Morphology is a Link to the Past: examining formative and secular galactic evolution through morphology

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Melanie A. Galloway

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advisor:
Lucy Fortson

month, 2017

Acknowledgements

Some acknowledgements

Dedication

Firstly dedicated to Zuko, for being the sweetest little bird.

Abstract

Galaxy morphology is one of the primary keys to understanding a galaxy's evolutionary history. External mechanisms (environment/clustering, mergers) have a strong impact on formative evolution of the major galactic components (disk, bulge, Hubble type), while internal instabilities created by bars, spiral arms, or other substructures drive secular evolution via the rearrangement of material within the disk. This thesis will explore several ways in which morphology may impact the dynamics and evolution of a galaxy using visual classifications from several Galaxy Zoo projects. Section 1 will focus on the present morphology of galaxies in the local Universe ($z < 0.2$) using data from Galaxy Zoo 2 and Galaxy Zoo UKIDSS. Section 2 will examine populations of morphologies at various lookback times, from $z = 0$ out to $z = 1$ using data from Galaxy Zoo Hubble.

We first explore the impact of bars in disc galaxies on channeling gas from the outer regions of the disk to the inner few kpc necessary to fuel an active galactice nucleus (AGN). Using a sample of 19,756 disk galaxies at $0.01 < z < 0.05$ imaged by the Sloan Digital Sky Survey and morphologically classified by Galaxy Zoo 2, the difference in AGN fraction in barred and unbarred disks was measured. A weak, but statistically significant, effect was found in that the population of AGN hosts exhibited a 16.0% increase in bar fraction as compared to their unbarred counterparts at fixed mass and color. These results are consistent with a cosmological model in which bar-driving fueling contributes to the fueling of growing black holes, but other dynamical mechanisms must also play a significant role.

We study the wavelength dependence on morphology by comparing the optical morphological classifications from GZ2 to classifications done on infrared images in GZ:UKIDSS. We find some cool result. [to be continued]

We examine more directly the morphological changes in galaxy populations as a function of their age using classifications from Galaxy Zoo: Hubble. A sample of XX,XXX disc galaxies from the COSMOS field at $0 < z < 1$ were identified as active or passive using a NUV-r / r -J diagnostic with rest-frame colors from the UltraVISTA catalog. We find that the fraction of disks that are passive increases/decreases from X.X% at

$z = 1$ to X.X% at $z = 0$. We intepret this result as [something having to do with the transformation of disk to elliptical, depending on result]. Addionally, we emphasize the challenges of visual classification that are particular to galaxies at high redshift. We present a correction technique to address these biases using simulated images of nearby SDSS galaxies which were artificially redshifted using the FERENGI code and classified in GZH.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Intro reminder to include: discussion on wavelength-dependence of morphological classifications. Discuss the transition of star formation coming through in optical, disappearing in near-IR, then reappearing in mid-IR (Galaxy Morphology Buta 2013 and briefly at end of Buta 2010).

1.1 Methods for morphological classification

Historically, most methods of morphological classification been done by visual inspection of small samples of images (e.g. Hubble (1926); Sandage (1961); de Vaucouleurs (1963); Block et al. (1994); Eskridge et al. (2002); Buta et al. (2010)), by either a single person or handful of experts. This method is becoming obsolete as we enter a new era of large data, with recent surveys such as SDSS and HST-Legacy, and upcoming JWST and LSST, producing high-quality images of hudreds of thousands of galaxies. To date, the largest morphological catalogs created by visual inspection from a small group of experts includes the Nair and Abraham catalog (Nair & Abraham, 2010) with $\sim 14{,}000$ galaxies, RC3 Catalog (de Vaucouleurs, 1991) with $\sim 23{,}000$ galaxies, and MOSES (Schawinski et al., 2007) with 50,000 galaxies. Even these catalogs, while successful, do not compare in size to the newly incoming data, and so more powerful and robust efforts are required to obtain morphological information on these scales.

One alternative to direct visual classification of morphologies is the use of proxies such as color, mass, surface brightness profile, or some combination of several. Color is commonly used as a proxy because of its mostly-tight relationship global morphology, in that spirals tend to be red and ellipticals tend to be blue. This type of morphological classification will always suffer from a high degree of inaccuracy, as there is no perfect physical measurement that is 100% correlated with shape. The morphology of a galaxy traces the dynamical history, where proxies such as color trace stellar growth; these two properties thus reveal different evolutionary histories on possibly very different timescales Fortson et al. (2011). Last, while there are several proxies which correlate somewhat with the probability of a galaxy being spiral or elliptical, very few could be used to identify finer substructures or more detailed morphological features within the overall shape.

An ideal method for handling the large amounts of data would be an automated classification scheme. Several such algorithms have been developed, with some success (Odewahn et al., 2002; Peng et al., 2002; Conselice, 2003) by using the stellar light distribution of the galaxy to assign it a morphological class. These approaches tend to be limited to identifying the global morphologies (ie, spiral or elliptical), and lack the precision to accurately identify finer, detailed features (such as bars or the number of spiral arms) (Beck et al. 2017). Further, they tend to incorporate proxies such as color as their input, which are often not accurate as previously noted. Much more promising techniques are currently being tested which incorporate the use of machine-learning algorithms and neural networks (Dieleman et al., 2015; Huertas-Company et al., 2015),(Beck et al.2017), but these require massive and accurate training-sets to perform properly.

