archival imaging has generated a sample with the potential for statistically robust studies of galaxy demographics and evolution.

One of the major difficulties in studying the morphologies of galaxies lies in the techniques used for measurement. Visual classification by experts has been used for many decades (e.g. 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 the advantage of 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 (e.g. see Lahav et al. 1995 for a discussion on the extent to which eight expert classifiers agree with each other). Automated measurements, both parametric (Peng et al. 2002; Simard et al. 2011; Lackner & Gunn 2012) and non-parametric (Abraham, van den Bergh & Nair 2003; Conselice 2003; Lotz, Primack & Madau 2004; Scarlata et al. 2007; Bamford et al. 2008; Freeman et al. 2013), scale well to very large 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; Fortson et al. 2012) utilizes crowdsourced visual classifications to measure galaxies in colour-composite images. The efforts of more than 200 000 classifiers allow 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 crowdsourced data require extensive calibration (Bamford et al. 2009; Willett et al. 2013), they have a proven reliability and have been used for a wide variety of scientific studies (e.g. 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; Smethurst et al. 2016).

This paper presents the classifications collected from the Galaxy Zoo: Hubble (GZH) project.[1] GZH was the third phase of Galaxy Zoo, following its initial results classifying $\sim$900 000 Sloan Digital Sky Survey (SDSS) images into primarily early/late types (Lintott et al. 2011) and Galaxy Zoo 2, which covered a subset of $\sim$250 000 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 *HST*. The Galaxy Zoo: CANDELS project has also classified morphologies of galaxies at high redshift, but using ACS and WFC3 rest-frame infrared imaging (Simmons et al. 2016).

The sample selection and creation of the images used for GZH is described in Section 2. Section 3 describes the GZH interface and the collection of classifications. Section 4 outlines the process used to calibrate and correct the crowdsourced vote fractions for redshift-dependent bias. Section 5 gives the main catalogue of results, 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 catalogues, with a summary in Section 8.

This paper assumes the *Wilkinson Microwave Anisotropy Probe* 9 cosmological parameters of $(\Omega_{\rm m}, \Omega_\Lambda, h) = (0.286, 0.714, 0.693)$ (Hinshaw et al. 2013).

[1] http://zoo3.galaxyzoo.org/

2 SAMPLE AND DATA

The GZH project contains images drawn from a number of different dedicated surveys and sample selection criteria. The majority of the data (as implied by the project name) were sourced directly from *HST* legacy surveys, all of which primarily used imaging from the Advanced Camera for Surveys (ACS). In addition to *HST* images, the project uses images from SDSS Stripe 82, as well as simulated *HST* images from multiple sources. Below we provide details on each source of imaging and describe the creation process for images shown to classifiers. We also detail the sources of metadata for galaxies in the sample, such as photometry and redshifts.

2.1 Hubble legacy surveys

Image information from multiple *HST* surveys was combined into a single photometric and morphological data base, the Advanced Camera for Surveys General Catalog (ACS-GC) by Griffith et al. (2012). A summary of the key parameters of the ACS-GC is given in Table 1.

The properties of the individual surveys are as follows.

(i) The All-wavelength Extended Groth strip International Survey (AEGIS; Davis et al. 2007) covers a strip centred at $\alpha = 14^{\rm h}17^{\rm m}$, $\delta = +52^\circ30'$. This area of the sky was selected for a deep survey due to a combination of low extinction and low Galactic/zodiacal emission. The ACS images covered 63 separate tiles over a total area of $\sim$710 arcmin^2. The two ACS bands for AEGIS had exposure times of 2300 s in *F*606*W* (V_{606W}) and 2100 s in *F*814*W* (I_{814W}). The final mosaic images were dithered to a resolution of 0.03 arcsec pixel^{-1}. For extended objects, the limiting magnitude of sources was 26.23 (AB) in V_{606W} and 25.61 (AB) in I_{814W}.

(ii) The Great Observatories Origins Deep Survey (GOODS; Giavalisco et al. 2004) covered two separate fields in the Northern and Southern Hemispheres: the *Hubble Deep Field*-North ($\alpha = 12^{\rm h}36^{\rm m}$, $\delta = +62^\circ14'$) and the *Chandra Deep Field*-South ($\alpha = 03^{\rm h}32^{\rm m}$, $\delta = -27^\circ48'$). The *HST* ACS imaging data from the two fields are referred to as GOODS-N and GOODS-S, respectively. ACS imaging in GOODS fields used four filters – *F*435*W* (B_{435W}), V_{606W}, *F*775*W* (i_{775W}), and *F*850*LP* (z_{850LP}). The mean exposure times for each epoch varied by band, from 1050 to 2100 s. The B_{435W} images were completed in a single epoch at the beginning of the survey, but the V_{606W}, i_{775W}, and z_{850LP} images were taken in five separate epochs separated by 40–50 d each. GZH includes co-added images from GOODS at both two-epoch and five-epoch depths. Images were dithered to a pixel scale of 0.03 arcsec pixel^{-1} and covered a total area of $\sim$320 arcmin^2 (160 arcmin^2 each for the north and south fields). The 5σ limiting magnitude for extended sources was 25.7 for V_{606W} and 25.0 for i_{775W}.

(iii) The Galaxy Evolution from Morphologies and SEDs (GEMS; Rix et al. 2004; Caldwell et al. 2008) survey was centred on the *Chandra Deep Field*-South. The GEMS data covered $\sim$800 arcmin^2, completely surrounding the area covered by GOODS-S. Images from ACS in GEMS had 1 orbit per pointing for a total of 63 pointings. The exposure times were 2160 and 2286 s in V_{606W} and z_{850LP}, respectively. The image resolution had a pixel scale of 0.03 arcsec pixel^{-1}. The 5σ limiting magnitude for source detection was 25.7 AB in V_{606W} and 24.2 AB in z_{850LP}.

(iv) The Cosmic Evolution Survey (COSMOS; Koekemoer et al. 2007; Scoville et al. 2007) covered an area of $\sim$1.8 deg^2 centred at $\alpha = 10^{\rm h}00^{\rm m}$, $\delta = +02^\circ12'$. Its location near the celestial equator was designed to enable coverage by ground-based telescopes in both the Northern and Southern Hemispheres, in addition to space-based observatories. The ACS data for COSMOS consisted of 1 orbit per pointing with an exposure time of 2028 s in I_{814W}; 590 total pointings were used to cover the entire field. The image resolution was dithered to 0.05 arcsec pixel^{-1}. The 50 per cent completeness magnitude for a galaxy with a half-light radius of 0.50 arcsec in I_{814W} was 24.7 mag.

In the ACS-GC, individual galaxies were identified using a combination of SExtractor (Bertin & Arnouts 1996) and the galaxy-profile fitting framework Galapagos (Barden et al. 2012). GZH included all galaxies with $m < 23.5$, where m is in the I_{814W}, z_{850LP}, or i_{775W} for the AEGIS and COSMOS, GEMS and GOODS-S (two-epoch), and GOODS-N (two-epoch) surveys, respectively. The full-depth GOODS images from both fields included galaxies with $m <$ 26.8. This yielded a total of 119 849 images (Table 1).

Images from *HST* legacy surveys were used to create multiple different colour and grey-scale images using different depths and filter combinations; these are described further in Section 2.4. Additionally, the use of a small set of *HST* images to create simulated images of active galactic nucleus (AGN) host galaxies is described in Section 2.3.1.

2.2 The Sloan Digital Sky Survey

The GZH project also includes images from the SDSS (York et al. 2000; Strauss et al. 2002), in particular those from Data Release 7 (Abazajian et al. 2009). A small number of images were used in the creation of simulated *HST* images, described in Sections 2.3.2 and 4 below. The majority of SDSS images used in GZH were from Stripe 82, including both single-epoch and co-added images. The single-epoch images provide a local sample for comparison to higher redshift galaxies (including the new measurements of clumpy structure), while the co-added images allow for analysis of morphological properties as a function of image depth.

Table 1. Summary of GZH *HST* imaging.

Survey	Total $t_{\rm exp}$ (s)	Filters	Resolution (arcsec pixel^{-1})	Area (arcmin^2)	$N_{\rm galaxies}$
AEGIS	2100–2300	V_{606W}, I_{814W}	0.03	710	8507
COSMOS	2028	I_{814W}	0.05	6480	84 954
GEMS	2160–2286	V_{606W}, z_{850LP}	0.03	800	9087
GOODS	–	–	–	–	–
GOODS-N two-epoch	2100–4200	V_{606W}, i_{775W}	0.03	320	2551
GOODS-S two-epoch	2100–4200	V_{606W}, z_{850LP}	0.03	320	3593
GOODS-N five-epoch	5100–10 500	B_{435W}, V_{606W}, i_{775W}, z_{850LP}	0.03	〃	6015
GOODS-S five-epoch	5100–10 500	B_{435W}, V_{606W}, i_{775W}, z_{850LP}	0.03	〃	5142
All *HST* surveys	–	–	–	8630	119 849

Single-epoch images from SDSS Stripe 82 were selected using the criteria from Willett et al. (2013), which required limits of `petroR90_r`>3 arcsec (where `petroR90_r` is the radius containing 90 per cent of the *r*-band Petrosian flux) and a magnitude brighter than $m_r < 17.77$. 21 522 galaxies in SDSS met these criteria. Co-added images from Stripe 82 were selected from the union of galaxies with co-added magnitudes brighter than 17.77 mag, and the galaxies detected in the `stripe-82-single` images and matched to a co-add source. This resulted in a total set of 30 339 images. Of the images in the co-added sample, 5144 (17 per cent) were dimmer than the initial cut of 17.77 mag.

2.3 Simulated *HST* images

To facilitate correction of classifications in the presence of known redshift bias (Section 4), GZH includes two different samples of simulated *HST* images: real *HST* galaxy images (Section 2.1) with nuclear emission added to simulate AGN host galaxies, and lower redshift galaxy images from SDSS (Section 2.2), artificially redshifted from $0.3 \leq z \leq 1$. Each is described separately below.

2.3.1 Images with simulated nuclear point sources

As AGN often have bright, unresolved optical emission, AGN have the potential to mimic or distort the identification of a bulge component. GZH thus includes a set of images designed to measure the effect of AGN on morphological classifications. The presence of an AGN was simulated by modelling the point spread function (PSF) of the telescope and then inserting a bright source near the centre of a real galaxy. For each image, the simulated AGN was assigned one of three colours – either blue, red, or flat (white) as seen in the colour images – and a range of brightnesses such that $L_{\rm ratio} \equiv L_{\rm galaxy}/L_{\rm AGN}$ is in (0.2, 1.0, 5.0, 10.0, 50.0). Combining these parameters generated 15 images with different simulated AGN for each host, in addition to the original galaxy image.

Two sets of simulated AGN were generated in GZH. The first set (version 1) was assembled from 95 galaxies from GOODS-S imaging and empirical PSFs made by combining stars in the GOODS fields using the PSF creation tools in DAOPHOT (Stetson 1987). The second set (version 2) was assembled from 96 galaxies in GOODS-S; this version used simulated PSFs from TINYTIM (Krist 1993), drizzled using the same procedures as those used in the reduction of the GOODS-S images (Koekemoer et al. 2002; Koekemoer, Fruchter & Hack 2003; Giavalisco et al. 2004). The use of these two versions facilitates comparisons between these different PSF creation methods, which are widely used in AGN host galaxy morphology studies (e.g. Sánchez et al. 2004; Simmons & Urry 2008; Pierce et al. 2010; Simmons et al. 2011). Each PSF creation method has advantages and disadvantages: the empirical PSFs better represent the nuances of the PSF in the specific data being used and look more realistic at lower luminosities, but the extended features of the noiseless TINYTIM PSFs are visually more realistic at higher luminosities.

Images with simulated AGN were classified in the interface in an identical manner and were evenly distributed with unaltered images of the galaxies. Volunteers were not explicitly told that the images had been altered during classification, as the goal was to measure the effect on normal classifications using the same technique as closely as possible. Following classification, a classifier could view a page with additional details about each galaxy; where applicable, these pages contained further information regarding image modifications. Simmons et al. (2013) used the simulated AGN host galaxy images for more reliable identification of bulgeless galaxies hosting AGN in SDSS data.

2.3.2 Generating images of artificially redshifted galaxies

The dimming and resolution effects of redshift can significantly affect galaxy classifications derived from any method. To facilitate corrections due to redshift, we include a sample of 288 galaxies with SDSS imaging that can be transformed to simulate *HST* imaging out to $z = 1$. The SDSS images were redshifted and processed to mimic simulated *HST* imaging parameters using the FERENGI code (Barden, Jahnke & Häußler 2008).

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 parametrized by a crude model that mimics the brightness increase of galaxies with increasing redshift (e.g. Lilly et al. 1998; Loveday et al. 2012). The effect on the redshifted images is simply an empirical addition to the magnitude of a galaxy of the form $M' = e \times z + M$, where M' is the corrected magnitude and e is the evolutionary correction in magnitudes ($e = -1$ essentially brightens the entire galaxy by 1 mag by $z = 1$). FERENGI was run on the images for values of e starting from $e = 0$ and decreasing to $e = -3.5$ in increments of $\Delta e = 0.5$. Fig. 1 shows several examples of the effects of 'losing' spiral/disc features with increasing redshift for two galaxies with no evolution corrections ($e = 0$); as the signal-to-noise ratio in the images decreases and the galaxies become fainter, the contrast between features like spiral arms goes down and $f_{\rm features}$ drops to the point where both galaxies would have been classified as likely ellipticals.

The final number of FERENGI images produced for each galaxy is ultimately a function of the 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. The use of these images to correct for redshift-dependent bias is described in Section 4.

2.4 Creating colour images

The images used for classification in GZH were colour-composite JPEGs made from multi-band data. These were created following the method of Lupton et al. (2004), which preserves colour information irrespective of intensity. An asinh intensity mapping was applied to enhance the appearance of faint features while avoiding saturating galaxy centres. The relative scalings of the filter bands were chosen to reproduce the colour appearance of the SDSS images in previous iterations of Galaxy Zoo.

Many of the legacy surveys described in Section 2.1 provided *HST* images in only two filters. For these, the shorter wavelength band was mapped to the blue channel, the longer wavelength band to the red channel, and the green channel created by taking the arithmetic mean of the red and blue. The bands used in each of the surveys are listed in Table 1. Although four bands were available for the GOODS survey, only two bands were used to create the original two-epoch images, for consistency with AEGIS and GEMS. The two-epoch GOODS-N and GOODS-S images were created using differing filters – this was a deliberate choice made so that the GEMS images could be directly compared with the overlapping coverage of GOODS-S (Fig. 2).

Only two-epoch GOODS images were included at the launch of GZH. Deeper, five-epoch GOODS images were added into GZH in 2015 March. The deeper images made use of the full four-band data by using the arithmetic mean of B_{435W} and V_{606W} in

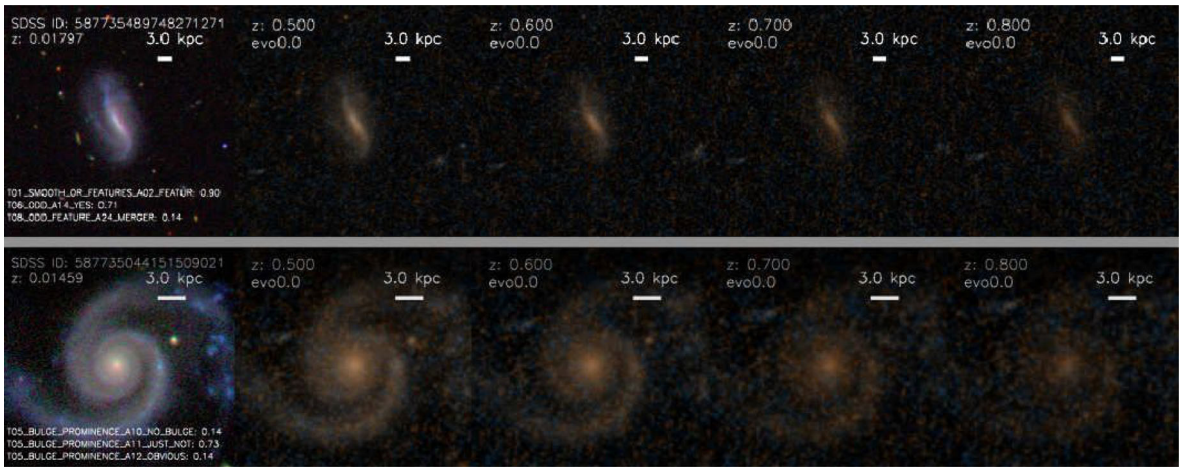

Figure 1. Examples of two galaxies which have been run through the FERENGI code to produce simulated *HST* images. The measured values of $f_{\rm features}$ from GZH for the images in each panel are (top row) $f_{\rm features}$= (0.900, 0.625, 0.350, 0.350, 0.225) and (bottom row) $f_{\rm features}$= (1.000, 0.875, 0.875, 0.625, 0.375).

