INITIAL ANALYSIS OF CROWDSOURCED MORPHOLOGICAL CLASSIFICATIONS OF DECALS IMAGES

K.W. Willett[1] et al.
[1]University of Minnesota
Draft version February 10, 2016

ABSTRACT

Abstract.
Keywords: keywords

1. SCIENCE CASE

The morphology of galaxies has long been a critical parameter in studying their formation and evolution over cosmic time. This includes basic divisions between early-type, late-type, and merging galaxies as well as more detailed measurements of bulge/disk ratio, spiral arm angle and orientation, and fainter features such as tidal tails or dust lanes. The former parameters measure the large scale integrated dynamical history of the galaxy, as well as its history of star formation (Schawinski et al. 2014), while the latter categories typically probe secular processes acting on longer timescales.

Automated measurements of galaxy morphology, however, are not yet sufficiently accurate to be used on large scale surveys, especially for faint targets and features with small angular sizes. The Galaxy Zoo project has provided reliable morphologies for several large surveys (including SDSS, UKIDSS, COSMOS, and CANDELS) by leveraging crowdsourced visual classifications into quantitative probabilities. This technique shows each image to dozens of people and seeks consensus answers, using our knowledge of user reliability to weight their responses, and has been proven (Lintott et al. 2011; Willett et al. 2013) to be as reliable (and in many cases better) that state-of-the art automated methods. By using the DECaLS images in the existing Galaxy Zoo platform, the team will provide morphologies that complement the other globally measured host galaxy properties such as colour and luminosity. Galaxy Zoo has a strong track record in publication from Galaxy Zoo morphologies, with 50 publications which collectively have ~ 2000 citations. Some of our science highlights include the discovery of new types of galaxies (e.g. green peas; Cardamone et al. 2009), as well as the investigation of red spirals (Bamford et al. 2009; Masters et al. 2010) and blue ellipticals (Schawinski et al. 2009).

The advantage of the DECaLS images over existing GZ measurements is primarily driven by two aspects: namely, the deeper imaging (especially in g- and z-bands, but also r-band), as well as improved spatial resolution compared to the SDSS which has provided the bulk of the low-redshift GZ images to date.

In addition, the Dark Energy Spectroscopic Instrument (DESI; Levi et al. 2013) is planning to use DECaLS for targets as part of major new galaxy redshift surveys covert 14000 $\deg^2$ of the northern sky. The low redshift part of this is the Bright Galaxy Sample (BGS) which will use bright time to observe 10 million nearby galaxies (median redshift of $z = 0.2$) from a simple magnitude-limited sample of $r < 19.5$. Galaxy Zoo morphologies of BGS galaxies will provide several scientific opportunities, especially studying the clustering (real and redshift space) of galaxies as a function of their

detailed morphologies. Such measurements will be significant improvements on existing SDSS-based GZ measurements (see Skibba et al. 2009, 2012) as it will provide greater statistics, larger scales and finer morphologies (e.g. better bar and bulge classifications). Moreover, morphologies may offer a new and unique way of helping to characterise the halo occupation statistics of galaxies thus assisting the detailed modeling of the DESI galaxies as a function of galaxy properties e.g., bright, massive, ellipticals are usually central halo galaxies. Moreover, morphologies may offer a new and unique way to help characterize the halo occupation statistics of galaxies, thus assisting the detailed modelling of the DESI galaxies as a function of galaxy properties, since bright massive ellipticals are usually central halo galaxies. In addition, they will enable analyses of the morphology-halo mass relation of galaxies with better precision than previous surveys.

We list below four example science cases that can be addressed using the combination of DECaLS imaging/metadata and GZ crowdsourced morphologies.