A best-of-both-worlds approach uses the power of crowdsourcing, which uses the input of thousands of individuals to visually classify galaxies in a fraction of the time achievable by a handful of experts; such a method was developed by Galaxy Zoo, the data from which is used throughout this thesis. The Galaxy Zoo project uses a simple online interface whereby images of galaxies are visually inspected by volunteers, which allows the identification of intricate morphological features to a higher degree of accuracy than computer algorithms today can achieve. Addionally, with a required 40+ independant classifications per galaxy, the resulting classifications carry a greater statistical significance than those generated from one or a few experts. The next chapter will describe how Galaxy Zoo collects data from volunteer citizen scientists, how the data is reduced and debiased, and finally how the data is used to assign morphological classifications to large samples of galaxies.

Chapter 2

Methodology

2.1 A Brief History of Galaxy Zoo

The increasing accessibility of the Internet in the last decade has allowed scientists to "outsource" tasks online using citizen science, with huge success. The project Seti@Home [1] (Anderson, 2002), launched in May 1999, was one of the first projects that revealed the massive number of people willing and excited to help contribute to science. Since launch, over 5 million participants donated idle time on their computers to assist SETI in analyzing radio telescope data to help in the search for extraterrestrial life. Citizen scientists were also extemely interested in taking in an even more active role in research, as seen in a later project Stardust@Home [2] , in which volunteers searched for dust grains in data via a web interface. This project engaged over 20,000 volunteers, and those who discovered dust grains were invited to become co-authors on the announcement papers. Early citizen science projects such as these inspired the launch of Galaxy Zoo.

The real need for a faster method of obtaining galaxy morphologies for large samples was realized in 2007 by graduate student Kevin Schawinksi, who was studying populations of elliptical galaxies as work for his PhD thesis at Oxford University. At the time, the accepted and fastest method for identifying early-type galaxies (in large quantities) was to select based on SDSS-measured spectra (**?**). He knew, however, that this sort of

[1] http://setiathome.berkeley.edu/
[2] http://stardustathome.ssl.berkeley.edu/

method would exclude potential star-forming ellipticals (as well as potentially include passive spirals), due to the non-perfect correlation between morphology and color, as mentioned in the previous Chapter. So, realizing that a visual inspection of the direct appearance of the galaxies was necessary to create a complete sample of ellipticals independent of color, Schawinksi devoted an entire week to classifying 50,000 galaxies by eye (MOSES, **?**).

The grueling task of classifying only a small fraction of the entire SDSS main sample ($\sim$ 900,000 galaxies) made it apparent that a better method for visual classification was becoming neccessary. Inspired by the 20,000 volunteers who participated in the Stardust@Home project, Schawinksi and Oxford colleague Chris Lintott realized that it would only take a few years to classify all of SDSS Main, assuming a similar participation response as StarDust. This led to the launch of Galaxy Zoo in July, 2007. This first phase (known now as Galaxy Zoo 1, or GZ1), included the brightest (Petrosian magnitude $r < 17.77ABmag$) 893,212 images from SDSS Data Release 6 (**??**). In this project users were asked to indentify simple features of a given galaxy, including whether it was elliptical or spiral, clockwise or anticlockwise, a merger, or star/other (the original Interface with options is shown in Figure 2.1.

-Classifying galaxies by morphology *independently* of proxies like color or mass is important because the morphology traces the dynamics while color traces stellar content, and these are only correlated on average and in the local universe. cite some motivations for crowdsourcing (ie Raddick; discussion of authentic scientific participation and learning)

2.2 Galaxy Zoo Data Reduction

2.2.1 User weighting by consistency

A typical Galaxy Zoo project collects classifications from over 10,000 unique volunteers. With such large numbers of classifiers, there exists the possibility that some fraction of these are “unreliable”, that is, their votes are consistent with random clicking. To ensure that all votes collected represent real classifications, a weighting technique is implemented to detect and down-weight unreliable votes.