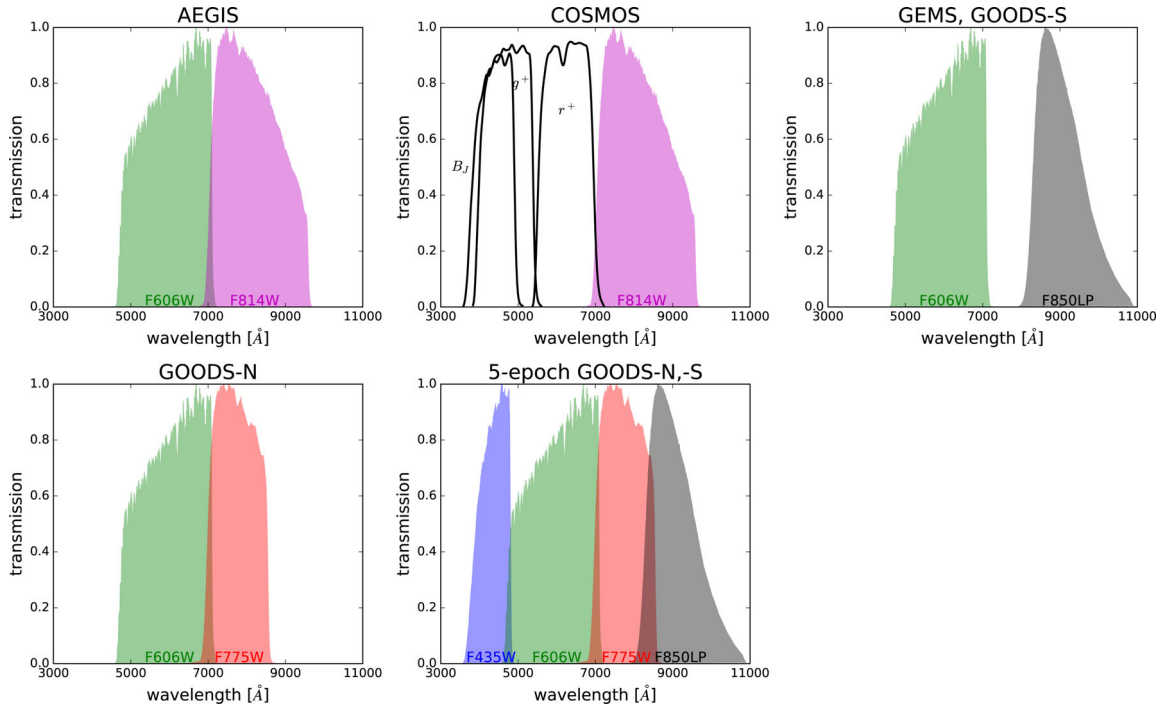

Figure 2. Transmission curves of the filters used by *HST* ACS in wide-field channel mode for the various surveys in GZH. The solid black curves show the filters for the Suprime Camera on Subaru, which was used to create colour gradients in the GZH COSMOS images.

the blue channel, I_{814W} in the green channel, and z_{850LP} in the red channel.

The COSMOS survey provides only I_{814W} *HST* imaging. For these galaxies, GZH used 'pseudo-colour' images created by using the ACS I_{814W} data as an illumination map and ground-based imaging from the Subaru telescope in B_J, r^+, and i^+ filters to provide colour information (see Griffith et al. 2012 for further details). This resulted in images with the angular resolution of *HST* (∼0.05 arcsec pixel^{-1}) for the overall intensity, but colour gradients at ground-based resolution, with seeing between 0.95 and 1.05 arcsec (Taniguchi et al. 2007).

Stripe 82 single-epoch images were taken directly from the DR7 SDSS SkyServer, which combined g', r', and i' exposures into the RGB channels (Nieto-Santisteban, Szalay & Gray 2004). The co-added Stripe 82 images were assembled from runs 106 and 206 in DR7 and processed into colour composites in the same manner as previous iterations of Galaxy Zoo.

In some cases, we found that attempting to emphasize faint features in the images resulted in the sky noise taking the appearance of brightly coloured speckles. This impaired the aesthetics of the images and was considered a potential distraction to visual classification. To counteract this, a soft-edged object mask was applied to the colour images and a desaturation operation performed. This masking procedure was effective in preserving the colour balance for galaxies and retained the visibility of faint features, while reducing the colour contrast in the sky noise and greatly improving the appearance of the images. This solution was applied to the co-added Stripe 82, COSMOS, and five-epoch GOODS images.

In addition to the primary *HST* legacy imaging sample with coloured images described as above, GZH also includes samples of *HST* galaxy images with different colour prescriptions. In particular, there are two sets of 3927 images each, drawn from the COSMOS sample; the first has a dramatically reduced colour saturation, and the second has reversed the colours so that the blue and red filters have exchanged places in the RGB image. The 'faded' galaxy set facilitates measuring of possible variations in classification due to the presence or absence of colour features. The 'recoloured' set enables an alternative test of potential colour biases in classifications.

The simulated AGN host galaxy colour images (described in Section 2.3.1) were created using the same prescription as for the GOODS two-epoch imaging. The artificially redshifted colour images (Section 2.3.2) were created using the same prescription as for the AEGIS images.

2.5 Galaxy sample labels

The full GZH sample is composed of eight different galaxy subsamples. Throughout this paper and in the published catalogues, they are referred to with the following labels.

(i) `main`: *HST* imaging with RGB colours making use of all available filters with typical saturation and in correct order, as described in Section 2.4. This sample includes AEGIS, COSMOS, GEMS, and full-depth GOODS (North and South) images (113 705 galaxies).

(ii) `faded`: a subset of COSMOS images with very low colour saturation (3927 galaxies).

(iii) `recoloured`: a subset of COSMOS images with red and blue channels reversed. Note that this subsample uses the same galaxies as the `faded` subsample (3927 galaxies).

(iv) `goods-shallow`: images from GOODS-North and GOODS-South, with colour from two filters and imaged at two-epoch depth (6144 galaxies).

(v) `stripe-82-single`: single-epoch images from SDSS Stripe 82 (21 522 galaxies).

(vi) `stripe-82-coadd`: co-added images from SDSS Stripe 82. Note that this subsample includes all galaxies in the `stripe-82-single` subsample, with additional sources detected in the deeper imaging (30 339 galaxies).

(vii) `redshifted`: simulated *HST* images constructed using SDSS images and artificially processed to redshifts between $0.3 \leq z \leq 1$ (288 original galaxies; 6624 redshifted images).

(viii) `simulated-agn`: simulated AGN host galaxies constructed using *HST* images and PSFs (96 original galaxies; 2961 images of simulated AGN hosts).

2.6 Galaxy metadata

Photometric data for the bulk of the GZH `main`, `faded`, `recoloured`, and `goods-shallow` samples were largely drawn from the tables in Griffith et al. (2012). This included photometric parameters such as the fluxes, magnitudes, radii, ellipticities, position angles, and positions drawn from both SExtractor and GALFIT. All photometric parameters were measured in both bands of the ACS imaging, with the exception of the single-band COSMOS images. Photometric data for the GOODS five-epoch imaging, including SExtractor parameters, are from Giavalisco (2012).

Redshifts for the GZH catalogue were compiled from a variety of sources. For each galaxy, the primary redshift is in the `Z_BEST` column of Table 2. The redshift type (spectroscopic: `SPEC_Z`, photometric: `PHOTO_Z`, or grism: `GRISM_Z`) is listed in the column `Z_BEST_TYPE`, and the source catalogue of the redshift is included as `Z_BEST_SOURCE`.

For galaxies which have published redshifts from multiple sources, the following algorithm was used to select the `Z_BEST` quantity. A high-quality spectroscopic redshift in the ACS-GC is the primary option, provided in the ACS-GC (Griffith et al. 2012), 3DHST (Momcheva et al. 2016), and MUSYC (Cardamone et al. 2010) catalogues and used in that order. For galaxies with multiple spectroscopic redshifts, more than 98 per cent are consistent ($\Delta z < 0.001$), and so the order of selection made no practical difference.

Table 2. GZH morphological classifications for *HST* images from AEGIS, COSMOS, GEMS, and GOODS.

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_						...
Zooniverse ID	Survey ID	Imaging	Correction[a]	N_{votes}	Fraction	Weighted	Debiased	Best	Lower limit	Upper limit	
AHZ100002g	10010842	AEGIS	0	127	0.118	0.128	0.089	0.089	0.032	0.106	
AHZ100002h	10010870	AEGIS	4	127	0.567	0.592	-0.126	0.592	–	–	
...											
AHZ20004kd	20014731	COSMOS	0	44	0.682	0.675	0.243	0.330	0.615	0.243	
AHZ20004ke	20014732	COSMOS	2	45	0.689	0.756	0.844	0.756	–	–	
...											
AHZ400043g	90022729	GEMS	3	121	0.702	0.733	0.486	0.733	–	–	
AHZ4000416	90022735	GEMS	0	127	0.646	0.698	0.509	0.509	0.347	0.394	
...											
AGZ0007z47	10014	GOODS-N-FULLDEPTH	1	40	0.475	0.475	0.203	0.475	0.126	0.475	
AGZ0007z48	10017	GOODS-N-FULLDEPTH	3	40	0.675	0.675	0.094	0.675	–	–	
...											
AGZ00083jb	8869	GOODS-S-FULLDEPTH	0	40	0.425	0.425	0.135	0.135	0.185	0.481	
AGZ00083jc	8878	GOODS-S-FULLDEPTH	0	40	0.205	0.205	0.050	0.050	−0.028	0.181	
...											

Notes. [a]Flag indicating how the vote fractions for this galaxy were corrected through debiasing (Section 4.1), if possible. 0 = correctable, 1 = lower limit ($f_{raw} - f_{adj}$ is not single-valued), 2 = uncorrected ($z < 0.3$), 3 = uncorrected (insufficient FERENGI galaxies in this z–μ bin), 4 = uncorrected (no galaxy redshift available).

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 113 705 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

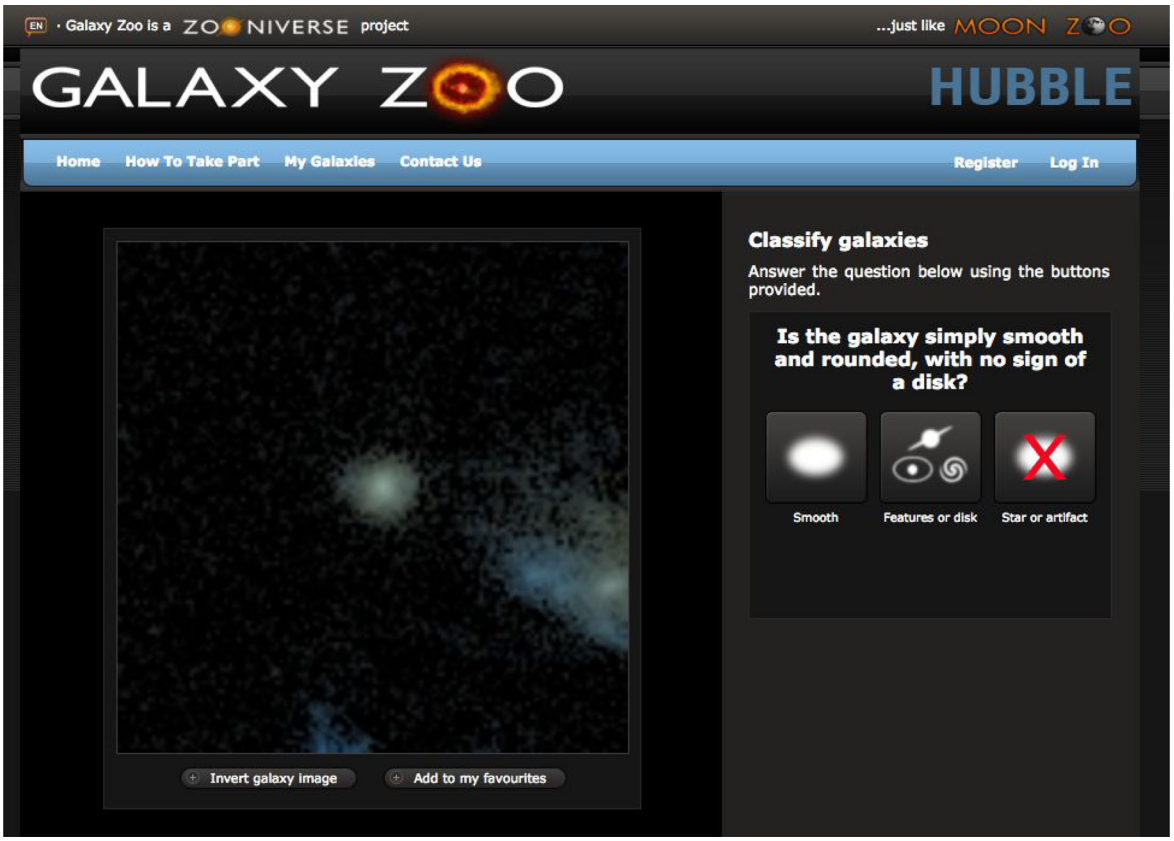

Figure 3. Screenshot of the GZH interface at the beginning of classifying a random galaxy, with the classifier ready to select an answer for the first question in the decision tree.

Galaxies with inconsistent spectroscopic redshifts between any pair of catalogues are marked with a flag in Table 2. If no spectroscopic redshifts were available, the 1σ errors of the photometric (ACS-GC, 3DHST, MUSYC, UltraVISTA; Ilbert et al. 2013) and UltraVISTA grism data were used. The measurement with the smallest reported 1σ error was selected in each case.

Photometric and spectroscopic data for the `stripe-82-single` and `stripe-82-coadd` galaxies were taken from the CasJobs DR7 tables. This included *ugriz* Petrosian magnitudes and fluxes, as well as the relative de Vaucouleurs and exponential fits from the model magnitudes. All redshifts used for SDSS galaxies were spectroscopic. 82.6 per cent of galaxies in the `stripe-82-single` images and 65.1 per cent of galaxies in the `stripe-82-coadd` images had a measured DR7 spectroscopic redshift.

The technique for redshift debiasing (Section 4) requires consistent measurements of the galaxy surface brightness. For both the `redshifted` calibration images and the *HST* images which have their morphologies corrected, we calculate the mean surface brightness μ within the effective radius (R_e) as

$$\mu = m + 2.5 * \log_{10}(2 \times (b/a) \times \pi R_e^2). \quad (1)$$

All photometric parameters are taken from SExtractor. For the FERENGI galaxies, m is `MAG_AUTO` in the I_{814W} band, (b/a) is the galaxy ellipticity (the profile rms along the semi-major and -minor axes), and R_e is the 50 per cent `FLUX_RADIUS` converted into arcsec (Melvin 2016). For the *HST* galaxies in the `main` sample, the parameters are identical except that we use `MAG_BEST` instead of `MAG_AUTO` (Griffith et al. 2012) in either the I_{814W}, i_{775W}, or z_{850LP} bands, depending on the available imaging (Table 1).

3 GZH INTERFACE AND CLASSIFICATIONS

Below we describe the classification structure of GZH, including the software interface and the hierarchical structure of a classification. Section 3.2 describes the process of combining individual classifications into vote fractions for each galaxy.

3.1 Interface and decision tree

Classifications for GZH were made using a web-based interface (Fig. 3), similar in design to Galaxy Zoo and Galaxy Zoo 2. The front-end runs on a Ruby on Rails framework with classifications stored in a MySQL backend. Classifiers were shown a randomly selected colour-composite image from the GZH sample; the default showed the image with a black sky background, although they had the option to invert the colour palette if desired. The questions and responses for morphology appeared on the right side of the image as a panel, including both text and icons. There was no tutorial required for participation, although classifiers could access an extensive 'Help' section containing example images and descriptive text for all the morphological labels.