1. **Statistical studies of galaxy evolution with morphology** *(Lead GZ scientists: Lintott, Masters, Skibba, Schawinski, Willett).* The complexity of galaxy evolution, along with the development of large surveys, like SDSS has driven forward the technique of statistical galaxy evolution in recent years, revolutionising our understanding of the galaxy population, and revealing the complex interdependence of galaxy properties such as mass, environment, dynamics, morphology and their star formation histories. The multi-dimensional correlations common in galaxy evolution (e.g. between stellar mass, colour, environment and morphology) makes larger samples vital to disentangle the primary variables. Galaxy Zoo has enable the addition of quantitative visual morphology to these techniques, and the involvement of a large number of Galaxy Zoo team members in this project underscores how this covers the core GZ:DECaLS science). Our previous work with GZ:SDSS has found that samples of $N \simeq 10^4$ may be needed for even a detection of some effects (Skibba et al. 2012). GZ:DECaLS data (especially when combined with GZ:DES, a proposal for which is under review within DES) will reduce the statistical uncertainty in such measurements. Questions we will address are:

   - How does star formation history correlate with morphology as a function of stellar mass and environment? Schawinski will provide significant expertise on constraining galaxy star formation histories and their links to morphology, also see Smethurst et al. (2015).

- Are bars triggered or destroyed by galaxy interactions in different environments? Follows on previous work by Skibba, Masters to greater statistical significance (Skibba et al. 2012; Masters et al. 2011, 2012).

- What is the contribution of bar dynamics to the feeding of AGN? e.g. an extension of Galloway et al. (2015) to greater statistical significance.

2. **Discovery of rare objects** *(Lead GZ Scientists: Keel, Maksym, Simmons and Lintott).* The Galaxy Zoo experience with the SDSS data showed the value of having classifiers able to note objects which fell outside the normal range of galaxy structure, both as regards color and morphology; the greater surface brightness sensitivity, in particular, the DECaLS data promises to similarly reward broad inspection. Galaxy Zoo is perhaps best well known by the public for its discovery of a variety of unusual objects with very low spatial densities. The most famous is "Hanny's Voorwerp" (an AGN-ionized gas cloud) (Lintott et al. 2009), but GZ has also discovered other classes of rare objects, including smaller versions of the Voorwerp (Keel et al. 2012), the "green peas" (compact star forming galaxies; Cardamone et al. 2009), bulgeless galaxies with AGN (Simmons et al. 2013), and overlapping galaxies for studying dust content (Keel et al. 2013). We expect that new classes of objects that may be revealed by deeper imaging, and intend to not only use catalogues on known galaxies in GZ:DECaLS, but also include Tractor identification of extended objects which are not otherwise catalogued.

3. **Galaxy morphology at low luminosities** *(Lead GZ scientists: Bamford and Willett).* Visual morphologies for large, low-redshift samples have so far been limited by the depth and resolution of SDSS and UKIDSS. Studies that have pushed down to low luminosities have therefore been restricted to small volumes and correspondingly small samples. For example, the GAMA survey study by Kelvin et al. (2014) studies only 3727 galaxies down to $M_r < -17.4$ mag, as it is limited to redshifts below 0.06. GAMA will soon gain greater depth with VST-KiDS imaging, and has complete spectroscopic coverage, but covers only 250 deg^2. DECaLS will reach similar depths, but over a much larger area. GZ:DECaLS data will therefore greatly improve upon both the quality and number of classifications at fainter luminosities. Combining results from these complementary surveys will enable us to make a definitive morphological census of the "faint end of the Hubble sequence", where the galaxy population is dominated by very late-type disks, irregulars and dwarf spheroidals.

4. **Exploring the role of minor mergers** *(Lead GZ Scientist: Kaviraj).* Minor merging is a fundamental process that is thought to drive around half of the local star formation budget ($\sim$ 40% in local spirals) (Kaviraj 2014). However, this process is poorly explored because minor mergers produce only faint tidal features that are invisible in typical imaging surveys (e.g. SDSS). We will use the deep DECaLS images to select minor merger remnants and study this process in detail in the local and intermediate redshift Universe ($z < 0.5$). Minor merger remnants will be selected by identifying spiral galaxies (i.e. galaxies which cannot have had a major merger) that are morphologically disturbed (so must have had a minor merger). This sample will be used to answer the following questions:

- What is the minor merger rate at $z = 0$ and does it show an evolution to intermediate redshift?