The weighting scheme used for all GZ projects represented in this thesis (GZ2,

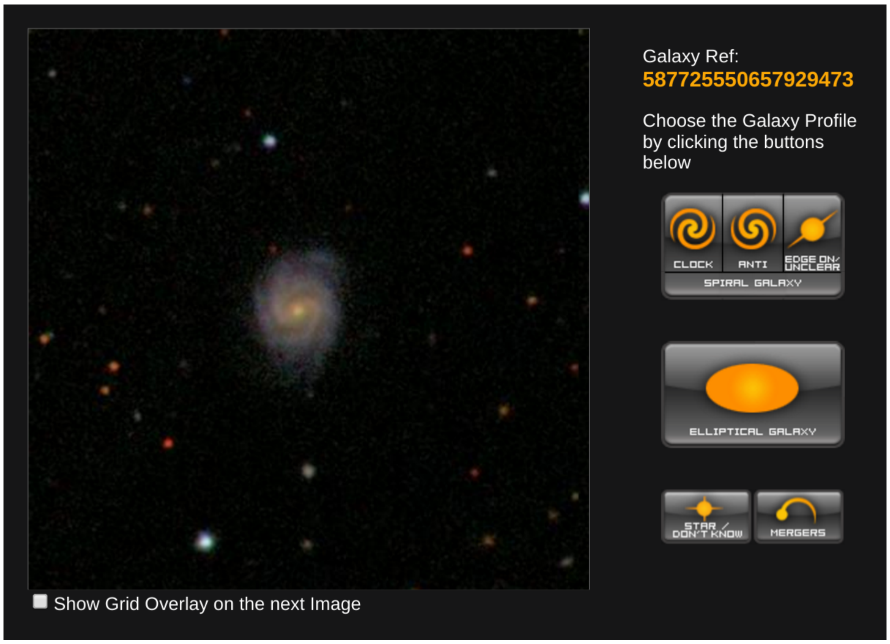

Figure 2.1 GZ1Interface

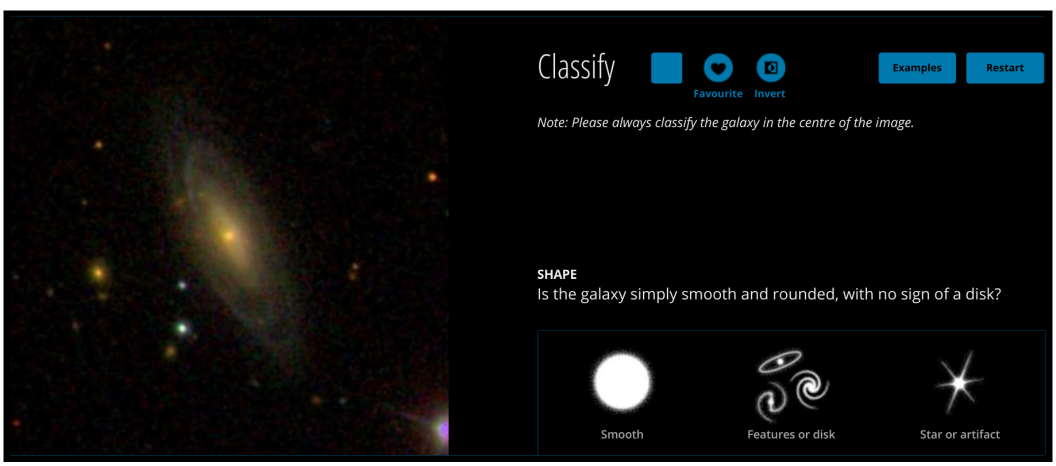

Figure 2.2 Interface

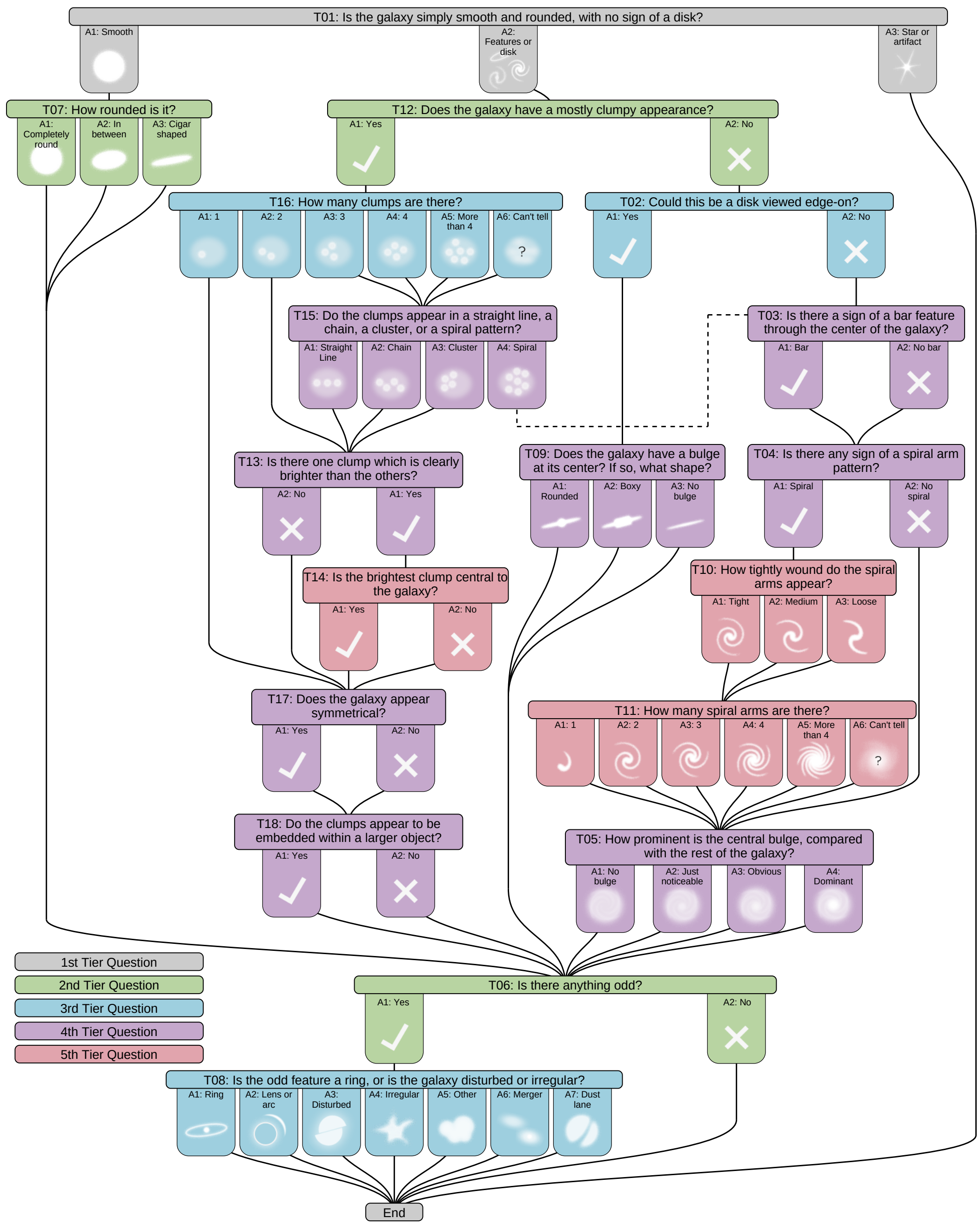

Figure 2.3 Decision tree for Galaxy Zoo:Hubble. Explain colors. Identical to GZ2 and UKIDSS with the addition of the clumpy question.

GZ:UKIDSS, and GZ:Hubble) evaluates the consistency of each user by how often their votes agree with the majority for each task in the decision tree. The consistency rating κ for a single task is defined as:

$$\kappa = \frac{1}{N_r}\sum_{i=1}^{N_r}\kappa_i \tag{2.1}$$

where f_r is the vote fraction for each response in the task, N_r represents the total number of responses to the task, $\kappa_i = f_r$ if the user's vote corresponds to response i, and $\kappa_i = (1 - f_r)$ if it does not. In this system, κ is then high if the vote agrees with the majority, and low if it does not.

The mean consistency computed for each response given is defined as the user's overall consistency $\bar{\kappa}$, and the user is assigned a weight w defined as:

$$w = \min(1.0, (\bar{\kappa}/0.6)^{8.5}) \tag{2.2}$$

All votes are then recalculated using the user weights, and the process is repeated as many as three times to ensure convergence. It can be seen in Equation 2.2 that a user's weight value is always less than or equal to one; in other words, users are only downweighted in cases of noticeable inconsistency, and never upweighted. Willett et al. (2013) show that most users with low consistencies tend to only have contributed a handful of classifications, which could either indicate that users become more accurate as they classify more galaxies, or that inconsistent users are inherently less likely to be interested in the project.

2.2.2 Classification bias in the local Universe

For samples of galaxies limited to the local universe ($z \lesssim 0.2$), there is no expected redshift dependence on the morphological classifications. Therefore, we would expect vote fractions representing different morphological features to be constant with respect to redshift. However, this is not the case - the average vote fraction for features, bars, spirals, and several others actually tend to *decrease* with redshift. Since we assume such features should be equally prevalent at any redshift in this small range, some bias unrelated to any true morphological evolution must be affecting the vote fractions.