The procedure for classifying an image in GZH followed a hierarchical decision tree (Fig. 4). Every classification began with the step of identifying whether the object at the centre of the image was a 'smooth' galaxy, a galaxy with a disc or other features, or a star/artefact. Subsequent questions in the tree depended on the previous answer(s) given by the classifier; the decision tree was designed so that every question relevant to the morphology in the process of being identified was answered. Questions that were not

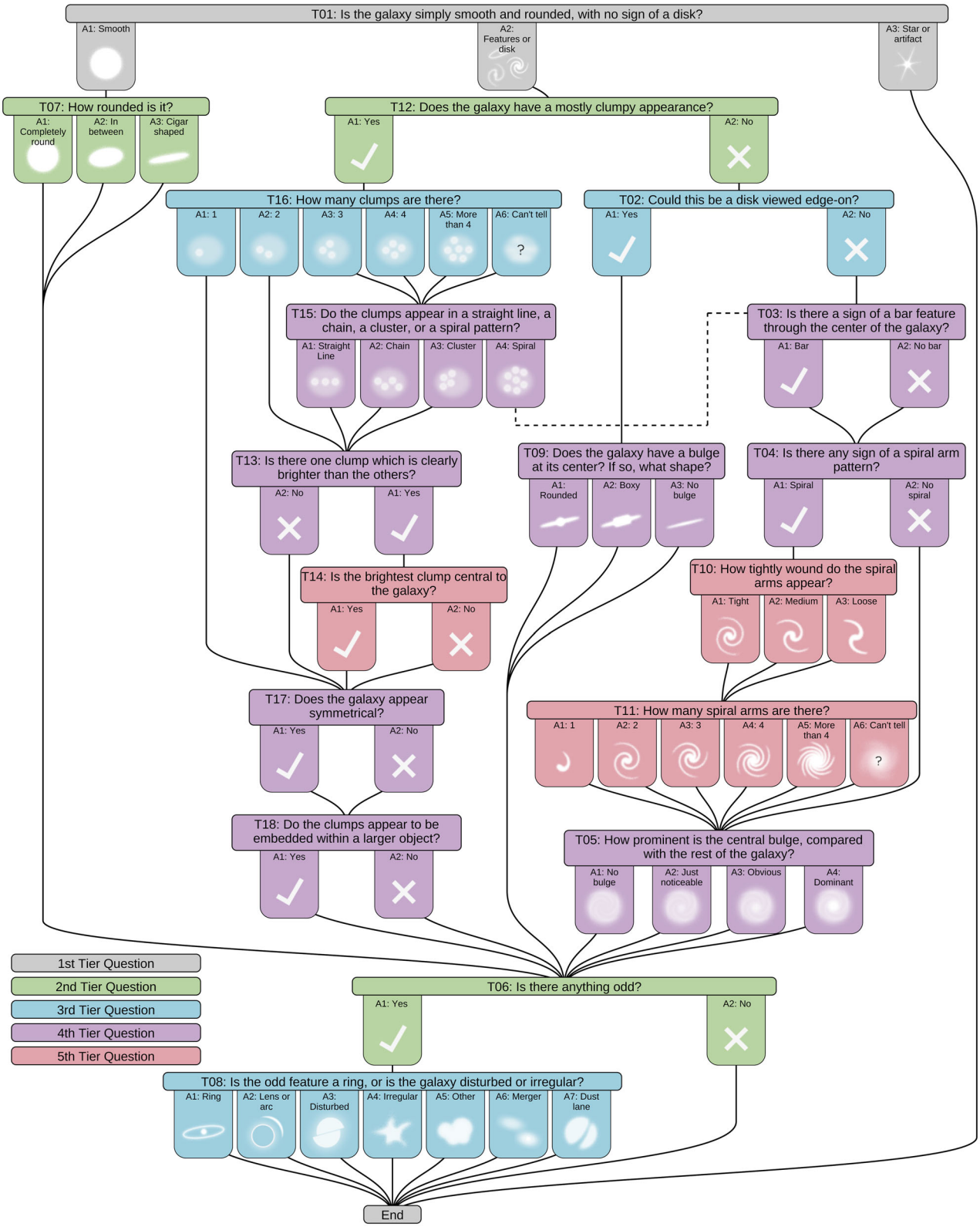

Figure 4. Flowchart of the questions presented to GZH classifiers, labelled with the corresponding task numbers. Tasks in the decision tree are colour-coded by tier level. Grey-coloured tasks are first-tier questions which are asked in every classification. Tasks coloured green, blue, purple, and pink (respectively) are one, two, three, or four steps below branching points in the decision tree. The dashed line between T15 and T03 indicates the unique case where volunteers could label both clumpy and disc-like morphologies for galaxies with clumps arranged in a spiral pattern.

answered were implicitly assumed to be absent in the image – for example, if the classifier identified a galaxy as being smooth, they were not asked to count the number of spiral arms. For every task, the classifier chose a single answer before continuing to the next question; they also had the option to restart any classification in progress.

The GZH decision tree was designed to be similar to that used by GZ2 while taking account of the likely differences in the morphologies of high-redshift galaxies. There are four broad sets of morphologies classified by the users in GZH. The first set identified stars or image artefacts (the result of either bad data or incorrect identification of an object as a galaxy by the ACS pipeline); in this case, the classification process ended and no further questions were asked. The second set was for 'smooth' galaxies, intended to select ellipticals/early-types; volunteers also indicate the relative axial ratio (roundness) for these galaxies. The third set was for disc/late-type galaxies, which labelled the features necessary to place a galaxy on the standard Hubble tuning fork (bars, spiral arm, strength of the central bulge). The final set, which was new in this phase of Galaxy Zoo and designed for high-redshift targets, identified objects dominated by clumpy morphologies. Further annotations for clumpy galaxies included assessing the number, arrangement, relative brightness, and location of the clumps within the galaxy. Finally, every classification had the option of identifying 'odd' features within the image; these labels were for relatively rare ($\lesssim$1 per cent) phenomena, including dust lanes, gravitational lenses, and mergers.

The number of independent classifications per subject collected by GZH was on average higher than GZ1 or GZ2, due to both the increased complexity of the decision tree and the relative difficulty of classifying images of small and distant galaxies. Images from the `main` AEGIS, GEMS, and GOODS data sets had a median of 122 independent classifications per image. The remaining images had fewer classifications either due to a later activation date (`main` COSMOS, `simulated-agn`) or a lower retirement limit (`stripe-82-single` and `stripe-82-coadd`). Images from these samples had a median of 46–48 classifications per image (Fig. 5).

The GZH project was launched on 2010 April 23 with the inclusion of the AEGIS, GEMS, GOODS two-epoch, and SDSS Stripe 82 images. Images from COSMOS and the simulated AGN were activated in 2010 December, as well as a small sample of images from AEGIS, GEMS, and GOODS that were previously excluded from the original sample due to cuts on blended and/or saturated objects and subsequently confirmed as classifiable galaxies. The GZH site collected data until its replacement, the fourth phase of Galaxy Zoo[2] (including data from both the *HST* CANDELS survey and SDSS DR8), began on 2012 September 10. Classifications for the GOODS five-epoch images were separately obtained from 2015 March–June using a new version of the Galaxy Zoo site but the same GZH decision tree. The GZH project finished with a total of 10 349 357 classifications from 93 898 registered participants.

3.2 Classifier weighting

As a first step to producing a consensus measurement of each galaxy, the votes of individual volunteers who classified galaxies in GZH were combined to make a vote fraction for each response ($f_{\rm response}$) to a question in the decision tree. Votes were subsequently weighted and re-combined in an iterative method similar to that in previous versions of Galaxy Zoo (Land et al. 2008; Willett et al. 2013), using a method chosen to be as egalitarian as possible while also identifying and downweighting classifiers who frequently disagreed with others. The weighting factor w was 1 for the top 95 per cent of classifiers as ranked by consistency. For the bottom 5 per cent of classifiers, w drops smoothly and is effectively zero for the bottom 1 per cent of the distribution function. Since downweighting only occurs for the bottom few per cent of classifiers (and an even smaller percentage of the classifications), the overall effect on the GZH data set was minimal. The method was effective, however, at filtering out contributions from randomly classifying or malicious participants.

Classifications for GZH were weighted only if the classifier was logged into the site under their username (this was encouraged, but not required for participation). Classifications by participants who were not logged in were marked as 'Anonymous' and receive the same $w = 1$ weighting as the vast majority of logged-in classifiers. Results from the Galaxy Zoo: CANDELS project (Simmons et al. 2016) show that the distributions of weights for anonymous and logged-in users are similar, supporting the default technique of GZH.

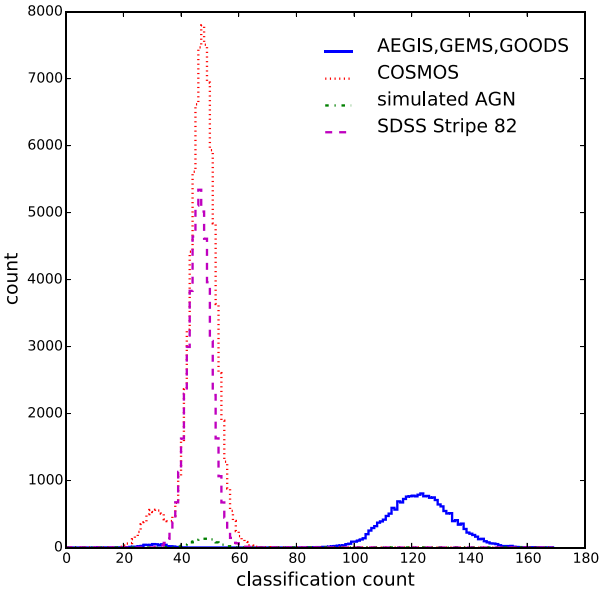

Figure 5. Distribution of the total number of classifications per image for GZH, split by survey.

4 CORRECTING FOR REDSHIFT-DEPENDENT CLASSIFICATION BIAS

The previous versions of Galaxy Zoo morphology classifications (Lintott et al. 2011; Willett et al. 2013) were based on observations of galaxies in the SDSS, which have a median redshift of $z < 0.2$. In these cases, it was assumed that there was no cosmological evolution of the morphologies of galaxies and therefore any observed changes in the morphological distribution were due to a redshift-dependent bias that affects image quality and classification accuracy (i.e. galaxies of a given mass/size will appear smaller and dimmer at higher redshifts). This bias is not unique to crowdsourcing techniques; its dependence on data quality is a potential problem for *both* automated and visual classifications, and must be addressed

[2] http://zoo4.galaxyzoo.org

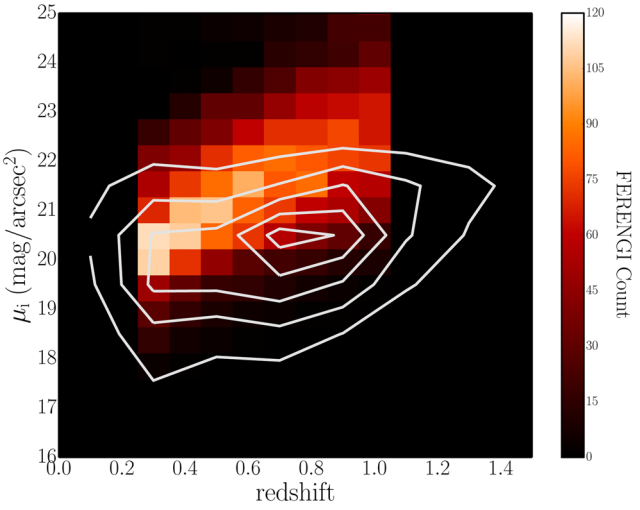

Figure 6. Surface brightness as a function of redshift for 3449 FERENGI images and the 102 548 `main` galaxies with measured μ and z values. The colour histogram shows the number of FERENGI images as a function of μ and $z_{\rm sim}$. White contours show counts for the galaxies in the `main` sample, with the outermost contour starting at $N = 1500$ and separated by intervals of 1500.

in order to accurately measure demographics over any significant redshift range.

For both previous releases of Galaxy Zoo morphologies, a correction for redshift-dependent bias was applied based on matching the mean classification fractions at the highest redshifts with those at the lowest redshift. Bamford et al. (2009) and Willett et al. (2013) provide complete descriptions of the process for GZ1 and GZ2, respectively.

Instead, in GZH the redshift range is large enough that cosmological evolution of the types and morphologies of galaxies is expected for the *HST* sample. In addition, the effects of band shifting will change the images even more across these redshift ranges. As a result, the previous methods of correcting for redshift-dependent bias do not work.

In order to test and correct for the effects of redshift, GZH includes a set of calibration images. These are simulated images of a set of nearby galaxies as they would appear observed at a variety of redshifts. The input images are from the SDSS (York et al. 2000; Strauss et al. 2002). The sample consists of 288 well-resolved galaxies at $z < 0.013$. The galaxies spanned a variety of morphologies (as selected by GZ2 classifications, including identifications of spiral structure, ellipticals, mergers, edge-on discs, bulge prominence/shape, and bars) and *r*-band surface brightnesses. The selection of galaxies spanned the redshift range of SDSS targets and maximized the number of *HST* galaxies at the same surface brightness and redshift. The FERENGI code was used to produce sets of images corresponding to observations of these galaxies out to $z_{\rm sim} = 1.0$ (see Section 2.3.2). The resulting images were classified in the GZH interface using the standard classification scheme.

Fig. 6 shows the distribution in μ and z of the artificially redshifted FERENGI images compared to the genuine *HST* images. The full bivariate distributions differ due to a combination of the detection limits of the *HST* surveys, intrinsic rareness of bright μ galaxies in the SDSS volume, and evolution of the stellar populations. Since the dependence of the debiasing correction is evaluated in separate bins, the main concern is the existence of an overlap between the two sets of images; this is the case for almost all of the *HST* images, which have $\mu > 19$ mag arcsec^{-2}. The brightest end of the distribution, which has fewer examples of comparable FERENGI images, corresponds to galaxies for which the debiased correction is expected to be minor.

4.1 Effects of morphological debiasing

The approach used in GZH for correcting the weighted classifications for redshift bias rests on the assumption that the *degree* of bias is a function of the apparent size and brightness of the galaxy. 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. The combination of all such parameters forms a high-dimensional space, and there is no obvious technique for separating these into individual effects. Instead, the method used here employs only two parameters as a simple model of the effect of redshift brightness on classification: the surface brightness (μ; intrinsic) and redshift (z; extrinsic).

Fig. 7 shows the change in $f_{\rm features}$ as a function of their lowest simulated redshift for the 3449 FERENGI images with robust photometric measurements across the full range of redshifts. For each simulated redshift value $z_{\rm sim}$ at a fixed surface brightness, $f_{{\rm features},z}$ is the value measured at that simulated redshift. We plot $f_{{\rm features},z}$ against $f_{{\rm features},z=0.3}$, the value measured for the same galaxy at $z_{\rm sim} = 0.3$.

The objective is to use these data to predict, for a galaxy with a measured $f_{{\rm features},z}$ value, what its $f_{\rm features}$ value *would have been* if it had been viewed at $z = 0.3$. This predicted value is defined as the 'debiased' vote fraction $f_{\rm features,\,debiased}$, and is calculated by applying a correction to the measured value of $f_{\rm features}$, determined by the ζ function described in the following section (equation 4). A reliable predicted value can be obtained so long as the relationship between $f_{{\rm features},z}$ and $f_{{\rm features},z=0.3}$ is single-valued; that is, for a given $f_{{\rm features},z}$, there is exactly one corresponding value of $f_{\rm features}$ at $z = 0.3$.

Fig. 7 shows that the relationship between $f_{{\rm features},z}$ and $f_{{\rm features},z=0.3}$ is *not* always single-valued; hence, it is not appropriate to correct galaxies that lie in certain regions of surface brightness/redshift/$f_{\rm features}$ space. Such regions tend to have low $f_{\rm features}$ values at high redshift, but a wide range of values at $z = 0.3$. These regions contain two morphological types of galaxies: the first set are genuine ellipticals, which have low values of $f_{\rm features}$ at both high and low redshift. The second group are discs whose features become indistinct at high redshift; hence, their $f_{\rm features}$ 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 (see the top-right panel of Fig. 7).

The criteria for determining whether a region of this space is single-valued, and therefore correctable, are as follows. In each surface brightness and redshift bin, the relationship between $f_{{\rm features},z}$ and $f_{{\rm features},z=0.3}$ is modelled by fitting the data with polynomials of degrees $n = 3$, 2, and 1, and using the best formal fit out of the three as measured by the sum of the residuals. These fits are shown as the dashed black lines in Fig. 7. Flat regions of the bins are areas in which there is *not* a clear single-valued relationship between $f_{{\rm features},z}$ and $f_{{\rm features},z=0.3}$. We quantify this by measuring the slope of the best-fitting polynomial to the vote fractions. Regions within the bins with a slope less than 0.4 (a boundary selected through manual inspection and testing) are considered *not* one-to-one, and therefore $f_{{\rm features},z}$ cannot be boosted to its $f_{{\rm features},z=0.3}$ value. Galaxies in this