- Do minor mergers trigger AGN activity? What role do they play in the low and high excitation modes in active galaxies?

- How does the minor merger rate depend on stellar mass and environment?

- What is the star formation enhancement due to minor merging in the host galaxies? Does this enhancement show any evolution with redshift?

- What fraction of the star formation budget is driven by minor merging in the nearby Universe?

The methodology for this project has already been developed in Kaviraj (2014) who performed an identical study using Stripe 82 (see Figure 1 in that paper for examples of deep images that will allow us to select minor merger remnants).

2. SELECTION CRITERIA

Galaxy Zoo 2 — main spectroscopic sample

- Galaxy in MGS or Stripe 82

- spectroscopic redshift available

- $0.0005 < z_{spec} < 0.25$

- $m_r < 17.0$

- `petro90_r` $> 3''$

- flag is not SATURATED, BRIGHT, or BLENDED

Galaxy Zoo 2 — Stripe 82 coadd

- version 1 of coadded data

- same as MGS with exception of $m_r < 17.77$

Dark Energy Camera Legacy Survey (DECaLS)

- Galaxies in NASA-Sloan (Extended) Atlas

- Good-quality measurements in g, r, z bands as of Jun 2015

- NSA: `brick_primary` = 1

- NSA: `decam_anymask` = 0

- NSA: `decam_nobs` ≥ 1

- $\theta_r > 10''$

- no magnitude cutoff

So the Sloan and DECaLS samples definitely were not selected in the same manner. However, all the relevant parameters are stored in the NASA-Sloan Atlas if we wanted to cut on that. More specifically, we can do a simple match (ideally through previously matched catalogs, but if necessary through a tight positional match) to directly compare morphological measurements for *the same galaxies* (Table 2).

There are 32,429 images of galaxies in GZ-DECaLS.

3. DATA

There is some overlap between the existing morphological classifications from Galaxy Zoo 2 (Willett et al. 2013). Most of the difference comes from the fact that SDSS was located at Apache Point Observatory in the Northern Hemisphere (latitude 32.780278°), while the DECaLS camera is mounted on the Blanco 4-m telescope at CTIO in the Southern Hemisphere (−30.169661°). They can cover a significant fraction of the same portion of the sky, but DECaLS will be limited at high northerly latitudes. Figure 1 shows the overlap between the NASA-Sloan ATLAS[1] (derived from SDSS) and DECaLS DR1.

Morphologies for this analysis are taken from the published GZ2 tables in Willett et al. (2013) for SDSS. The DECaLS morphologies have been collated and weighted, but not systematically debiased to account for changes in morphological fraction as a function of apparent size and brightness. Therefore, we only compare *weighted vote fractions* ($f_{w,morph}$) for galaxies in each sample.

In the GZ2 main spectroscopic sample (243,500 galaxies), we matched galaxies within a $3''.0$ radius and find 9,281 subjects appearing in both catalogs. We match a further 5,800 subjects using the same radius for the Stripe 82 data. There is overlap of 2,814 bright Stripe 82 galaxies that are included in both. The unique total of 12,267 subjects is only 38% of the GZ-DECaLS catalog, despite the fact that the original SDSS Legacy sky coverage (Strauss et al. 2002) overlaps with all of the current DECaLS bricks.

Part of the mismatch comes from the limited spatial coverage of the Stripe 82 region in SDSS, which only covered a declination range of $-1.26° < \delta < +1.26°$ (Annis et al. 2014). The DECaLS imaging bricks have NSA targets in a larger area, extending between roughly $-2.5° < \delta < +2.5°$ (Figure 2). These are presumably targets imaged in SDSS DR8 or later, since otherwise they would have been included in the original GZ2 selection.