The source of this bias comes from the apparent size and brightness of the images of the galaxies being classified, which are strongly affected by redshift. Images of more distant galaxies appear smaller and dimmer, and therefore finer features are simply more difficult to detect. This sort of classification bias is a problem with any morphological classification, whether it be expert classifiers, automated dection, or crowd-sourced visual inpsection.

This section will describe the methods used to correct this type of classification bias for galaxies in the local Universe, where no true morphological evolution is a factor. Beyond the local Universe this assumption is no longer valed, so techniques implimenting classifications of artificially-redshifted galaxies are used for calibration; these are described in detail in Chapter **??**.

Debiasing Galaxy Zoo 2: W13 method

The debiasing technique used in GZ2 assumed firstly that galaxies with similar brightnesses and sizes should, on average, share similar mixes of morphologies at any redshift. Using this assumption, galaxies were grouped into bins of absolute magnitude M_r, Petrosian effective radius R_{50}, and redshift. For each task in the GZ2 decision tree, the vote fractions for each response in any size/magnitude bin were adjusted so that their average matched the average vote fraction of its lowest-redshift bin. This method is described in detail in Willett et al. (2013), but the main approach is as follows:

For a given size/magnitude bin, the ratio of vote fractions for a pair of responses i and j for a single task can be written as f_i/f_j. Due to the classification bias described above, this ratio may not reflect the "true" ratio for this size/magnitude range, but can be written in terms of the true ratio with a multiplicitave constant $K_{i,j}$:

$$\left(\frac{f_i}{f_j}\right)_{z=z'} = \left(\frac{f_i}{f_j}\right)_{z=0} \times K_{i,j} \tag{2.3}$$

Where $(f_i/f_j)_{z=z'}$ represents the ratio measured in a size/magnitude bin at $z = z'$, and $(f_i/f_j)_{z=0}$ is the "true," or intrinsic ratio of vote fractions, defined as the ratio measured in the lowest redshift bin.