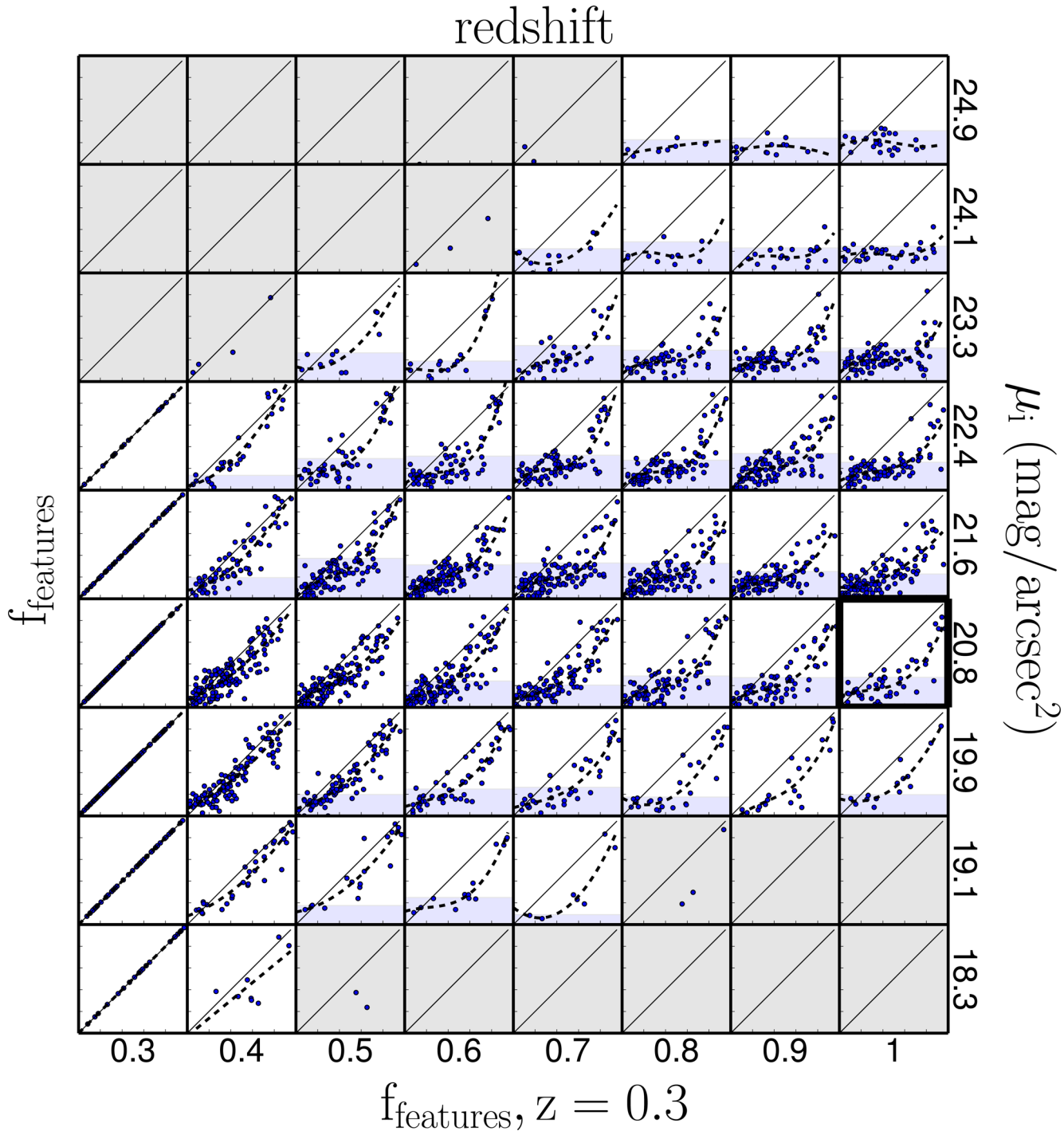

Figure 7. Effects of redshift bias in 3449 images in the FERENGI sample. Each point *in a given redshift and surface brightness bin* represents a unique galaxy. On the *y*-axis in each bin is the $f_{features}$ value of the image of that galaxy redshifted to the value corresponding to that redshift bin. On the *x*-axis is the $f_{features}$ value of the image of the same galaxy redshifted to $z = 0.3$. The dashed black lines represent the best-fitting polynomials to the data in each square. The solid black line represents $f_{features,z}$=$f_{features,z=0.3}$. Regions in which there is a single-valued relationship between $f_{features}$ 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 grey. A larger version of the bin outlined at $z = 1.0$ and $20.3 < \mu < 21.0$ (mag arcsec^{-2}) is shown in Fig. 8.

region are referred to as the *lower limit* sample, because the most stringent correction available is that the weighted $f_{features}$ is a lower limit to the true value. These regions are highlighted in blue in Fig. 7. Uncoloured (white) regions of the plot have sufficiently high slopes to consider the relationship as 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 five galaxies were considered; regions with fewer than five galaxies are considered to have 'not enough information' to determine the $f_{features,z}$ and $f_{features,z=0.3}$ relationship, coloured grey in Fig. 7. These galaxies are hereafter referred to as the 'NEI' sample.

Table 3. Number of correctable galaxies for the top-level task in GZH, split by *HST* survey.

	Correction type	AEGIS	COSMOS	GEMS	GOODS-N five-epoch	GOODS-S five-epoch	Total
Correctable	0	2908	21 169	2802	1459	1189	29 527
Lower limit	1	833	5169	1021	1377	1267	9667
No correction needed ($z \leq 0.3$)	2	955	10 870	1175	415	400	13 815
NEI	3	2677	43 058	3559	2077	2184	53 555
No redshift information	4	1134	4688	530	687	102	7141
Total		8507	84 954	9087	6015	5142	113 705

The unshaded regions in Fig. 7 define discrete ranges of redshift, surface brightness, and $f_{\rm features}$ within which a galaxy must lie in order for the debiased correction to be confidently applied. While the appropriate correctable regions were defined as discrete bins, the true correctable region is assumed to be a smooth function of z, μ, and $f_{\rm features}$. To define this smooth space, we calculate the shape of the convex hull that encloses the correctable and lower limit FERENGI galaxies in z–μ–$f_{\rm features}$ space. The boundaries are then adjusted until the contamination from both groups is minimized. The resulting hulls define the correctable and lower limit regions for categorizing the *HST* galaxies. The results of this method and final categorization of the *HST* sample are in Table 3. Of the galaxies at redshift higher than $z = 0.3$, 17 per cent can be debiased using the ζ method, 27 per cent cannot be debiased since the relationship between $f_{\rm features,z}$ and $f_{\rm features,z=0.3}$ is not monotonic, and 56 per cent cannot be debiased since they either have an unknown redshift or insufficient numbers of FERENGI images in their z–μ bin to determine a reliable correction term (ζ, described in Section 4.2).

For the 'lower limit' galaxies for which a single debiased $f_{\rm features}$ value cannot be confidently assigned, the *range* of debiased values is estimated and included as a data product. This uses the FERENGI simulated data to analyse the range of intrinsic $f_{\rm features,z=0.3}$ values for any given observed $f_{\rm features}$ value, again as a function of surface brightness and redshift. In each unclassifiable (shaded blue) z, μ bin, we compute the spread of intrinsic values of $f_{\rm features,z=0.3}$ for five quantiles of observed $f_{\rm features}$ (corresponding to the 'clean' thresholds used in prior GZ publications). The range of intrinsic values for GZH is defined by the upper and lower 1σ limits, enclosing the inner 68 per cent of the data. This range is represented by the orange bars in Fig. 8. For any galaxy which cannot be directly debiased by the ζ method, these ranges are used to denote the upper and lower limits on the expected values $f_{\rm features,z=0.3}$ as a function of the observed $f_{\rm features}$.

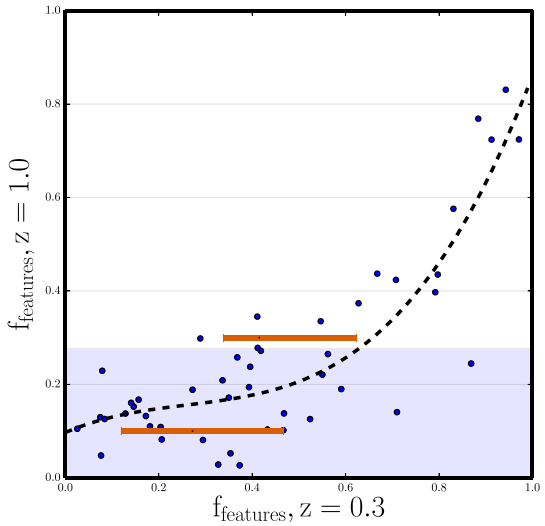

Figure 8. A larger version of the dark-outlined square in Fig. 7, containing FERENGI galaxies that have been artificially redshifted to $z = 1.0$ and have surface brightnesses between $20.3 < \mu < 21.0$ (mag/arcsec^2). The orange bars represent the inner 68 per cent (1σ) of the uncorrectable $f_{\rm features}$ quantiles, which are used to compute the limits on the range of debiased values.

4.2 Correcting morphologies for classification bias

For the 'correctable' sample of galaxies, we observe a decline in the vote fraction $f_{\rm features}$ with increasing simulated redshift for each unique galaxy. We model this relationship for artificially redshifted images with a simple exponential function which bounds $f_{\mu,\,z>0.3}$ between $f_{\mu,\,z=0.3}$ and 0:

$$f_{\mu,z} = 1 - (1 - f_{\mu,z=0.3})e^{\frac{z-z_0}{\zeta}}, \quad (2)$$

where $f_{\mu,\,z=0.3}$ is the vote fraction at the lowest redshift in the artificially redshifted sample ($z_0 = 0.3$). Here, ζ is a parameter that controls the rate at which $f_{\rm features}$ decreases with increasing redshift (and that may depend on other galaxy properties). This function bounds the observed vote fractions between $f_{\mu,\,z=0.3}$ and zero based on two assumptions: (1) the vote fractions for featured galaxies decrease monotonically with increasing redshift, therefore restricting vote fractions for a given galaxy to be less than $f_{\mu,\,z=0.3}$ (which is almost always true of the data), and (2) the vote fractions cannot be less than zero (which is always true).

Fig. 9 shows the change in vote fraction and the best-fitting model for a random set of galaxies in the FERENGI sample. The results show that there is a clear dependence on redshift for the observed changes in $f_{\rm features}$. To examine whether a better fit should involve additional parameters, we tested the global dependence of ζ on the galaxy surface brightness. This is motivated by the fact that brighter galaxies would presumably be easier to identify; Bamford et al. (2009) and Willett et al. (2013) found that changes in $f_{\rm features}$ depend on a combination of physical size and absolute magnitude. To simplify our model, we subsume these into a single parameter of surface brightness, μ.

Fig. 10 examines the results of fitting the FERENGI images with equation (2). This fit only includes galaxies where SExtractor robustly measured the photometry in all images between $0.3 < z_{\rm sim} < 1.0$ (such that $\sigma_{M_I} < 1.0$ mag; Melvin 2016). Interestingly, the derived correction is only a very weak function of surface brightness. Higher surface brightness galaxies have on average slightly stronger corrections, possibly because these galaxies have larger $f_{\rm features}$ values at high redshifts. Low surface brightness

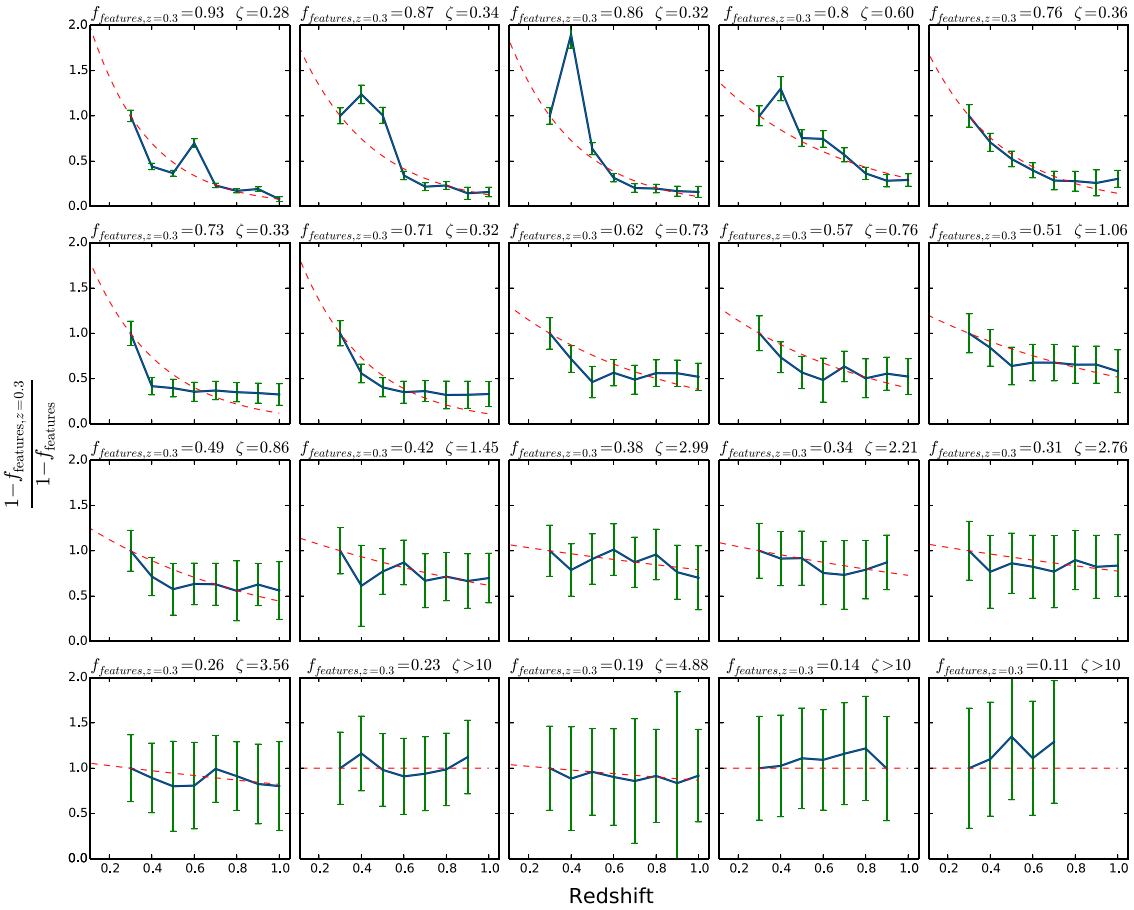

Figure 9. Behaviour of the normalized, weighted vote fractions of features visible in a galaxy ($f_{features}$) as a function of redshift in the artificial FERENGI images. These plots demonstrate overall trends in the sensitivity to feature detection as a function of redshift. Galaxies in this plot were randomly selected from a distribution with evolutionary correction $e = 0$ and at least three detectable images in redshift bins of $z \geq 0.3$. The displayed bins are sorted by $f_{features,z=0.3}$, labelled above each plot. Measured vote fractions (blue solid line) are fitted with an exponential function (red dashed line; equation 2); the best-fitting parameter for ζ is given above each plot.

galaxies are more likely to begin low and remain low; the bounded nature of the drop-off (and variance among the individual voters) means that the average magnitude of ζ will be lower.

Since there is little evidence for any strong systematic dependence of ζ on μ, we do not include any additional parameters in fitting to our calibration model. We fit the data in Fig. 10 with a linear function:

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

where $\hat{\zeta}$ is the correction factor applied to each galaxy as a function of surface brightness. The best-fitting parameters to the linear fit from least-squares optimization are $\zeta_0 = 0.50$, $\zeta_1 = -0.03$. To make the final debiased correction for the genuine *HST* data, we apply a correction similar to equation (2):

$$f_{\text{features,debiased}} = 1 - (1 - f_{\text{features, weighted}})e^{\frac{-(z-z_0)}{\hat{\zeta}}}, \quad (4)$$

where $f_{\text{features, weighted}}$ is the weighted vote fraction described in Section 3.2, and $f_{\text{features, debiased}}$ is bounded between $f_{\text{features, weighted}}$ and 1.

Fig. 11 shows the distribution of the fraction of galaxies that are ultimately correctable as a function of $f_{features}$ and redshift. The distributions for individual samples (AEGIS, COSMOS, GEMS, GOODS) are individually very similar. We emphasize that a correction is more likely to be derived and applied for galaxies at higher redshift and with higher weighted values of $f_{features}$. This has important consequences for selection of physically meaningful samples from GZH (see Section 6), meaning that comparative studies of galaxies should use a threshold on the likelihood of features that evolves as a function of redshift.

Few galaxies in the sample have sufficiently high corrections to completely change them from being confidently 'smooth' to 'featured' following the correction for redshift bias (Fig. 12). As a check, we compare morphologies of highly boosted galaxies to the expert visual classifications in CANDELS (Kartaltepe et al. 2015). Only nine galaxies that were strongly boosted ($f_{features} < 0.25$ and $f_{features,\ debiased} > 0.50$) in GZH also appear in the CANDELS expert sample. Of those, the CANDELS expert sample classifies four as spheroids/discy, two as discs, one as a spheroid, one as irregular/discy, and one as unclassifiable. The $f_{features,best}$ values for GZH are all between 0.5 and 0.7, making them all intermediate disc candidates. Visual inspection of the GZH images shows extended disc structure in at least five images; the remaining four are either extremely faint or have imaging artefacts in the ACS data. Since the surveys use different rest-frame filters and the overlap sample is small, though, detailed comparisons between the overall morphologies are highly difficult (see Sections 7.2 and 7.3).

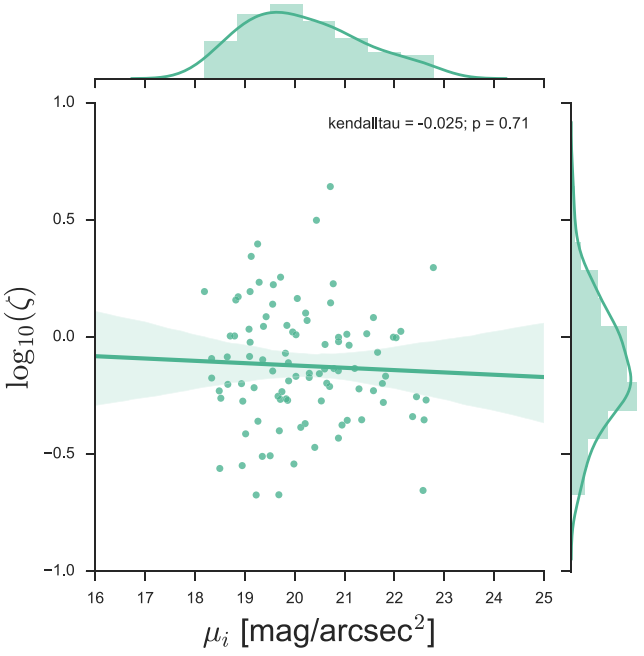

Figure 10. All fits for the FERENGI galaxies of the vote fraction drop-off parameter ζ for $f_{\rm features}$ as a function of surface brightness. This includes only the simulated galaxies with a bounded range on the drop-off ($-10 < \zeta < 10$) and sufficient points to fit each function (28 original galaxies, each with varying images artificially redshifted in one to eight bins over a range from $0.3 \lesssim z_{\rm sim} \lesssim 1.0$).