However, there are many DECaLS galaxies in the imaging area covered by the main Legacy survey. Galaxies in DECaLS but *not* GZ2 have $\langle m_r \rangle = 17.27$ mag, $\langle r_{petro} \rangle = 6''.64$, $\langle z \rangle = 0.093$. Galaxies in *both* DECaLS and GZ2 are on average brighter ($\langle m_r \rangle = 16.29$ mag), larger ($\langle r_{petro} \rangle = 7''.99$), and lower-redshift ($\langle z \rangle = 0.080$). The vast majority of the DECaLS images with no GZ2 counterpart are galaxies with $17.0 < m_r < 17.77$ — the fainter magnitude limit is that set by the GZ2 main sample, and the brighter was the spectroscopic targeting limit for SDSS (required for a redshift and inclusion in the NSA; Figure 3). The few remaining galaxies brighter than 17.0 mag but not in GZ2 may be the result of positional matching errors, very low-redshift ($z < 0.0005$) galaxies or targets with a large angular size that were shredded in the initial SDSS pipeline.

Summary: roughly 40% of the DECaLS galaxies have morphological measurements from GZ2, and can be used for direct comparison. We believe we understand the reasons for the remainder of DECaLS images that are not matched to GZ2; these will be valuable scientific additions as new targets, and can serve as independent checks on the accuracy of the classifications.

REFERENCES

Annis, J., Soares-Santos, M., Strauss, M. A., et al. 2014, ApJ, 794, 120
Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1324
Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191
Galloway, M. A., Willett, K. W., Fortson, L. F., et al. 2015, MNRAS, 448, 3442
Kaviraj, S. 2014, MNRAS, 437, L41
Keel, W. C., Manning, A. M., Holwerda, B. W., et al. 2013, PASP, 125, 2
Keel, W. C., Lintott, C. J., Schawinski, K., et al. 2012, AJ, 144, 66
Kelvin, L. S., Driver, S. P., Robotham, A. S. G., et al. 2014, MNRAS, 444, 1647
Levi, M., Bebek, C., Beers, T., et al. 2013, ArXiv e-prints, arXiv:1308.0847 [astro-ph.CO]
Lintott, C., Schawinski, K., Bamford, S., et al. 2011, MNRAS, 410, 166
Lintott, C. J., Schawinski, K., Keel, W., et al. 2009, MNRAS, 399, 129
Masters, K. L., Mosleh, M., Romer, A. K., et al. 2010, MNRAS, 405, 783
Masters, K. L., Nichol, R. C., Hoyle, B., et al. 2011, MNRAS, 411, 2026
Masters, K. L., Nichol, R. C., Haynes, M. P., et al. 2012, MNRAS, 424, 2180
Schawinski, K., Lintott, C., Thomas, D., et al. 2009, MNRAS, 396, 818
Schawinski, K., Urry, C. M., Simmons, B. D., et al. 2014, MNRAS, 440, 889
Simmons, B. D., Lintott, C., Schawinski, K., et al. 2013, MNRAS, 429, 2199
Skibba, R. A., Bamford, S. P., Nichol, R. C., et al. 2009, MNRAS, 399, 966
Skibba, R. A., Masters, K. L., Nichol, R. C., et al. 2012, MNRAS, 423, 1485
Smethurst, R. J., Lintott, C. J., Simmons, B. D., et al. 2015, MNRAS, 450, 435
Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810
Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, MNRAS, 435, 2835