Figure 2.4 shows the local ($z = 0$) ratios of f_i/f_j for the first two responses i and j for the first three tasks of the GZ2 decision tree, which are used to calculate the debiased

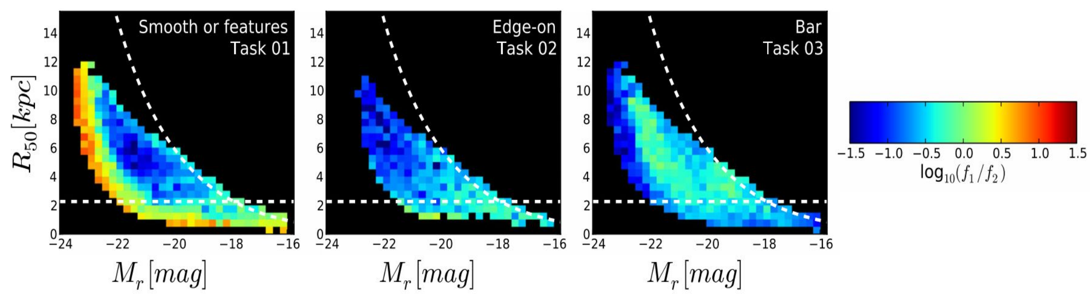

Figure 2.4 Local ratios of morphologies for the first three tasks in the GZ2 decision tree, used to derive debiased votes for the GZ2 sample. The full figure which includes baseline ratios for all tasks in the GZ2 decision tree is shown in Willett et al. (2013), Figure 5.

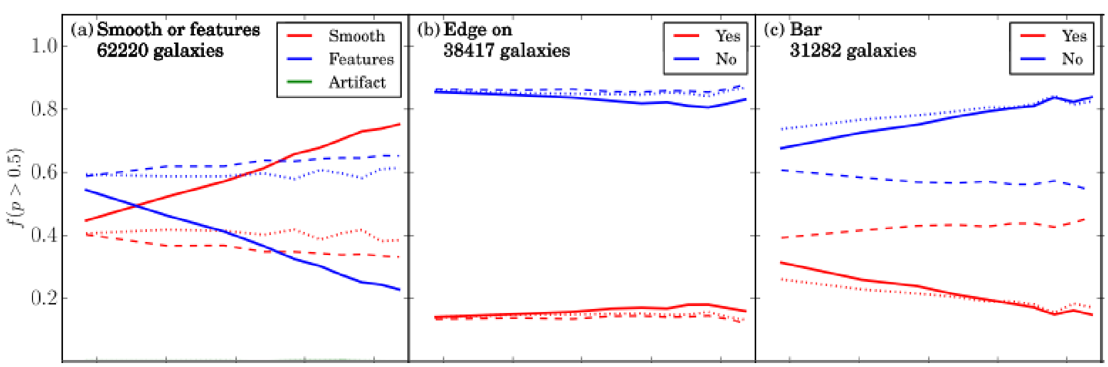

Figure 2.5 Placeholder - will make original later. Shows results of 2 debiasing methods for the first 3 tasks in GZ2

.

vote fractions as outlined above. For Task 01, f_i/f_j corresponds to $f_{smooth}/f_{features}$, for Task 02 $f_{edgeon}/f_{not\ edgeon}$, and for Task 03 $f_{bar}/f_{no\ bar}$. The figure demonstrates the size and magnitude dependence of the most local morphological populations: for example, in Task 01, the largest and brightest galaxies tend to have more votes for "smooth" than "featured", which is consistent with our current understanding that ellipticals tend to be larger and more massive than spirals.

The results of this method for the first three Tasks in the GZ2 decision tree are shown in Figure 2.5. For each response in each Task, the average vote fraction is calculated as a function of redshift. Solid lines represent the weighted/non-debiased votes and the dotted lines are the debiased votes using this method (hereafter W13). The redshift dependence on vote fraction is very evident in the downward trend of the solid lines corresponding to responses which detect features, such as $f_{features}$ and f_{bar} in this example. The dashed lines show the effect of the debiasing which attempts to flatten out the distribution. Full figures showing the results for all Tasks in the tree are available in Willett et al. (2013) (Figure 3) and Hart et al. (2016) (Figure 8). From 2013-2017, the debiased vote fractions calculated in this method were used in the majority of published Galaxy Zoo papers, and are used in the study described by Chapter 3.