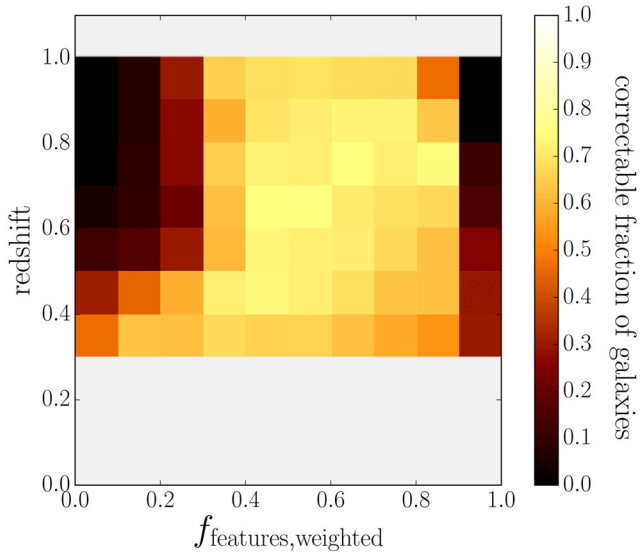

Figure 11. Histogram showing the fraction of galaxies that have a finite correction for the debiased vote fractions $f_{\rm features,\,debiased}$ as a function of $f_{\rm features}$ and redshift. The parameter space for corrections is limited to $0.3 \leq z \leq 1.0$ due to the sampling of the parent SDSS galaxies and detectability in the FERENGI images.

4.3 Challenges of debiasing questions beyond 'smooth or features'

As with the *HST* images, each FERENGI subject had a varying number of classifiers answering the various questions in the hierarchical decision tree. Every classifier answers the first question, *'Is the galaxy smooth and rounded, with no sign of a disc?'*; as such, the vote fractions $f_{\rm smooth}$, $f_{\rm features}$, and $f_{\rm artefact}$ all have the lowest statistical error for any question in the tree, based only on the total number of responses (between 40 and 120; see Section 3). The number of participants answering subsequent questions, however, is always equal to or less than the number who answered 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 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 that can be extracted for the higher tier questions. The debiasing technique used (Section 4.2) requires that at least 10 classifiers answer each question for a galaxy with $z_{\rm sim} = 0.3$ *and* the corresponding image at higher redshift. This requirement is (by default) met by all galaxies for the smooth/features question. However, this is often *not* met for questions beyond Task 01. On average, 60 per cent $\pm$ 24 per cent of the galaxies do not have sufficient data to measure a correction, as compared to 2.0 per cent achieved for Task 01 (Table 4). This leaves the average surface brightness/redshift bin with insufficient data points to confidently measure the change in vote fraction ($\lesssim$10 galaxies per bin). For these reasons, debiased vote fractions are only provided in the GZH catalogue for Task 01 (smooth/features). We suggest that use of morphological data for higher order tasks should instead use the weighted vote fractions (see Section 6).

5 THE GZH CATALOGUE

The catalogue for GZH includes morphological data for 189 149 images (generated from a total of 145 741 unique galaxies). The full table can be accessed at http://data.galaxyzoo.org. The online data also include a secondary metadata table, which is drawn from a variety of sources detailed in Section 2.

Each image is listed under a unique project ID (e.g. 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 are provided: for each question, $N_{\rm votes}$ is the number of classifiers who answered that question. For each unique answer, `fraction` is the fraction of classifiers who selected that answer ($N_{\rm answer}/N_{\rm votes}$), and `weighted` is the weighted fraction, which takes into account overall consistency (Section 3.2).

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, disc structure, etc.). For this reason, caution must be taken when taking vote fractions as cut-offs to determine morphological structure; guidelines for careful classification are given in Section 6.

The GZH catalogue is corrected for redshift bias only for the first question of the GZH decision tree (Section 4), which asks *'Is the galaxy smooth and round, with no sign of a disc?'*. For this question, the catalogue provides the additional parameters `debiased`, `lowerlimit`, `upperlimit`, and `best` vote fractions. The `best` fraction for $f_{\rm features}$ is chosen based on the categorization of the galaxy: if it is 'correctable', `best` = `debiased`; if it is a lower limit, `best` = `lowerlimit`; if neither condition applies, then `best` = `weighted`.

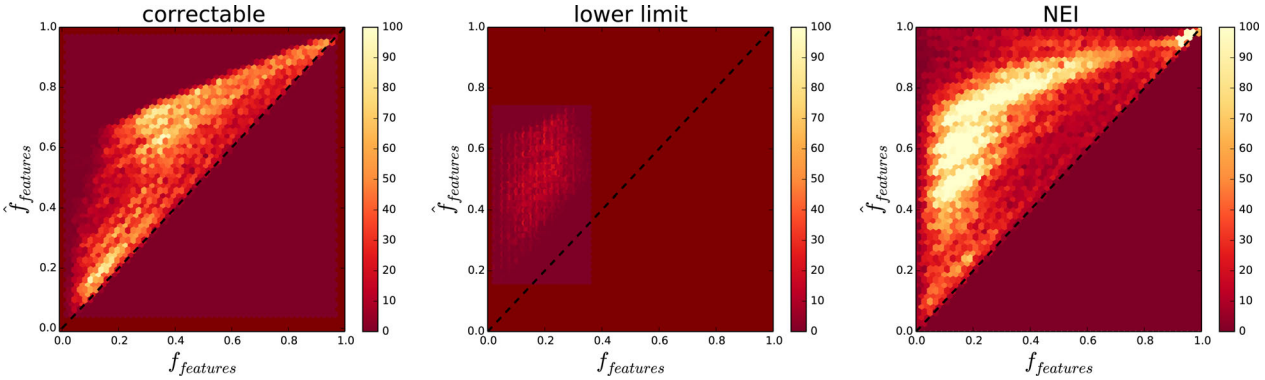

Figure 12. Debiased $f_{\rm features}$ corrected to $z = 0.3$ versus weighted $f_{\rm features}$ for the correctable (left), lower limit (middle), and 'not enough information' (NEI; right) galaxies in the GZH sample.

Table 4. Distribution of the ability of images to measure morphological bias for the FERENGI data (see Fig. 7).

	N	per cent
Correctable	1690	49 per cent
Lower limit	1678	49 per cent
NEI	81	2 per cent
Total	3449	100 per cent

The debiased and best vote fractions for $f_{\rm smooth}$ are calculated on the criteria that vote fractions for all answers must sum to unity:

$$f_{\rm smooth} \equiv 1 - f_{\rm features,best} - f_{\rm artefact}. \quad (5)$$

In rare cases (1.2 per cent of the `main` sample), this requirement resulted in negative vote fractions for $f_{\rm smooth}$; these were cases in which the $f_{\rm features}$ vote fraction was boosted to a high value relative to $f_{\rm artefact}$. In these cases, the constraint of equation (5) is met by setting $f_{\rm smooth}$=0.0 and decreasing $f_{\rm features,best}$ accordingly. This correction is typically small, with a median change of $f_{\rm features}$ or $f_{\rm smooth}$ of $\Delta f = 0.04$.

Tables 2, 5–10 also contain flags for Task 01 that identify clean (but not complete) samples of galaxies with high likelihoods of being either smooth, featured, or a star/artefact. These are set as `smooth/featured/artefact_flag = 1` if $f_{\rm smooth/features/artefact\_best} > 0.8$. For a galaxy to be flagged as 'smooth', an additional criterion of `correction_type` =0, 1, or 2 is applied. This accounts for the uncertainty in distinguishing between genuine ellipticals and discs whose features have been washed out due to surface brightness and redshift effects (Section 4).

Data products for GZH are split by the type of image being classified. Each sample in Section 2.5 corresponds to the data in Tables 2, 5–10.

Table 5. GZH morphological classifications for colour-faded *HST* images.

			t01_smooth_or_features_			t01_smooth_or_features_a01_smooth_						...
Zooniverse ID	Survey ID	Imaging	Correction	$N_{\rm votes}$	Fraction	Weighted	Debiased	Best	Lower limit	Upper limit	Flag	
AHZF000001	20000002	COSMOS	1	48	0.708	0.755	0.294	0.755	0.450	0.831	0	
AHZF000003	20000004	COSMOS	0	49	0.367	0.379	0.136	0.136	0.161	0.289	0	
AHZF000004	20000006	COSMOS	3	49	0.265	0.271	0.038	0.271	–	–	0	
AHZF00000z	20000102	COSMOS	1	44	0.727	0.780	0.296	0.780	0.421	0.797	0	
AHZF000010	20000104	COSMOS	2	53	0.811	0.849	0.894	0.849	–	–	0	
...												

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 3927 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 6. GZH morphological classifications for colour-inverted *HST* images.

			t01_smooth_or_features_			t01_smooth_or_features_a01_smooth_						...
Zooniverse ID	Survey ID	Imaging	Correction	$N_{\rm votes}$	Fraction	Weighted	Debiased	Best	Lower limit	Upper limit	Flag	
AHZC000001	20000002	COSMOS	1	52	0.615	0.664	0.223	0.664	0.396	0.777	0	
AHZC000003	20000004	COSMOS	0	48	0.333	0.364	0.049	0.049	0.026	0.154	0	
AHZC000004	20000006	COSMOS	3	51	0.235	0.252	0.016	0.252	–	–	0	
AHZC00000z	20000102	COSMOS	1	49	0.755	0.757	0.327	0.620	0.279	0.620	0	
AHZC000010	20000104	COSMOS	2	51	0.843	0.882	0.926	0.881	–	–	1	
...												

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 3927 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 7. GZH morphological classifications for GOODS two-epoch images.

			t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_						...
Zooniverse ID	Survey ID	Imaging	Correction	N_{votes}	Fraction	Weighted	Debiased	Best	Lower limit	Upper limit	Flag
AHZ3000001	50000000	GOODS-N	3	123	0.390	0.415	0.079	0.415	–	–	0
AHZ3000002	50000001	GOODS-N	2	126	0.341	0.355	0.356	0.356	–	–	0
AHZ3000003	50000005	GOODS-N	0	129	0.760	0.826	0.641	0.641	0.622	0.835	0
AHZ3000004	50000008	GOODS-N	3	120	0.758	0.787	0.629	0.787	–	–	0
AHZ3000005	50000010	GOODS-N	3	123	0.854	0.890	0.605	0.890	–	–	0
...											

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 6144 galaxies and morphological information for all tasks in the tree. A subset of the information is shown here to illustrate form and content.

Table 8. GZH morphological classifications for SDSS Stripe 82 single-epoch images.

			t01_smooth_or_features_	t01_smooth_or_features_a01_smooth_		...
Zooniverse ID	Survey ID	Imaging	N_{votes}	Fraction	Weighted	Flag
AHZ5000001	587730845812064684	SDSS	41	0.585	0.595	0
AHZ5000002	587730845812065247	SDSS	46	0.609	0.651	0
AHZ5000003	587730845812196092	SDSS	51	0.039	0.044	0
AHZ5000004	587730845812196825	SDSS	35	0.514	0.605	0
AHZ5000005	587730845812524122	SDSS	47	0.766	0.812	1
AHZ5000006	587730845812654984	SDSS	42	0.500	0.542	0
AHZ5000007	587730845812655541	SDSS	41	0.488	0.526	0
AHZ5000008	587730845812720365	SDSS	53	0.792	0.84	1
AHZ5000009	587730845812720640	SDSS	43	0.000	0.0	0
AHZ500000a	587730845812720699	SDSS	40	0.425	0.478	0
...						

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 9. GZH morphological classifications for SDSS Stripe 82 co-added images.

			t01_smooth_or_features_	t01_smooth_or_features_a01_smooth_		...
Zooniverse ID	Survey ID	Imaging	N_{votes}	Fraction	Weighted	Flag
AHZ6000001	8647474690312306978	SDSS	40	0.275	0.289	0
AHZ6000002	8647474690312307154	SDSS	43	0.605	0.634	0
AHZ6000003	8647474690312307877	SDSS	51	0.608	0.627	0
AHZ6000004	8647474690312308301	SDSS	52	0.038	0.038	0
AHZ6000005	8647474690312308318	SDSS	44	0.614	0.632	0
AHZ6000006	8647474690312308880	SDSS	36	0.667	0.683	0
AHZ6000007	8647474690312372644	SDSS	48	0.646	0.674	0
AHZ6000008	8647474690312372789	SDSS	45	0.489	0.571	0
AHZ6000009	8647474690312372931	SDSS	47	0.553	0.587	0
AHZ600000a	8647474690312373190	SDSS	47	0.574	0.559	0
...						

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.

6 USING THE GZH CATALOGUE

The primary purpose of the GZH catalogue is to provide a reliable method of selecting galaxies of a desired morphological type. This section provides instructions for creating such pure samples using the vote fractions corresponding to the tasks shown in Fig. 4 (e.g. Masters et al. 2011; Cheung et al. 2015; Galloway et al. 2015). Increasing the levels of the thresholds can create purer, but not necessarily complete, morphologically selected samples. These are useful for selecting galaxies of rare or unique types that merit individual study. Looser cuts can be applied to obtain samples with a higher level of completeness, although the rate of false positives must be closely monitored. We stress that the details of any selection process will vary based on the particular science case; for example, Bamford et al. (2009), Skibba et al. (2009), and Smethurst et al. (2015) all demonstrate the advantages of using the vote fractions directly as weights rather than applying discrete thresholds.

To select galaxies of a morphological type identified with a particular task, a cut is placed on the vote fraction for that task (f_{task}), *as well as* the vote fractions for the tasks preceding it, because of the dependence induced by the decision-tree structure. For

Table 10. GZH morphological classifications for *HST* images with simulated AGN.

					t01_smooth_or_features_		t01_smooth_or_features_a01_smooth_		...
Zooniverse ID	Survey ID	Imaging	Version	$L_{\rm ratio}$	AGN colour[a]	$N_{\rm votes}$	Fraction	Weighted	Flag
AHZ7000001	9002470011	GEMS	1	0.2	1	42	0.238	0.239	0
AHZ7000002	9002470012	GEMS	1	1.0	1	51	0.255	0.265	0
AHZ7000003	9002470013	GEMS	1	5.0	1	47	0.170	0.167	0
AHZ7000004	9002470014	GEMS	1	10.0	1	41	0.195	0.195	0
AHZ7000005	9002470015	GEMS	1	50.0	1	47	0.170	0.178	0
...									
AHZ70002a4	9010781875	GOODS-S	2	50.0	2	33	0.242	0.285	0
AHZ70002a5	9010781881	GOODS-S	2	0.2	3	32	0.312	0.323	0
AHZ70002a6	9010781882	GOODS-S	2	1.0	3	35	0.543	0.559	0
AHZ70002a7	9010781883	GOODS-S	2	5.0	3	28	0.429	0.460	0
AHZ70002a8	9010781884	GOODS-S	2	10.0	3	25	0.200	0.200	0
AHZ70002a9	9010781885	GOODS-S	2	50.0	3	30	0.167	0.167	0
...									

Notes. [a]Flag indicating the colour of the PSF in the simulated AGN. 0 = no simulated AGN, 1 = blue, 2 = flat, 3 = red.
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 2961 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. Suggested thresholds for selecting morphological samples from GZH.

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

example, to select barred galaxies, a cut may be placed on $f_{\rm bar}$ 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 classifiers 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 $f_{\rm bar}$ is well sampled, cuts on all previous tasks must be applied.

The flexibility of this catalogue allows users to set their own selection criteria for vote fraction thresholds to create a morphologically pure sample. Table 11 provides suggested cuts for selecting galaxies with a variety of morphologies. These thresholds were determined by visual inspection of various subsamples of GZH data. The thresholds employ a combination of 20 votes for the task being considered, as well as limits on the vote fractions for previous response(s) in the decision tree.

We visually analysed subsamples of 50 galaxies meeting both criteria, as well as a control sample of galaxies which had 20 classifiers 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 per cent purity.

To use data from Table 11 to create, for example, a sample of three-armed spiral galaxies, we suggest selecting objects with $N_{\rm armnumber} \geq 20$, $f_{\rm features} > 0.23$, $f_{\rm clumpy,\ no} > 0.30$, $f_{\rm edge\text{-}on,\ no} > 0.25$, and $f_{\rm spiral,yes} > 0.25$. These cuts define a sample of galaxies of 'arm number candidates'; i.e. galaxies for which answering the arm number question makes physical sense and the vote fraction $f_{\rm arm\ number}$ is well sampled. Such galaxies will be featured, non-clumpy, non-edge-on spirals. At this point, a final cut can be made on $f_{\rm arm\ number\ =\ 3}$ to select spirals with three arms.

Tasks 03, 04, and 05 have an additional possible pathway; as shown in Fig. 4, a classifier might also be shown this question if they select 'featured/disc' in Task 01, 'clumpy' in Task 12, two or more clumps in Task 16, and 'spiral arrangement' in Task 15. After applying the appropriate thresholds for this path, <0.5 per cent of the galaxies which have ≥20 answers to these questions used this pathway to arrive at these tasks. None of these images show obvious disc structure, although the clumps within are arranged in a spiral pattern.