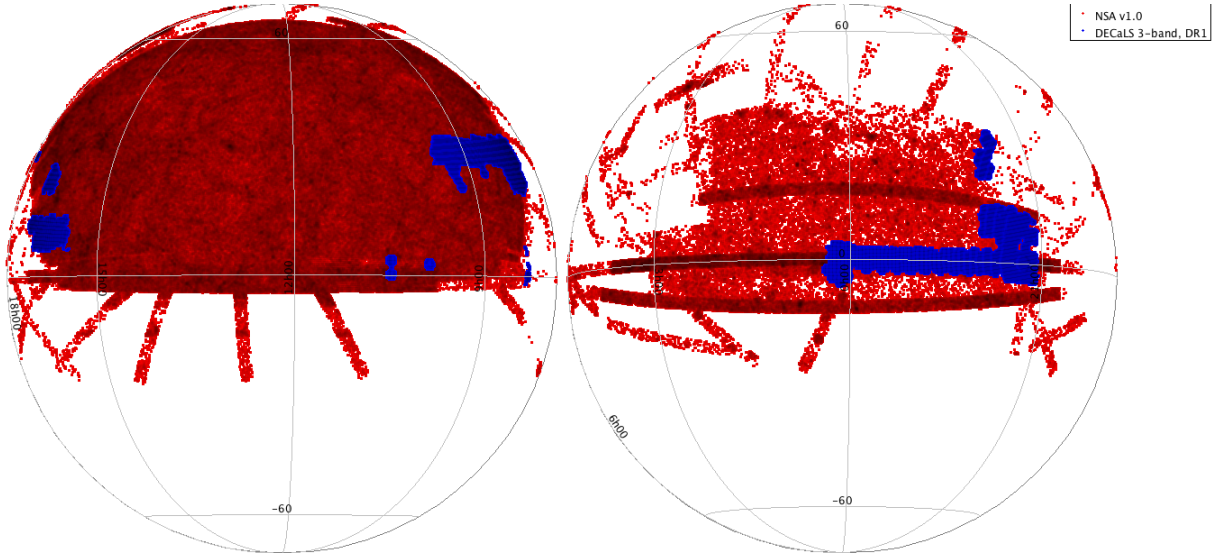## 4. ANALYSIS

[1] http://www.nsatlas.org/

Figure 1. Overlap between galaxies in the NASA-Sloan Atlas (red) and selected targets for Galaxy Zoo from DECaLS DR1 (blue).
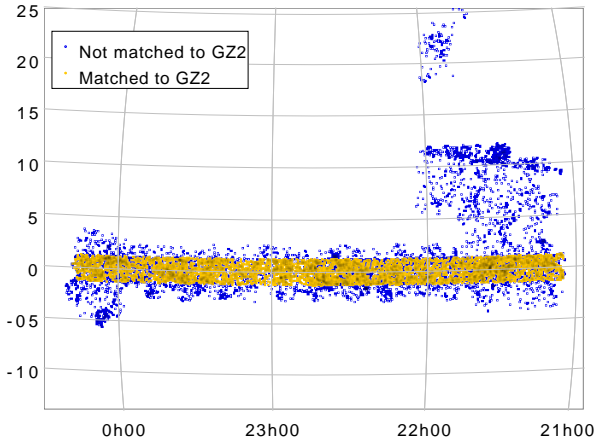