Debiasing GZ2 and UKIDSS: H16 method

The W13 debiasing method is successful at adjusting the vote fractions to more accurately resemble the "true" distribution of morphologies at low redshift, but has two primary limitations. First, the rectangular binning of all three parameters (size, magnitude, and redshift) is only effective when the parent sample is large enough that sufficient data per bin remains available after the three dimentional binning. (For example, to requre 10 bins in each parameter with at least 50 galaxies per bin, a parent sample must contain at minimum N=10x10x10x20=50,000 galaxies, assuming a perfectly even distribution of values in each parameter). GZ2 is not so affected by this limitation, with a parent sample size of $\sim$ 250,000 galaxies. However, this is only true when considering the debiasing of the first Task, which is asked of every galaxy. After this Task, the parent sample for computing a correction term decreases as not all Tasks are asked of every galaxy; for example, the Tier 4 Task which asks for the number of spiral arms is only seen by the majority of volunteers in 33,000 galaxies of the full GZ2 sample. Thus debiasing this Task would require a smaller limit on the number of bins per dimension or the number of galaxies per bin, both of which decrease the robustness of the method. Even with a large parent sample for any Task, the rectangular binning is also limited by the inability to account for data which lie on the outer edges of the parameter space, as there tends to be insufficient data in the outer bins.

A new debiasing technique (hereafter H16) was developed by Galaxy Zoo member Ross Hart (Hart et al., 2016) which substitutes Voronoi binning for the rectangular method. Voronoi binning optimizes the shape and location of bins based on the desired signal for each bin; in this case, the number of galaxies per bin is set initially, and the bins are drawn to fulfill that requirement. In this way, the number of galaxies available for measuring the change in vote fractions for each bin is maximized. Thus, this method is more effective at debiasing smaller samples, where the three dimensional binning preserves the signal in each bin. This method was therefore used to debias the UKIDSS sample which is much smaller than GZ2, with only ∼70,000 galaxies. An example of Voronoi binning the UKIDSS data in size and magnitude is shown in the left panel of Figure 2.6. Each size and magnitude bin is then Voronoi-binned by redshift.

The second limitation of the W13 method is that while it effectively corrects the vote fractions for any Task so that the average morphology is constant as a function of redshift, it does not account for the *distribution* of morphologies at low redshift. This produces good results when the corrected values are used for population studies, where the percentage of galaxies exhibiting a particular morphology are desired, but may not always reproduce accurate *individual* vote fractions. The R16 method instead corrects the high redshift vote fractions based on the change in distribution of vote fractions observed at low redshift, rather than comparing to only the average values. The first step of this method is shown in the right panel of Figure 2.6. For the low redshift bin of a given task, the cumulative distribution of vote fractions for each response is fit with a continuous function, which is used as the baseline distribution (similar to teh baseline average votes in the W13 method.) The vote fractions making up the cumulative distributions at higher redshifts are then adjusted as needed to match the low redshift distribution as closely as possible.

Results of this method are shown and compared to W13 in Figure 2.5. The top row shows the average vote fractions as a function of redshift for the raw (solid lines), W13 debiased (dashed lines), and R15 debiased data (dotted lines). Both methods are succesful in stabalizing the average morphologies across redshift. The bottom row shows the distribution of vote fractions of a low-redshift bin and high redshift bin. It can be seen here that while both methods can reproduce the average vote fractions at low redshift, the R15 method is more succesful in reproducing the distribution of votes at low redshift.

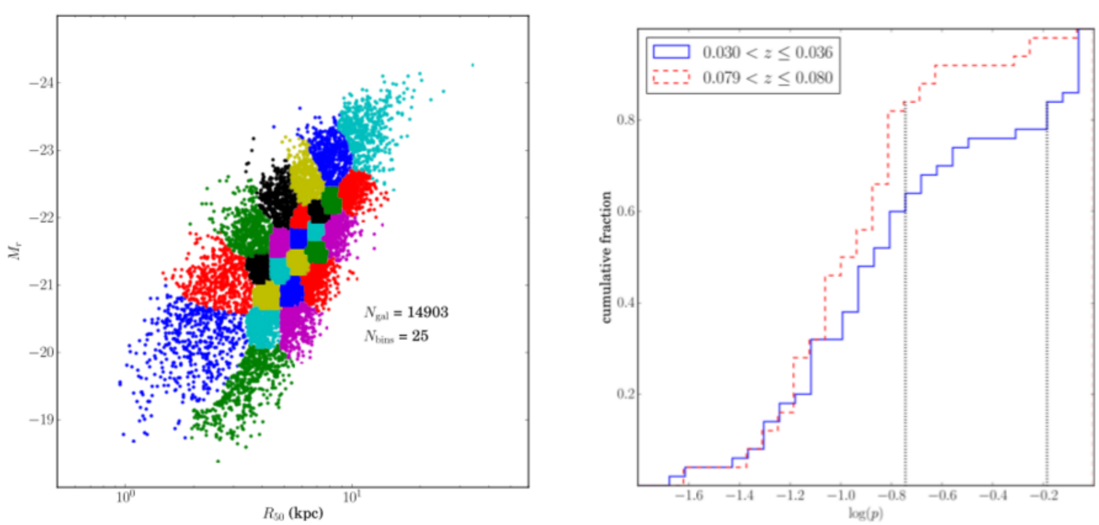

Figure 2.6 Placeholer - make my own later. Left shows how voronoi binning is done; compare to rectangular binning in W13. Right is cumulative distribution of some vote fraction at low and high redshift.