This section described how to use Table 11 to select galaxies for which a particular task is reliably sampled. The following two examples extend this and show how to use the vote fractions to obtain samples of galaxies with a specific morphological type.

6.1 Example 1: selecting barred galaxies

Bars in galaxies are a trace of the dynamical state of the disc (e.g. Combes 2009; Athanassoula 2012). Discs which have significant vertical motions typically do not form bars (with the exception of tidally triggered bars; Barnes & Hernquist 1991), but once a disc settles to a thin, dynamically cool state, the formation of bars proceeds quickly. Most theoretical predictions show that bars are long-lasting in the absence of significant galaxy interactions

(Athanassoula, Lambert & Dehnen 2005), so the fraction of bars in disc galaxies can measure the dynamical maturity of that population. For this reason, tracking the fraction of bars in disc galaxies as a function of redshift has attracted significant interest since the first resolved images of high-redshift galaxies were obtained (e.g. Abraham et al. 1999; Elmegreen, Elmegreen & Hirst 2004; Jogee et al. 2004; Sheth et al. 2008; Melvin et al. 2014; Simmons et al. 2014).

We create a sample of barred disc galaxies in GZH by applying cuts on the previous tasks in Table 11. We first identify 11 049 'bar candidates', which are disc galaxies that are sufficiently face-on to attempt visual identification of a bar. These galaxies were selected by applying the cuts $N_{\rm bar} \geq 20, f_{\rm features} > 0.23, f_{\rm clumpy,\ no} > 0.30$, and $f_{\rm edge\text{-}on,\ no} > 0.25$. These galaxies are featured, non-clumpy, non-edge-on galaxies. Of these, a pure sample of 730 barred discs was identified by applying a cut of $f_{\rm bar} > 0.7$. A subsample of 50 galaxies was visually inspected and 94 per cent were found to contain strong bars. A complete sample of strong and weak bars was created by applying a cut of $f_{\rm bar} > 0.3$. This sample contained 3218 galaxies, 86 per cent of which were found to contain weak or strong bars through visual inspection.

The resulting bar sample can be used to estimate the redshift evolution of bar fraction; we find a steady decrease of $f_{\rm bar} \sim 0.32$ at $z = 0.4$ to $f_{\rm bar} \sim 0.24$ at $z = 1.0$. This decrease agrees with Melvin et al. (2014), although they report a lower overall bar fraction going from $f_{\rm bar} = 0.22$ at $z = 0.4$ to $f_{\rm bar} = 0.11$ at $z = 1.0$. The difference in total bar fraction is expected, as this analysis used a looser cut on $f_{\rm bar}$, there is no luminosity cut, and the use of debiased values for $f_{\rm features}$ increases the total amount of discs in the sample compared to Melvin et al. (2014). However, the results from GZH indicate that if features are reliably identified, the bar is always visible; for galaxies at increasing redshifts, the discs fade first but the galactic bar remains visible (Melvin 2016). As a result, we do *not* recommend using an evolving cut on $f_{\rm bar}$ as a function of redshift for the selection of barred discs.

Another existing study using GZH to select barred galaxies was presented in Cheung et al. (2015). This study used a slightly different bar sample selection to demonstrate that AGN hosts show no statistically significant enhancement in bar fraction or average bar likelihood compared to closely matched inactive galaxies. We note that their technique matched test and control galaxies over the same redshift ranges, which minimizes the possible impact of redshift bias.

6.2 Example 2: identifying clump multiplicity

Clumps are known to be a characteristic feature of galaxies at high redshift, and there is evidence that they play a crucial 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, Burkert & Schartmann 2016). Simulations show clumps migrate from the outer disc to the galactic centre in only a few orbital periods (Mandelker et al. 2017), 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 disc galaxies. GZH includes a 'clumpy' path in the decision tree for the purposes of identifying clumps and investigating their evolution with redshift.

For galaxies identified as 'clumpy' in GZH, the number of clumps can be determined using Task 16. Table 11 can be used to select 8444 galaxies measured as 'clumpy' using $f_{\rm features} > 0.23$ and $f_{\rm clumpy,yes} > 0.80$ to ensure that the vote fractions for Task 16 are well sampled. The clump number can be reasonably identified for 1112 of the clumpy galaxies; for 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 1112 which did have distinguishable clumps, there are 61 one-clump galaxies using $f_{\rm 1clump} > 0.50$, 442 two-clump galaxies using $f_{\rm 2clumps} > 0.80$, 275 three-clump galaxies using $f_{\rm 3clumps} > 0.75$, 71 four-clump galaxies using $f_{\rm 4clumps} > 0.70$, and 263 galaxies with more than four clumps using $f_{\rm >4clumps} > 0.70$. Alternatively, these data may be used to create more general samples of clumpy galaxies with few clumps and many clumps. A sample of 989 'few clumps' galaxies can be made using $(f_{\rm 1clump} + f_{\rm 2clumps}) > 0.5$ and 2910 'many clumps' galaxies using $(f_{\rm 3clumps} + f_{\rm 4clumps} + f_{\rm >4clumps}) > 0.5$.

7 ANALYSIS

7.1 Demographics of morphology

Any analysis of the morphological distribution of galaxies must properly consider morphology with respect to other physical properties of the sample, such as colour, mass, size, environment, and redshift. We defer such analyses to later papers, and comment here only on a few broad characteristics of the overall GZH sample.

As an example of visualizing the overall morphological distribution, Fig. 13 shows the breakdown of GZH morphologies as a flow diagram. In order to show a physically meaningful sample, Fig. 13 only includes images from a volume-limited sample ($z < 1.0$, $m < 22.5$ in the *I* band) in a single *HST* survey (COSMOS, which has the largest number of total galaxies). We use a simple plurality vote for the responses to each task to characterize the morphologies. This emphasizes one of the basic results of the GZH project, namely that there are comparable numbers of spheroid and disc galaxies in the Universe at $z \sim 1$, with significant diversity in the arrangement of spiral arms and internal clumps.

The overall distribution of galaxy types is significantly different from the low-redshift sample classified in SDSS imaging from GZ1 and GZ2. Lintott et al. (2011) found that elliptical galaxies exceeded spiral galaxies by a factor of ∼2: 1 in their spectroscopic sample when using a plurality vote criterion [although Bamford et al. (2009) show that this strongly depends on the selection method; spirals are the dominant population in a volume-limited sample at $z < 0.088$]. The data for GZH show that smooth galaxies have a comparable total population to 'featured' galaxies of which roughly 20 per cent are dominated by clumps rather than well-organized discs. The fraction of objects identified as stars or artefacts is also much higher in the *HST* imaging. By plurality votes, these encompass only ∼0.1 per cent of images in SDSS (Willett et al. 2013), but 6 per cent of images in GZH.

Within the sample of galaxies identified as 'not smooth', it is clear that the addition of the clumpy branch is necessary to describe a large fraction of the sample, since disc-dominated galaxies outnumber clumpy morphologies by less than a factor of 2. Disc galaxies are mostly unbarred (Melvin et al. 2014) and possess two visible spiral arms over a flat distribution of pitch angles and bulge prominence. Clumpy galaxies are identified across the full range of clump multiplicities, with the exception of one-clump galaxies (which would be difficult to differentiate from compact spheroids). Roughly half of the galaxies have at least one clump identified as the brightest. Clumps are most commonly asymmetrically arranged in clusters and usually embedded within larger objects.

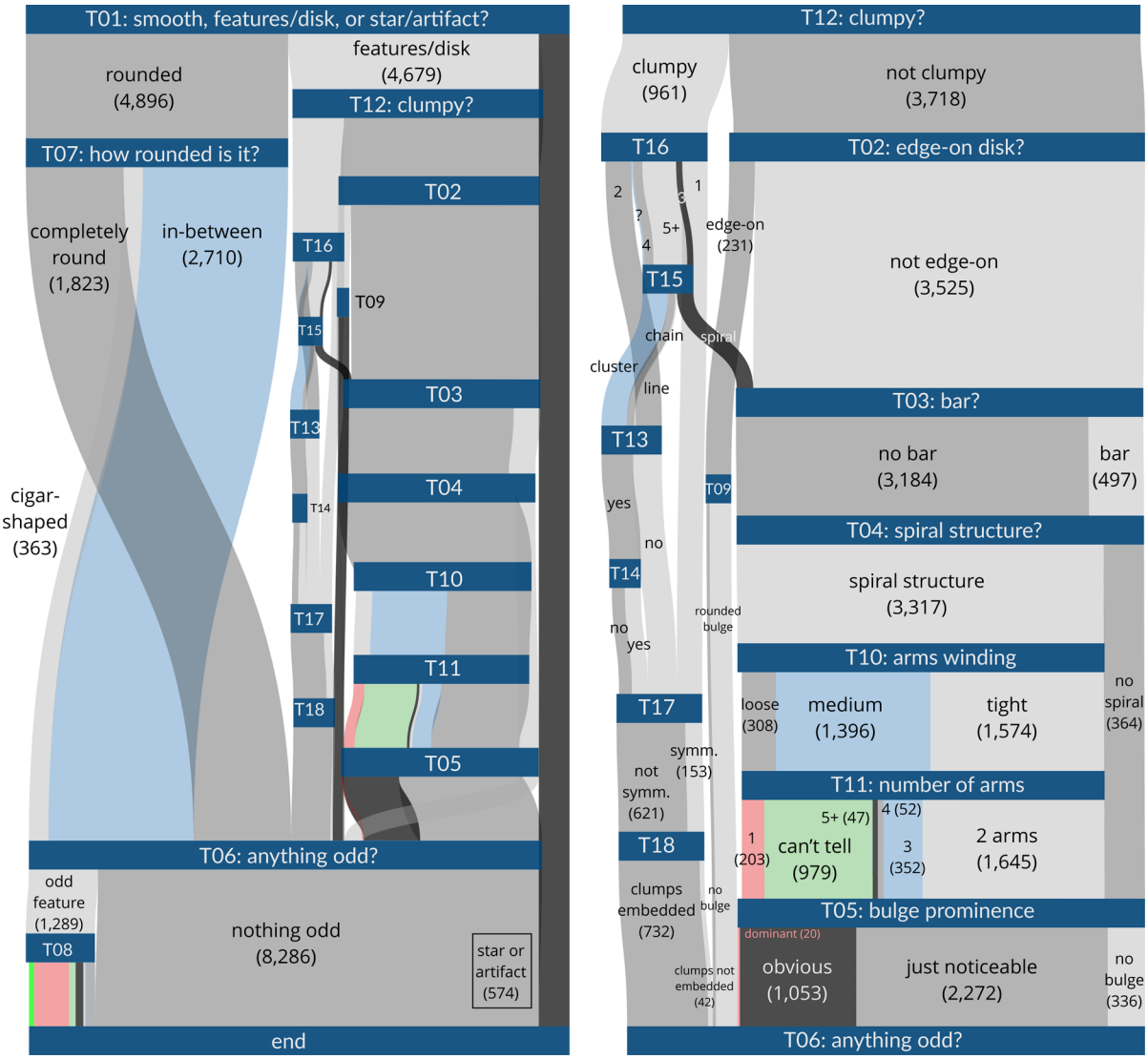

Figure 13. Demographics of the morphologies for a volume-limited sample ($z < 1.0$, $m_I < 22.5$) of the COSMOS galaxies in GZH. Each node in the diagram (dark blue horizontal bars) represents a task in the tree (Fig. 4). The left diagram shows the full decision tree. The right diagram zooms in on the features/clumpy tasks only. Paths between tasks represent each possible answer to the task, flowing from top to bottom between their origin and the subsequent task in the tree. Labels are assigned to galaxies based on the plurality answer for each task, with each galaxy assigned only one label per node. Widths of the paths are proportional to the number of galaxies assigned to that path. The widths of the nodes are proportional to the number of galaxies for which the question was reliably answered.

7.2 GZH and Galaxy Zoo: CANDELS

There are 7681 galaxies in the GOODS-S field with morphological classifications in both GZH and the Galaxy Zoo: CANDELS project (Simmons et al. 2016). Since both the sensitivity and filters for the two sets of images are significantly different (and there is no correction for redshift bias applied to GZC), there is no prior reason to expect a perfect correlation between the separate vote fractions for the projects. Briefly, we note that the $f_{\rm features}$ value for GZH is on average higher than GZC; the effect is strongest at $f_{\rm features,\,GZC} < 0.3$, for which roughly half the galaxies have $f_{\rm features,\,GZH} > 0.5$. However, the correlation between vote fractions is single-valued (although not linear, with a Pearson correlation coefficient of $r = 0.6$), and should be possible to calibrate using a similar approach to that described in Section 4; the correlation between other tasks, such as edge-on galaxies, is significantly stronger ($r = 0.9$). While the raw vote fractions are not directly comparable, the initial analysis indicates that the broad morphologies are at least consistent.

7.3 Comparing GZH morphologies to other catalogues

All of the legacy surveys included in the GZH imaging have had morphological catalogues previously published; these have significant differences in the number of galaxies, size and magnitude limits, classification scheme, and the methods used for measuring morphology. These catalogues have been cross-matched to GZH to compare results.

The types and accuracy of morphological classification strongly depend on the sample and methods being used. In an attempt to make a consistent comparison between different techniques, galaxies are broadly grouped into three categories:

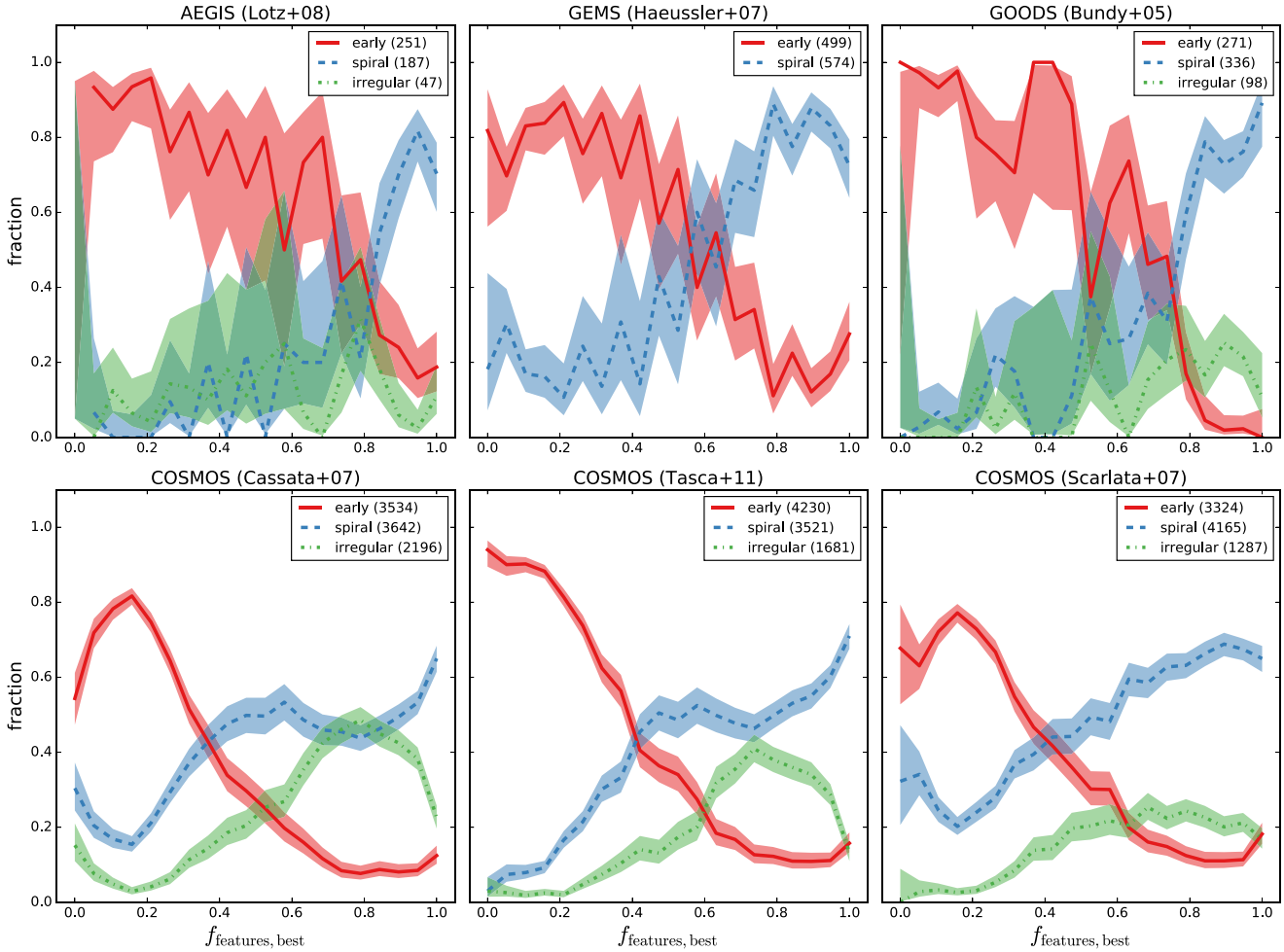

Figure 14. Distributions of morphological parameters for a volume-limited sample ($z < 1.0$, $m < 22.5$) of galaxies matched between GZH and six published morphological catalogues, split by survey (AEGIS, COSMOS, GEMS, and GOODS). This plot shows the fraction of overall galaxies for each of the external morphological categories as a function of $f_{\mathrm{features,best}}$ in GZH. The shaded regions around the binned fractions are confidence intervals calculated for a binomial population (Cameron 2011).

bulge-dominated/elliptical/smooth, disc-dominated/spiral, and irregular/clumpy. These categories are matched to two GZH parameters: $f_{\mathrm{features,best}}$, which identifies disc-dominated and clumpy galaxies, and f_{odd}, which identifies deviations from well-formed spirals or S0s and which constitutes a 'catch-all' for the variety of asymmetric morphologies that can constitute an irregular galaxy.