Figure 2. GZ-DECaLS galaxies in the Stripe 82 region. Galaxies with a match in the main GZ2 sample are shown the filled yellow symbols. Galaxies without a GZ2 match (open blue symbols) are due to a combination of lying outside the SDSS DR7 footprint and/or being fainter than the $m_r = 17.0$ magnitude limit for GZ2.
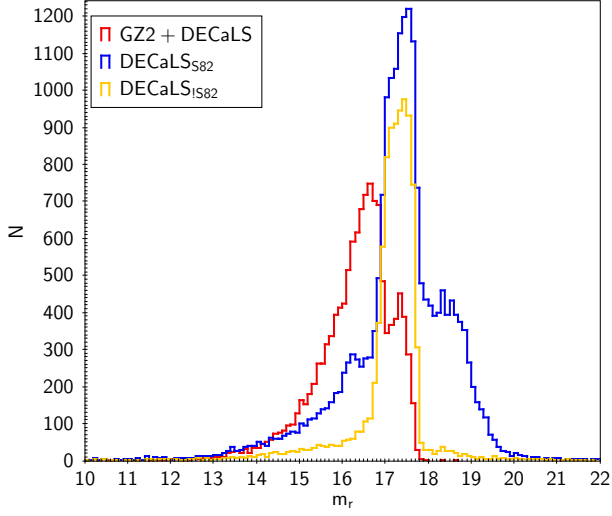
Figure 3. Histogram of the apparent *r*-band magnitude distribution for various GZ-DECaLS subsamples, binned in $\Delta m_r = 0.1$. The red histogram shows galaxies classified in both GZ2 and GZ-DECaLS; the blue represents GZ-DECaLS galaxies from the Stripe 82 region but not detected in GZ2, and the yellow represents GZ-DECaLS galaxies from outside the Stripe 82 region but not detected in GZ2
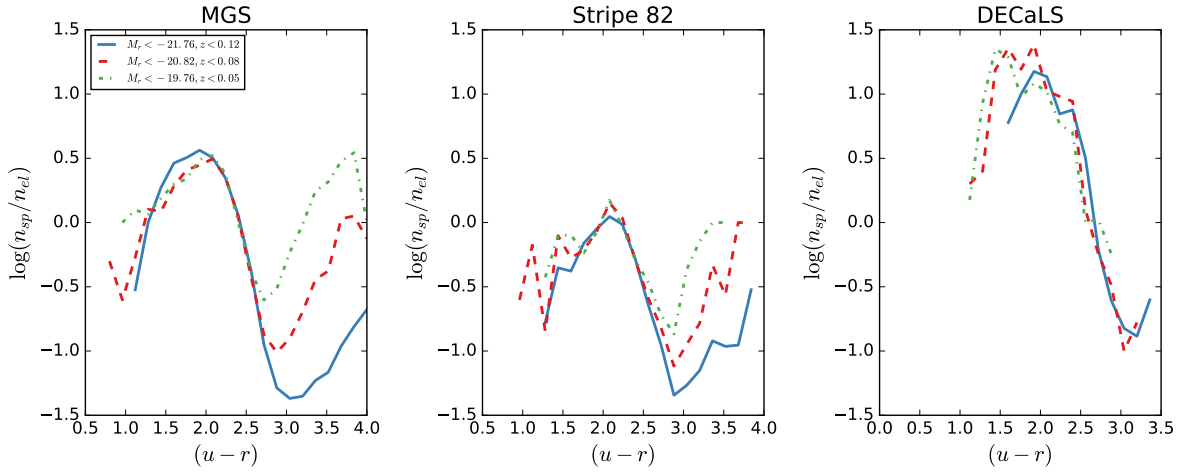

Figure 4. Average spiral-elliptical ratio for various galaxy samples as a function of optical $(u-r)$ color. From left to right, curves are for the GZ2 main spectroscopic sample, the deeper Stripe 82 coadded images, and the DECaLS images. Colors/linestyles show different volume/absolute magnitude limits for each sample.

Table 1
Galaxy Zoo morphological demographics for low-z optical imaging — all galaxies