Chapter 3

Bar AGN project

Basically Galloway et al. (2015)

Chapter 4

UKIDSS

4.1 Intro: wavelength dependence on morphology: optical and IR

Historically, visual morphological classification of galaxies has been conducted on optical images. Blue B-band images were the primary source dating back to Hubble's classic tuning-fork classification scheme (Hubble, 1926) and in the subsequent modifications by Sandage (1961) and de Vaucouleurs (1963). The more recent and larger morphological catalogs also derive their classifications from rest-frame optical images, either single-band (de Vaucouleurs (1991) (B-band), Scarlata et al. (2007) (ACS I-F814W), Fukugita et al. (2007) and Nair & Abraham (2010) (SDSS g-band)) or color-composite (Lintott et al. (2008), Willett et al. (2013) (SDSS-gri)).

In the optical regime, the flux is dominated by young, hot stars; this results in an emphasis of spiral structure in the images, but they tend to have patchy appearances due to the abundance of star-formation regions in the arms. Optical images also are impacted by extinction due to dust, which can obscure features that tend to be composed of older stellar components (such as bars and bulges). Longer wavelenths are free of these effects, making them ideal for revealing the underlying "stellar backbone" of galaxies.

It is possible, then, to consider two morphologically distinct components of a galaxy: a gas-dominated Population I disk, and a star-dominated Population II disk. The Population I disk is most easily seen in the optical, revealing HII regions, cold HI gas, and emission from young OB stars; these regions will tend to highlight flocculance in spiral structure. The Population II disk, on the other hand, traces the underlying mass distribution; consisting of the old, cooler stellar population, it is more easily seen at longer wavelengths. Block & Puerari (1999) even suggests that two separate classifications schemes should be required for all galaxies; one for the Population I disk, which can be probed in optical and ultraviolet images, and a Population II disk, for which longer wavelength images, free of dust extinction, would be required.

The extent to which the morphologies of the younger and older stellar populations are decoupled, however, is not yet clear. Early studies which directly compared optical and near-IR images found very significant differences between the two morphologies (Hackwell & Schweizer, 1983; Thronson et al., 1989; D. Block, 1991; Block et al., 1994). Block & Puerari (1999) goes as far as to sugget that there is no correlation between the two, and that the optically-definied Hubble tuning fork "does not constrain the morphology of the old steallar Population II disks." However, all of the aforementioned studies only compared morphologies of either a single galaxy, or at most a handful, so these conclusions cannot be applied generally.

The advent of larger surveys incorporating near and mid-IR detectors enabled morphological comparisons between the two wavelength regimes on a much grander scale than had previously been achieved. New results contradicted those of the previous case-studies: in general, IR morphology was found to be well-correlated with optical morphology in larger samples of galaxies. Eskridge et al. (2002) compared near-IR H-band (1.65μm) Hubble-type classifications to B-band in a sample of 205 nearby spiral galaxies from the Ohio State University Bright Spiral Galaxy Survey (OSUBSGS). Applying deVaucouler's classification system, they found an overall good correlation between the two morphologies, but on average galaxies from Sa through Scd appeared one T type earlier in the H band than in the B band. In the IR images the bulge tended to appear more prominent and the spiral arms less knotty, which resulted in the slightly earlier classifications. For the earliest (optically S0/a and Sa) and latest-type galaxies (optically Scd through Sm), no difference in morphologies was found. This is an expected result for the earlier-types, since these have little ongoing star formation and very little dust, so it is expected that both optical and IR morphologies are dominated by old stars. This result is less intuitive for the later-type galaxies, as these are dominated by ongoing star formation. However, these galaxies are defined as having very weak or nonexistant bulges and poorly defined spiral structure. Since the main driver in the differences in morphology across wavebands was found in the intermediate spirals to be the relative prevalance of a bulge and difference in contrast and appearance of spiral arms, galaxies lacking these features should not, in fact, be expected to look different in the IR than the optical.

Buta et al. (2010) obtained similar results comparing optical and mid-IR (3.6 μm) images from the *Spitzer* Survey of Stellar Structure in Galaxies (S^4G, Sheth et al. (2010)) in a large sample of 2,331 spiral galaxies. Like Eskridge et al. (2002), the optical and IR classifications were very well correlated, with the most significant differences occuring for S0/a to Sc galaxies, where the 3.6 μm were on average slightly earlier than the B-band classifications.

Infrared imaging is also often used in place of (or in addition to) optical to indentify stellar bars (e.g. Mulchaey & Regan (1997); Knapen et al. (2000); Block et al. (2004); Sheth et al. (2008)). Like bulges, bars are primarily composed of old, red stars, and therefore better traced by longer wavelengths. In fact, it is not uncommon for an infrared bar to be completely invisible in the optical. Notable examples include NGC 1566 (Hackwell & Schweizer, 1983), NGC 1068 (Thronson et al., 1989; Scoville et al., 1988), NGC 309 (D. Block, 1991), NGC 4736 (Block et al., 1994), and NGC 4303 (Figure 1, Sheth et al. (2003)). This trend is not only limited to case-studies; for example, in a larger sample of 29 galaxies classified as unbarred in the optical, *50%* of these were found to be barred in the near-IR images (Mulchaey & Regan, 1997).