The comparison and analysis of the GZH morphologies is done on a set of matched, volume-limited samples at $z < 1.0$ and $m_{I|i|z} < 22.5$ mag. The redshift and magnitude limits for the sample are chosen to match the shared constraints for the shallowest depths (GEMS; Bundy, Ellis & Conselice 2005) and the limits on morphological reliability (COSMOS/ZEST; Scarlata et al. 2007) in the comparison catalogues. We match all catalogues against GZH using a positional radius of 0.5 arcsec.

Fig. 14 shows the proportion of galaxies as split by their automated/expert visual morphologies for each of the six catalogues matched to GZH. Galaxies from every catalogue show a strong correlation between $f_{\mathrm{features,best}}$ and the fraction of galaxies identified as spirals, with a corresponding anti-correlation between $f_{\mathrm{features,best}}$ and the fraction of ellipticals.

AEGIS galaxies were morphologically classified using non-parametric measurements by Lotz et al. (2008). Their method used a combination of the Gini coefficient (G), which measures the relative inequality in pixel brightness, and M_{20}, the second-order moment of the brightest 20 per cent of the light (Lotz et al. 2004). A linear combination of G and M_{20} delineates three broad categories of galaxy morphology: E/S0/Sa ('elliptical'), Sb/Sc/Ir ('spiral'), and mergers ('irregular'). We limit the AEGIS galaxies to those with reliably measured morphologies and S/N > 3 in both the V and I bands in Lotz et al. (2008).

The AEGIS sample had only 485 total galaxies matched with GZH, and so the statistical uncertainties in the analysis are higher. However, at $f_{\mathrm{features,best}} > 0.8$ there is a clear separation of featured and unfeatured galaxies in Fig. 14, though this is not seen at lower $f_{\mathrm{features,best}}$.

Morphologies for GEMS galaxies were measured by Häußler et al. (2007), who used single-component Sérsic fits to the *F*850*LP* imaging. There are 8846 galaxies with measurements in both Häußler et al. (2007) and GZH. The primary morphological parameter in the automated catalogue is the Sérsic index n defining the radial surface brightness profile. 'Elliptical' galaxies are selected by $n > 2.5$ and 'spirals' by $n < 2.5$. There is no automatic measurement of irregular or clumpy structure.

Galaxies from GEMS also have a strong correlation with $f_{\mathrm{features,best}}$ (Fig. 14), which is somewhat surprising (but encouraging) considering the known limits of assigning galaxy morphology

based only on a single Sérsic parameter. Visual inspection of images of the exceptions – where both $f_{\rm features,best}$ and n are high – shows that most are obvious spirals but with prominent bulges. This means that the single-component Sérsic fit is likely choosing too small of an effective radius and missing the extended disc structure for a large population of galaxies. This is a known issue with using n as a stand-alone measure of galaxy morphology – it is relatively common, for example, to measure high n for an object with an exponential disc but a central cusp in the light profile (e.g. due to an AGN, cluster, foreground star, noise, etc.).

Galaxies in both of the GOODS fields down to a limit of $z_{\rm AB} = 22.5$ were visually classified by a single expert (R. S. Ellis), inspecting both z-band and composite *Viz* colour images (Bundy et al. 2005). These morphologies are assigned a numerical value based on categories in Brinchmann et al. (1998): the corresponding morphologies used are 'elliptical' (classes 0,1,2), 'spirals' (classes 3,4,5), and 'irregular' (classes 6,7,8).

The GOODS galaxies show a very similar trend to the AEGIS images in Fig. 14, with nearly 100 per cent pure samples at the lowest and highest values of $f_{\rm features,best}$. We conjecture that the strong agreement is at least partially due to the shared method of visual classification, and reinforces the findings from several phases of Galaxy Zoo that crowdsourced classifications can be competitive with dedicated experts.

COSMOS galaxies have multiple published data sets automatically classifying morphology, all using a variation of non-parametric measurements. Cassata et al. (2007) used a combination of concentration (C), asymmetry (A), G, and M_{20}. Cassata et al. (2005) make empirical divisions based on several hundred training images to assign galaxy morphology. A similar method was employed by Tasca (2011), using the same non-parametric indices but with a different method of calculating the Petrosian radius and total light profile, and using a nearest-neighbours method to label morphologies. Scarlata et al. (2007, ZEST) used C, A, G, M_{20}, the galaxy ellipticity (ϵ), and Sérsic index (n) to quantify morphology; a principal component analysis is used to assign galaxies to categories. All three COSMOS catalogues use discrete morphological categories of ellipticals, spirals, and peculiar/irregulars. We use the same categories with the exception of assigning S0/bulge-dominated discs (type 2.0) from ZEST into the unfeatured category.

Galaxies from the COSMOS catalogues all have morphologies that are reasonably well predicted by the GZH based on $f_{\rm features,best}$. The relation is strongest for early-type galaxies, since the comparable fractions of irregulars and spirals show that clumpy/asymmetric morphologies are clearly also being captured by the $f_{\rm features,best}$ parameter. The small drop-off in the fraction of early-types for the Cassata and ZEST samples at $f_{\rm features,best}<0.1$ in Fig. 14 is somewhat puzzling, but could potentially be due to the presence of a small population of extremely compact objects (stars or quasars) that would be classified as artefacts in GZH.

A detailed comparison of all external catalogues is outside the scope of this paper, but we emphasize only that it is impossible for any catalogue to be fully consistent with *all* of the previously published data sets [for example, only 64 per cent of the galaxies in Tasca (2011) have the same morphological class as the same galaxies in Cassata et al. (2007)]. We interpret the broad agreements between early-types, spirals, and irregulars between GZH and all three methods of automated classification as a validation of both the crowdsourced visual classifications provided by GZH and those automated methods.

It has been proposed that automated methods are preferable to visual classification due to being quantified and more easily reproducible (e.g. Conselice 2014; Pawlik et al. 2016). We argue that this distinction applies only to visual classification done by individual and/or small numbers of experts. The technique of crowdsourcing using a large number of independent classifiers provides visual classification which is both quantified and reproducible, and (as demonstrated in this section) broadly agrees with all the automated methods and expert visual classifications.

We repeated all analyses comparing morphologies between GZH and external publications on the entire GZH data set, rather than the volume-limited sample ($z < 1.0$, $m < 22.5$) discussed above. If galaxies at all redshifts and at the original magnitude limit of $m < 23.5$ are used, the correlations between all of the automatic morphological populations and GZH decrease significantly. The exceptions are the GEMS galaxies, which are the only sample that was visually classified and had a shallower magnitude cut on their catalogue. Based on these comparisons, we suggest that use of morphological catalogues with these *HST* images be limited to galaxies with $m_{I[i]z} < 22.5$; fainter targets have been demonstrated to be less reliable, regardless of the method of morphological classification being used.

8 SUMMARY

This paper presents the catalogue release for the GZH project, which uses crowdsourced visual classifications to measure galaxy morphologies. The first two phases of Galaxy Zoo (Lintott et al. 2011; Willett et al. 2013) used images of low-redshift galaxies from SDSS; this is the first result of the project with space-based images of high-redshift targets (in addition to the Galaxy Zoo: CANDELS collaboration; Simmons et al. 2016). The final sample includes classifications for 189 149 images generated from 145 741 unique galaxies (of which 115 402 are in *HST* legacy imaging).

Galaxies were selected from a brightness-limited sample from multiple legacy surveys using the ACS on the *HST*, including AEGIS, GEMS, GOODS-N, GOODS-S, and COSMOS. The catalogue also includes classifications for SDSS images from Stripe 82 at $z < 0.25$; these images serve both as a low-redshift anchor for cosmological studies and a potential comparison for the different epochs of classification between GZH and Galaxy Zoo 2 (Willett et al. 2013).

The data in the GZH catalogues have been extensively reduced and tested. A dominant effect is the known bias against identifying discy and asymmetric substructures at either low resolution or surface brightness. This can be the result either of genuinely small or dim galaxies, or a perceived effect from observing galaxies at further distances (higher redshift). To calibrate this *without* potentially overcorrecting for the genuine morphological evolution of galaxies over cosmic time, the GZH project uses SDSS images of low-redshift galaxies, processes them to appear as if they were at higher redshift, and classifies them through the GZH interface in an identical fashion. The resulting change in $f_{\rm features}$ as a function of z and μ is applied as a multiplicative correction to the top-level vote fractions for ∼50 per cent of the GZH galaxies. However, any population studies using GZH morphological data must use a combination of debiased vote fractions and upper limits from the 'uncorrectable' galaxies, due to the evolving nature of the thresholds as a function of redshift.

Galaxies in GZH show significant changes in the disc/elliptical fraction as a function of redshift, along with an increasing number of galaxies dominated by smaller clumps and presumed to be in the process of building up their baryonic mass through a combination of hierarchical merging and *in situ* star formation. While the majority of scientific interpretation is left to future work, this paper confirms

the decrease in observed bar fraction with increasing redshift and identifies a new way of selecting clumpy galaxies as a function of clump multiplicity.

The full data tables for the catalogues can be accessed in machine-readable form from both the journal and at http://data.galaxyzoo.org. All the code and data tables used to generate this manuscript can be found at https://github.com/willettk/gzhubble.

ACKNOWLEDGEMENTS

KW, MG, CS, MB, and LF gratefully acknowledge support from the US National Science Foundation Grant AST1413610. Support for BDS was provided by the National Aeronautics and Space Administration through Einstein Postdoctoral Fellowship Award Number PF5-160143 issued by the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of the National Aeronautics Space Administration under contract NAS8-03060. KS gratefully acknowledges support from Swiss National Science Foundation Grant PP00P2_138979/1. TM acknowledges funding from the Science and Technology Facilities Council ST/J500665/1. RJS is supported by the STFC grant code ST/K502236/1. The development and hosting of Galaxy Zoo: Hubble was supported by a grant from the Alfred P. Sloan Foundation. The Zooniverse acknowledges support from a Google Global Impact Award.

We thank Meg Schwamb and the ASIAA for hosting the 'Citizen Science in Astronomy' workshop, 2014 March 3–7 in Taipei, Taiwan, at which some of this analysis was initiated. We also thank Jennifer Lotz for sharing her G–M_{20} measurements for the AEGIS sample. We thank Coleman Krawczyk for his assistance in producing Fig. 4. We thank Nathan Cloutier and Brent Hilgart for useful discussions. We also thank the referee for thoughtful comments which improved the quality of this paper.

This project made heavy use of the ASTROPY packages in PYTHON (Astropy Collaboration et al. 2013), the SEABORN plotting package (Waskom et al. 2015), astroML (Vanderplas et al. 2012), and TOPCAT (Taylor 2005, 2011). Modified code from Nick Wherry and David Schlegel was used to create the JPG images. Fig. 13 was generated with http://sankeymatic.com/. Holwerda (2005) provided valuable assistance in interpreting SEXTRACTOR output.

This work is based on (GO-10134, GO-09822, GO-09425.01, GO-09583.01, GO-9500) programme observations with the NASA/ESA *Hubble Space Telescope*, obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS website is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Abraham R. G., Merrifield M. R., Ellis R. S., Tanvir N. R., Brinchmann J., 1999, MNRAS, 308, 569
Abraham R. G., van den Bergh S., Nair P., 2003, ApJ, 588, 218
Astropy Collaboration et al., 2013, A&A, 558, A33
Athanassoula E., 2012, MNRAS, 426, L46
Athanassoula E., Lambert J. C., Dehnen W., 2005, MNRAS, 363, 496
Baillard A. et al., 2011, A&A, 532, A74
Bamford S. P., Rojas A. L., Nichol R. C., Miller C. J., Wasserman L., Genovese C. R., Freeman P. E., 2008, MNRAS, 391, 607
Bamford S. P. et al., 2009, MNRAS, 393, 1324
Barden M., Jahnke K., Häußler B., 2008, ApJS, 175, 105
Barden M., Häußler B., Peng C. Y., McIntosh D. H., Guo Y., 2012, MNRAS, 422, 449
Barnes J. E., Hernquist L. E., 1991, ApJ, 370, L65
Beckwith S. V. W. et al., 2006, AJ, 132, 1729
Behrendt M., Burkert A., Schartmann M., 2016, ApJ, 819, L2
Bell E. F. et al., 2012, ApJ, 753, 167
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
Buta R. J., 2013, Galaxy Morphology. Springer-Verlag, Berlin, p. 1
Caldwell J. A. R. et al., 2008, ApJS, 174, 136
Cameron E., 2011, PASA, 28, 128
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
Cassata P. et al., 2005, MNRAS, 357, 903
Cassata P. et al., 2007, ApJS, 172, 270
Cheung E. et al., 2015, MNRAS, 447, 506
Chevance M., Weijmans A.-M., Damjanov I., Abraham R. G., Simard L., van den Bergh S., Caris E., Glazebrook K., 2012, ApJ, 754, L24
Combes F., 2009, in Jogee S., Marinova I., Hao L., Blanc G. A., eds, ASP Conf. Ser. Vol. 419, Galaxy Evolution: Emerging Insights and Future Challenges. Astron. Soc. Pac., San Francisco, p. 31
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, ApJ, 660, L1
de Vaucouleurs G., 1959, Handbuch Phys., 53, 275
Dressler A., 1980, ApJ, 236, 351
Elmegreen D. M., Elmegreen B. G., 2014, ApJ, 781, 11
Elmegreen B. G., Elmegreen D. M., Hirst A. C., 2004, ApJ, 612, 191
Elmegreen D. M., Elmegreen B. G., Rubin D. S., Schaffer M. A., 2005, ApJ, 631, 85
Elmegreen D. M., Elmegreen B. G., Ferguson T., Mullan B., 2007, ApJ, 663, 734
Elmegreen D. M., Elmegreen B. G., Marcus M. T., Shahinyan K., Yau A., Petersen M., 2009, ApJ, 701, 306
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
Fortson L. et al., 2012, Galaxy Zoo: Morphological Classification and Citizen Science. CRC Press, Boca Raton, FL, p. 213
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. et al., 2015, MNRAS, 448, 3442
Genel S. et al., 2014, MNRAS, 445, 175

Giavalisco M., 2012, VizieR Online Data Catalog, 2318
Giavalisco M. et al., 2004, ApJ, 600, L93
Griffith R. L. et al., 2012, ApJS, 200, 9
Grogin N. A. et al., 2011, ApJS, 197, 35
Guo Y. et al., 2015, ApJ, 800, 39
Häußler B. et al., 2007, ApJS, 172, 615
Hinshaw G. et al., 2013, ApJS, 208, 19
Holwerda B. W., 2005, preprint (arXiv:e-prints)
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, New Haven, CT
Ilbert O. et al., 2013, A&A, 556, A55
Jogee S. et al., 2004, ApJ, 615, L105
Kartaltepe J. S. et al., 2015, ApJS, 221, 11
Kaviraj S., 2014a, MNRAS, 437, L41
Kaviraj S., 2014b, MNRAS, 440, 2944
Koekemoer A. M. et al., 2007, ApJS, 172, 196
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Koekemoer A. M., Fruchter A. S., Hook R. N., Hack W., 2002, in Arribas S., Koekemoer A., Whitmore B., eds, The 2002 HST Calibration Workshop: Hubble after the Installation of the ACS and the NICMOS Cooling System. Space Telescope Science Institute, Baltimore, MD, p. 339
Koekemoer A., Fruchter A., Hack W., 2003, Space Telesc. Eur. Coord. Facil. Newsl., 33, 10
Krist J., 1993, in Hanisch R. J., Brissenden R. J. V., Barnes J., eds, ASP Conf. Ser. Vol. 52, Astronomical Data Analysis Software and Systems II. Astron. Soc. Pac., San Francisco, p. 536
Lackner C. N., Gunn J. E., 2012, MNRAS, 421, 2277
Lahav O. et al., 1995, Science, 267, 859
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
Lilly S. et al., 1998, ApJ, 500, 75
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Lintott C. et al., 2011, MNRAS, 410, 166
Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163
Lotz J. M. et al., 2008, ApJ, 672, 177
Loveday J. et al., 2012, MNRAS, 420, 1239
Lupton R., Blanton M. R., Fekete G., Hogg D. W., O'Mullane W., Szalay A., Wherry N., 2004, PASP, 116, 133
Mandelker N., Dekel A., Ceverino D., DeGraf C., Guo Y., Primack J., 2017, MNRAS, 464, 635
Mao S., Mo H. J., White S. D. M., 1998, MNRAS, 297, L71
Masters K. L. et al., 2011, MNRAS, 411, 2026
Melvin T., 2016, PhD thesis, Univ. Portsmouth, Portsmouth
Melvin T. et al., 2014, MNRAS, 438, 2882
Momcheva I. G. et al., 2016, ApJS, 225, 27
Mortlock A. et al., 2013, MNRAS, 433, 1185
Nair P. B., Abraham R. G., 2010, ApJS, 186, 427
Nieto-Santisteban M. A., Szalay A. S., Gray J., 2004, in Ochsenbein F., Allen M. G., Egret D., eds, ASP Conf. Ser. Vol. 314, Astronomical Data Analysis Software and Systems (ADASS) XIII. Astron. Soc. Pac., San Francisco, p. 666
Pawlik M. M., Wild V., Walcher C. J., Johansson P. H., Villforth C., Rowlands K., Mendez-Abreu J., Hewlett T., 2016, MNRAS, 456, 3032
Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266
Pierce C. M. et al., 2010, MNRAS, 405, 718
Rix H.-W. et al., 2004, ApJS, 152, 163
Sánchez S. F. et al., 2004, ApJ, 614, 586
Sandage A., 1961, The Hubble Atlas of Galaxies. Carnegie Institute of Washington, 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
Sheth K. et al., 2008, ApJ, 675, 1141
Silk J., Mamon G. A., 2012, Res. Astron. Astrophys., 12, 917
Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11
Simmons B. D., Urry C. M., 2008, ApJ, 683, 644
Simmons B. D., Van Duyne J., Urry C. M., Treister E., Koekemoer A. M., Grogin N. A., GOODS Team, 2011, ApJ, 734, 121
Simmons B. D. et al., 2013, MNRAS, 429, 2199
Simmons B. D. et al., 2014, MNRAS, 445, 3466
Simmons B. D. et al., 2016, MNRAS, preprint (arXiv:1610.03070)
Skibba R. A. et al., 2009, MNRAS, 399, 966
Skibba R. A. et al., 2012, MNRAS, 423, 1485
Smethurst R. J. et al., 2015, MNRAS, 450, 435
Smethurst R. J. et al., 2016, MNRAS, 463, 2986
Steinmetz M., Navarro J. F., 2002, New Astron., 7, 155
Stetson P. B., 1987, PASP, 99, 191
Strauss M. A. et al., 2002, AJ, 124, 1810
Taniguchi Y. et al., 2007, ApJS, 172, 9
Tasca L. A. M., 2011, VizieR Online Data Catalog, 7265
Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, ASP Conf. Ser. Vol. 347, Astronomical Data Analysis Software and Systems XIV. Astron. Soc. Pac., San Francisco, p. 29
Taylor M., 2011, Astrophysics Source Code Library, record ascl:1101.010
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 Proceedings of Conference on Intelligent Data Understanding (CIDU), Introduction to astroML: Machine learning for astrophysics, p. 47
Vogelsberger M. et al., 2014, MNRAS, 444, 1518
Waskom M. et al., 2015, seaborn: v0.6.0 (June 2015), Available at http://dx.doi.org/10.5281/zenodo.19108
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