Task	SDSS main sample			Stripe 82 coadd			DECaLS		
	N_{tot}	f_{tot}	$f_{prevtask}$	N_{tot}	f_{tot}	$f_{prevtask}$	N_{tot}	f_{tot}	$f_{prevtask}$
smooth	179153	0.74	0.74	16209	0.82	0.82	23292	0.72	0.72
features/disk	64067	0.26	0.26	3346	0.17	0.17	7967	0.25	0.25
star/artifact	280	< 0.01	< 0.01	210	0.01	0.01	1170	0.04	0.04
edge-on	9932	0.04	0.16	624	0.03	0.19	1726	0.05	0.22
not edge-on	54135	0.22	0.84	2722	0.14	0.81	6241	0.19	0.78
barred disk	14366	0.06	0.26	801	0.04	0.29	1174	0.03	0.19
no bar	39887	0.16	0.74	1932	0.10	0.71	5167	0.15	0.81
spiral	45462	0.19	0.84	2520	0.13	0.92	4973	0.15	0.80
no spiral	8791	0.04	0.16	213	0.01	0.08	1368	0.04	0.20
tight spiral arms	17322	0.07	0.39	1113	0.06	0.45	2279	0.07	0.46
medium spiral arms	20691	0.08	0.46	981	0.05	0.40	1637	0.05	0.33
loose spiral arms	6821	0.03	0.15	388	0.02	0.16	871	0.02	0.18
1 spiral arm	1879	0.01	0.04	119	0.01	0.05	237	< 0.01	0.05
2 spiral arms	26413	0.11	0.59	1602	0.08	0.65	3566	0.10	0.72
3 spiral arms	3025	0.01	0.07	188	0.01	0.08	625	0.01	0.13
4 spiral arms	837	< 0.01	0.02	51	< 0.01	0.02	192	< 0.01	0.04
5+ spiral arms	758	< 0.01	0.02	44	< 0.01	0.02	167	< 0.01	0.03
?? spiral arms	11922	0.05	0.27	478	0.02	0.19	—	—	—
no bulge	3962	0.02	0.07	103	0.01	0.04	593	0.01	0.10
noticeable bulge	34139	0.14	0.63	1139	0.06	0.42	—	—	—
obvious bulge	15791	0.06	0.29	1321	0.07	0.48	5316	0.16	0.85
dominant bulge	361	< 0.01	0.01	170	0.01	0.06	432	0.01	0.07
round edge-on bulge	6506	0.03	0.66	524	0.03	0.85	1244	0.03	0.76
boxy edge-on bulge	173	< 0.01	0.02	5	< 0.01	0.01	53	< 0.01	0.03
no edge-on bulge	3135	0.01	0.32	84	< 0.01	0.14	329	0.01	0.20
round elliptical	62308	0.26	0.35	6092	0.31	0.38	9279	0.28	0.39
in-between elliptical	91284	0.37	0.51	8331	0.42	0.51	11369	0.35	0.48
cigar-shaped elliptical	25561	0.10	0.14	1786	0.09	0.11	2644	0.08	0.11
odd feature	23795	0.10	0.10	1713	0.09	0.09	—	—	—
no odd features	219425	0.90	0.90	17842	0.90	0.91	—	—	—
ring	4099	0.02	0.18	178	0.01	0.11	317	0.01	0.01
lens/arc	155	< 0.01	0.01	17	< 0.01	0.01	4	< 0.01	< 0.01
disturbed	720	< 0.01	0.03	47	< 0.01	0.03	—	—	—
irregular	5761	0.02	0.25	113	0.01	0.07	44	< 0.01	< 0.01
other	4919	0.02	0.21	589	0.03	0.38	14	< 0.01	< 0.01
merger	7018	0.03	0.31	599	0.03	0.39	—	—	—
dust lane	220	< 0.01	0.01	6	< 0.01	< 0.01	141	< 0.01	< 0.01
overlapping	—	—	—	—	—	—	52	< 0.01	< 0.01
nothing	—	—	—	—	—	—	979	0.03	0.03
merger	—	—	—	—	—	—	1543	0.04	0.04
tidal debris	—	—	—	—	—	—	343	0.01	0.01
merger and tidal debris	—	—	—	—	—	—	141	< 0.01	< 0.01
neither	—	—	—	—	—	—	29232	0.93	0.93

Table 2
Galaxy Zoo morphological demographics for low-z optical imaging — overlaps only