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table 2. GZH morphological classifications for *HST* images from AEGIS, COSMOS, GEMS, and GOODS.
Table 5. GZH morphological classifications for colour-faded *HST* images.
Table 6. GZH morphological classifications for colour-inverted *HST* images.
Table 7. GZH morphological classifications for GOODS two-epoch images.
Table 8. GZH morphological classifications for SDSS Stripe 82 single-epoch images.
Table 9. GZH morphological classifications for SDSS Stripe 82 co-added images.
Table 10. GZH morphological classifications for *HST* images with simulated AGN.
(http://www.mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stw2568/-/DC1).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: GOODS SHALLOW-DEPTH DATA

GZH used both five-epoch and two-epoch sets of data to construct the GOODS set of images. The 11 157 full-depth five-epoch images are used in the `main` catalogue; the classifications for the 6144 `goods-shallow` images are provided as a supplementary table. This section analyses the effect of image depth on the ability of the Galaxy Zoo classifiers to identify features or disc structure in the images.

A1 Comparing shallow- and full-depth morphologies

Of the 11 157 galaxies in the GOODS-N and GOODS-S full-depth sample, 4460 are also in the shallow-depth sample. Fig. A1 shows a strong correlation between $f_{\rm features}$ for both sets of images. The mean change in $f_{\rm features}$ from the shallow- to full-depth images $f_{\rm features,\ full} - f_{\rm features,\ shallow} \equiv \Delta f = -0.01$, with a standard deviation of $\sigma = 0.18$.

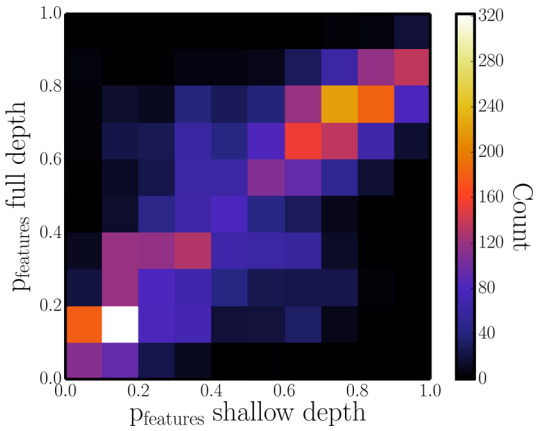

Figure A1. Distribution of $f_{\rm features}$ for the 4460 GOODS galaxies with both shallow-depth (two-epoch; `goods-shallow`) and full-depth (five-epoch; `main`) images morphologically classified in GZH. For most galaxies, the value of $f_{\rm features}$ is consistent ($\Delta f_{\rm features} < 0.2$) between depths. Examples of galaxies with sharp changes in $f_{\rm features}$, as well as those with little to no change, are shown in Figs A2–A4.

Table A1. Correctable fractions for the top-level task in GZH in the GOODS shallow-depth (two-epoch) images.

	GOODS-N	GOODS-S	Total
Correctable	1051	730	1781
Lower limit	131	334	465
No correction needed ($z \leq 0.3$)	267	267	534
NEI	943	2078	3021
No redshift information	159	184	343
Total	2551	3593	6144

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

	N	per cent	$\langle \Delta f \rangle$	$\langle z \rangle$
Smooth to smooth	708	15.9	0.00	0.72
Smooth to intermediate	346	7.8	0.21	0.70
Smooth to featured	6	0.1	0.80	0.45
Intermediate to smooth	266	6.0	−0.24	0.66
Intermediate to intermediate	2363	53.0	0.00	0.75
Intermediate to featured	121	2.7	0.17	0.82
Featured to smooth	3	0.1	−0.73	0.71
Featured to intermediate	370	8.3	−0.15	0.69
Featured to featured	277	6.2	−0.05	0.71
Total	4460	100		

While there is some variance in Δf in the whole sample, the change is usually small and not often significant enough to change a morphological classification. Defining a clean sample of disc galaxies as those with $f_{\rm features,best} > 0.8$, elliptical galaxies as those with $f_{\rm smooth,\ best} < 0.2$, and intermediate as those in between, 75 per cent of the sample would not change morphology. Of the remaining 25 per cent that would change morphology, only 0.2 per cent (nine galaxies) drastically change morphology either from smooth to featured or vice versa, while the rest transition to or from the 'intermediate' morphology. Details can be seen in Table A2, and examples of images representing the nine possible changes (or lack of) in morphology are shown in Figs A2–A4.

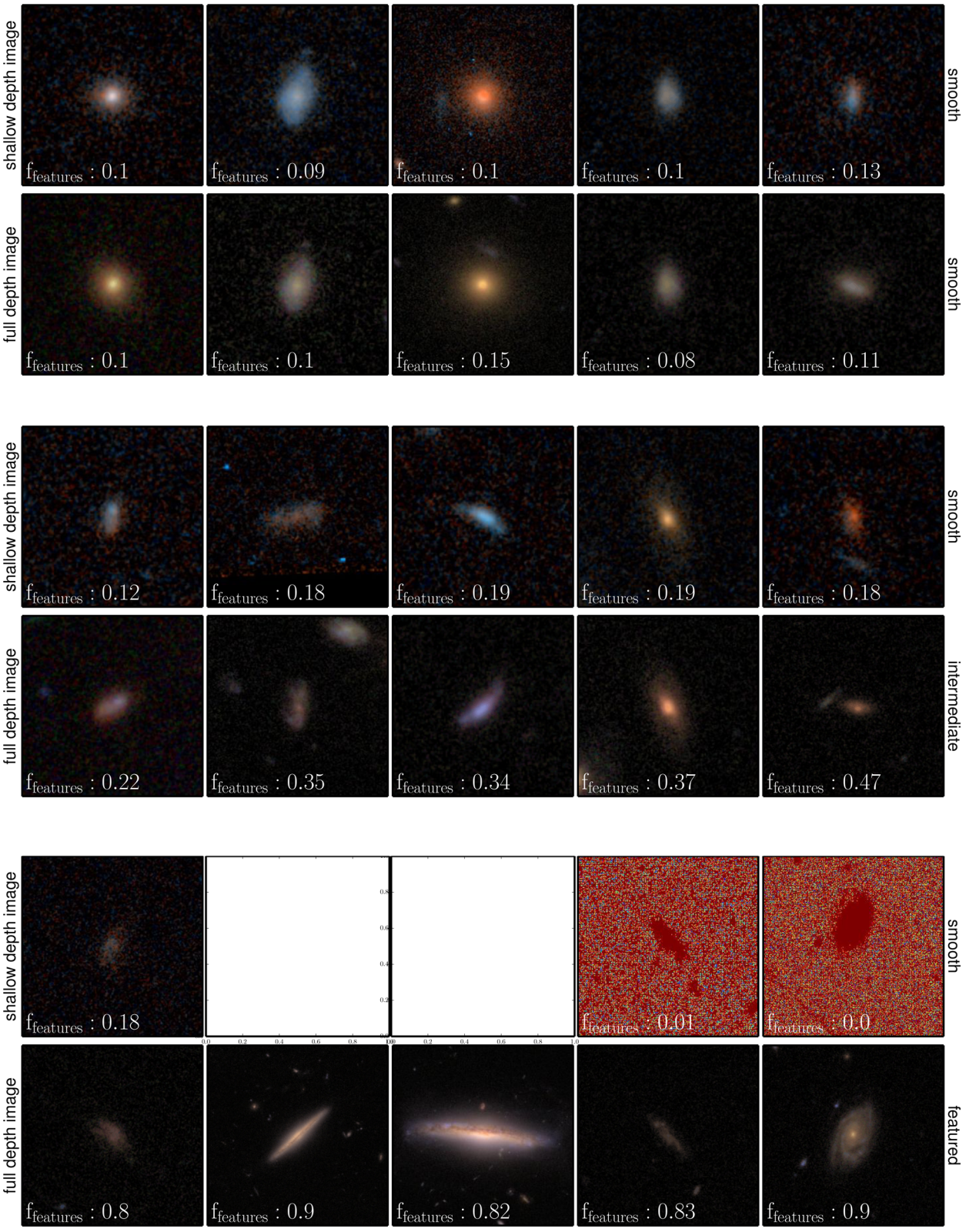

Figure A2. Example images of GOODS galaxies that compare GZH morphological classifications between their two-epoch and five-epoch imaging. The top set was classified as 'smooth' in both the two- and five-epoch imaging. The middle set was classified as 'smooth' in the two-epoch imaging and as intermediate between 'smooth' and 'featured' in the five-epoch imaging. The bottom set was classified as 'smooth' in the two-epoch imaging and 'featured' in the five-epoch imaging (there are only seven such images in the sample).

Figure A3. Similar to Fig. A2. The top set (a) was classified as 'intermediate' in the two-epoch imaging and 'smooth' in the five-epoch imaging. The middle set (b) was classified as intermediate in both the two- and five-epoch imaging. The bottom set (c) was classified as 'intermediate' in the two-epoch imaging and 'featured' in the five-epoch imaging.

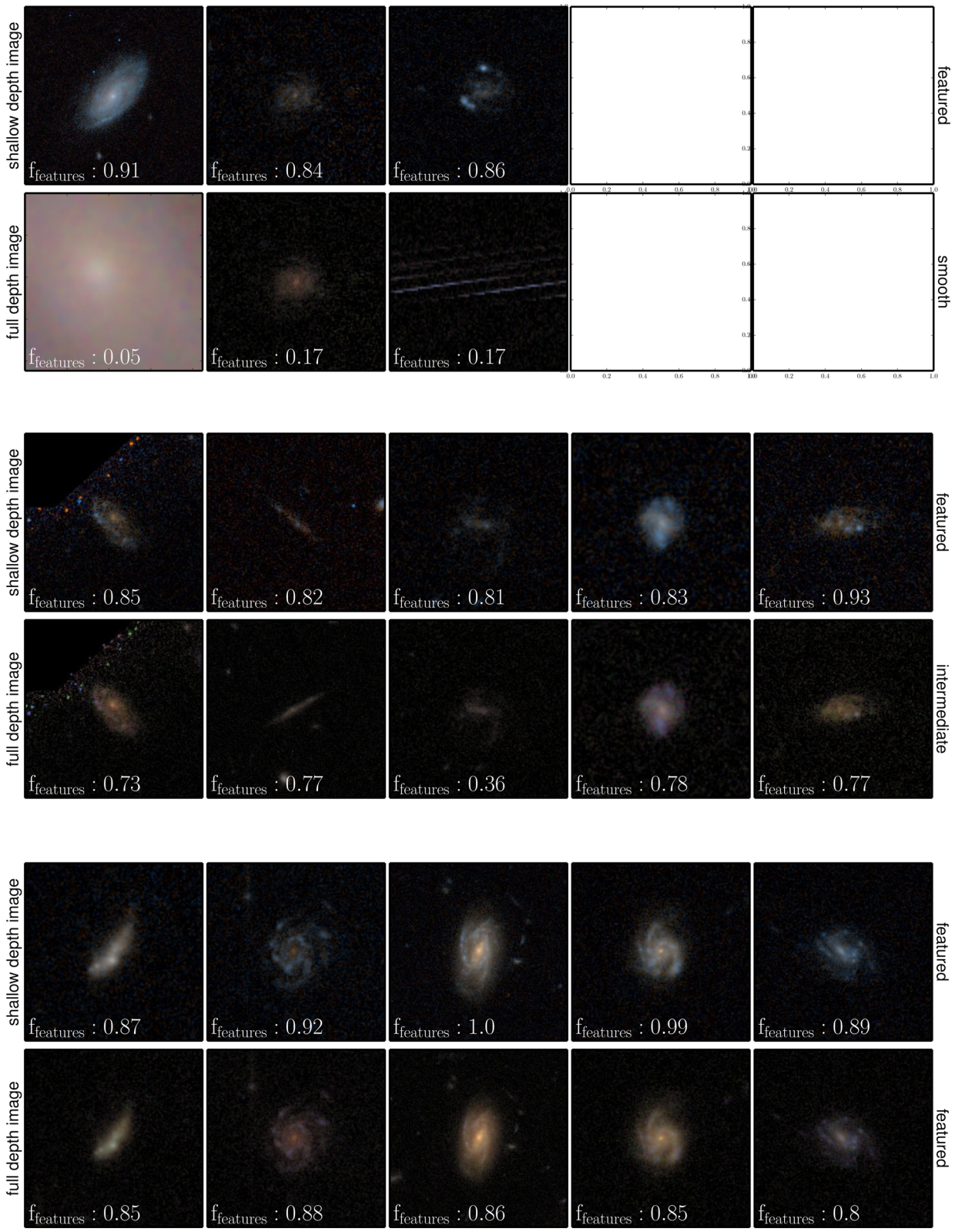

Figure A4. Similar to Fig. A2. The top set (a) was classified as 'featured' in the two-epoch imaging and 'smooth' in the five-epoch imaging (there were only six such images in the sample). The middle set (b) was classified as 'featured' in the two-epoch imaging and 'intermediate' in the five-epoch imaging. The bottom set (c) was classified as 'featured' in both the two- and five-epoch imaging.

[1]*School of Physics and Astronomy, University of Minnesota, 116 Church St SE, Minneapolis, MN 55455, USA*
[2]*Department of Physics and Astronomy, University of Kentucky, 505 Rose St, Lexington, KY 40506, USA*
[3]*School of Physics and Astronomy, The University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[4]*Oxford Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK*
[5]*Institute for Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Burnaby Road, Portsmouth PO1 3FX, UK*
[6]*SEPnet, South East Physics Network, UK*
[7]*Center for Astrophysics and Space Sciences, Department of Physics, University of California, San Diego, CA 92093, USA*
[8]*Department of Mathematics and Science, Wheelock College, Boston, MA 02215, USA*
[9]*Kavli Institute for the Physics and Mathematics of the Universe (WPI), Todai Institutes for Advanced Study, The University of Tokyo, Kashiwa 277-8583, Japan*
[10]*Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA*
[11]*Department of Astronomy & Astrophysics, 525 Davey Lab, The Pennsylvania State University, University Park, PA 16802, USA*
[12]*Centre for Astrophysics, Science & Technology Research Institute, University of Hertfordshire, Hatfield, AL10 9AB, UK*
[13]*European Southern Observatory, Alonso de Córdova 3107, Vitacura, Casilla 19001, Santiago, Chile*
[14]*Department of Astronomy, Yale University, New Haven, CT 06520, USA*
[15]*Adler Planetarium, 1300 S. Lake Shore Drive, Chicago, IL 60605, USA*
[16]*Institute for Astronomy, Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland*
[17]*GitHub, 88 Colin P Kelly Jr St, San Francisco, CA 94107*

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