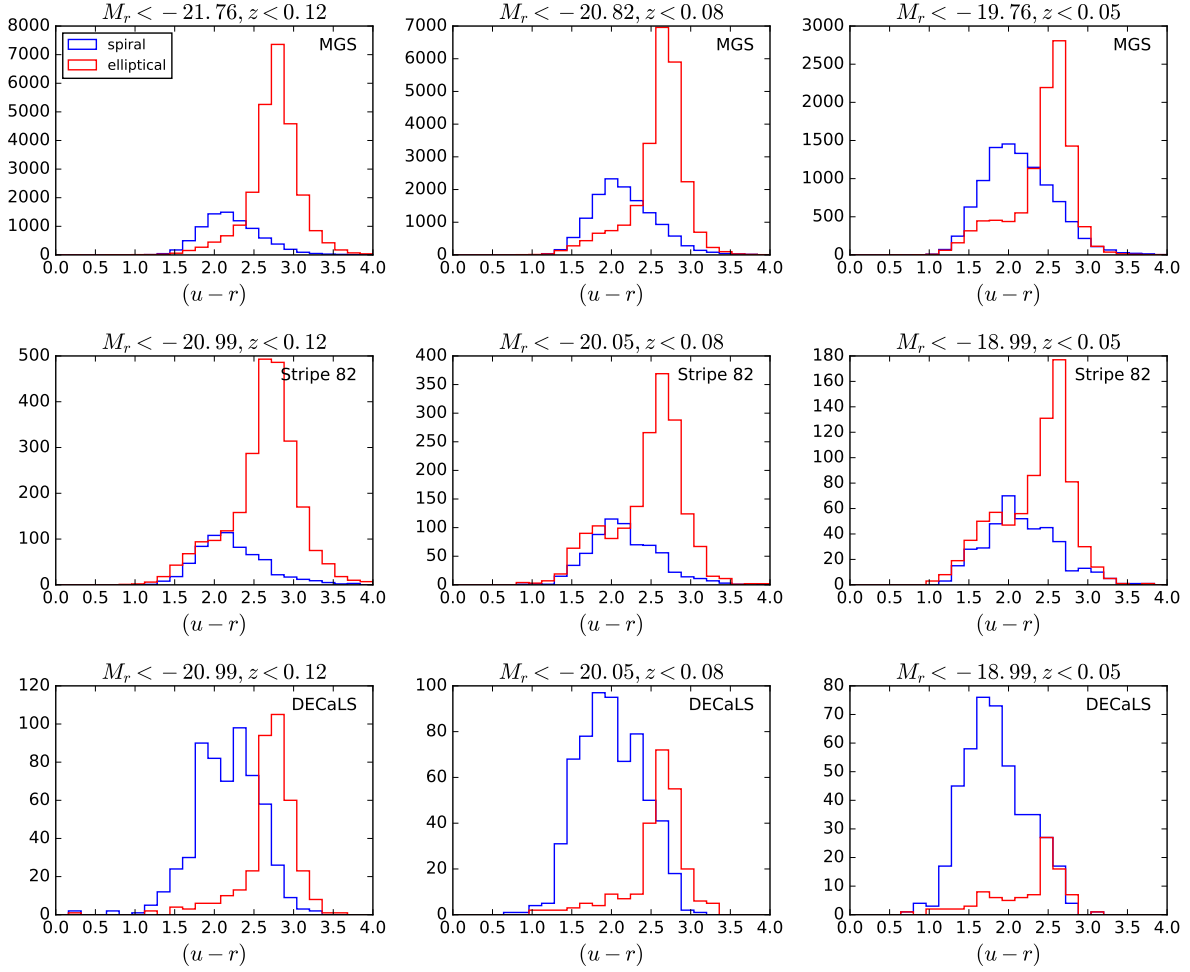| Task	SDSS main sample			Stripe 82 coadd			DECaLS		
	N_{tot}	f_{tot}	$f_{prevtask}$	N_{tot}	f_{tot}	$f_{prevtask}$	N_{tot}	f_{tot}	$f_{prevtask}$
smooth	179153	0.74	0.74	16209	0.82	0.82	23292	0.72	0.72

Figure 5. Histograms of the optical $(u−r)$ color distribution for various volume-limited and sample choices, separated by highly-confident ($p >= 0.8$) morphological classifications into spiral and ellipticals. Galaxies with intermediate morphologies ($0.2 < p < 0.8$) are not shown. **Top row**: GZ2 main spectroscopic sample. **Middle row**: Stripe 82 coadded. **Bottom row**: DECaLS.