The fraction of spiral galaxies which exhibit bars (defined as the bar fraction) has been measured extensively in optical images, and typically falls near 50% when bars of all strengths are considered (Masters et al., 2010)(should probably cite more). Since it is much more common to find an infrared bar in an optically unbarred galaxy than the reverse, it is expected that the bar fraction in the infrared will, in general, be higher than what has been measured in the optical. Some studies find a substantial increase: Seigar & James (1998) for example spectulate that "bars may always be present in disks at some level", based on finding a bar fraction of 90% when using infrared images (as compared to their optical measurement of 68%). Although their sample consisted of only 45 galaxies total, they claim this measurement should represent the general population of spirals, because their selection was not biased towards barred galaxies. Other studies report similar increases in bar fraction in the infrared, albeit not quite as large. Knapen et al. (2000) in a similar sample size of 50 galaxies find a bar fraction in the infrared of 70%, a strong increase from the optical 50%. Eskridge et al. (2000) in sample of 186 galaxies measure a bar fraction of 72% in the infrared which is *double* that of their optical measurement. While these studies report significant increases in bar fraction as a function of wavelength, they do dispute the claim by Seigar & James (1998), emphasizing that at least 30% of galaxies in their sample are truly unbarred across all wavelengths.

Other more recent studies find larger bar fractions in the infrared, not significantly so. Whyte et al. (2002) measure an increase from 72% to 79% in a sample of 72 galaxies, while Sheth et al. (2008) reports 60% for both wavelengths. MenendezDelmestre et al. (2007) also found a slight increase from 63% to 67% in a sample of 151 galaxies, noting that although bars tended to appear stronger in the near-IR, on average they were not so weak in the optical as to become undetectable. Finally, Buta et al. (2010) also reported a similar result of 60% barred spirals, which was consistent with the fraction computed in optical RC3 classifications.

Now: segue into describing how we'll investigate these using a *much* larger sample than previously done, using GZ classifications. 2 goals: 1) investigate change in hubble-ish type in UKIDSS vs GZ2 (by looking at bulge question and arms-windyness), and 2) bar fraction, plus probably some case studies of galaxies whose morphologies change drastically.

4.2 UKIDSS sample

The UKIDSS sample is comprised of 71,052 infrared images of galaxies which had been previously optically classified in GZ2. The images were taken with the United Kingdom Infrared Telescope (UKIRT) as part of the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. (2007); Warren et al. (2007). The Large Area Survey (LAS) portion of UKIDSS covered the SDSS observations at high Galactic altitudes, allowing for full YZJHK coverage.

Morphological classifications for the UKIDSS sample were obtained via Galaxy Zoo,

UKIDSS		GZ2	
Filter	Depth (AB mag)	Filter	Depth (AB mag)
Y	21.13	g	22.2
J	20.91	r	22.2
K	20.25	i	21.3
seeing:	<1.2”	PSF width:	1.4” (median in r)
pixel scale:	0.4”	pixel scale:	0.396”

Table 4.1 Comparison of depth and resolution of the UKDISS and GZ2 images. The resolution between the two surveys is comparable, but the UKIDSS images are an average of ∼1 magnitude shallower in all bands used to create the color-composite images that where classified.

where users were shown YJK color-composite images. The classification tree used was identical to that in GZ2, allowing a direct comparison of morphologies using the same vote fractions. Raw votes were counted and weighted by user consistency in the same manner as the GZ2 sample (details of this process are given in Chapter 2).

One major challenge in comparing the UKIDSS and GZ2 morphologies is to ensure that any differences measured are mostly driven by actual morphological differences between wavebands, and not due to varying instrumental parameters. Details of the instrumentation for both samples is shown in Table 4.1. The resolution of both sets are comparible - with similar pixel size and PSF widths, the ability to resolve finer features in the images should be consistent for both. The difference in depth, however, is significant: the SDSS gri bands used to create the color-composite images in GZ2 are on average ∼1 magnitude deeper than what is achieved for the LAS YJK bands in UKIDSS. To minimize the impact the difference the difference in depth may have in comparing the two sets of images, the comparison sample is limited to the nearest and brightest galaxies. The sample is thus restricted to a volume-limit of $z < 0.06$ and $M_{r,petro} < -20.0$, which consists of 10,395 galaxies of the 54,238 with spectroscopic redshifts.

4.3 Comparison of Hubble Types in Spirals

In this section the global morphologies seen in the infrared and optical are compared. As described above, the most recent studies found similar results when comparing the Hubble T-types of both wavelengths; in general, the morphologies are well-correlated, with the IR T-types being on average one T-type earlier than in the optical. The strongest difference occured for the optically intermediate-type spirals. In the most early type spirals (with very dominant bulge and very tight spiral arms), these features showed up equally well in the infrared. On the other extreme end, the very late type spirals (with almost no bulge and not well-defined arms) also showed no large change, since the relative size of the bulge and relative tightness of the arms were the main driver of the morphological differences between wavelengths. For the intermediate T-types, there was much more "wiggle room" for the bulges and arms to show more significant differences.

The first portion of this comparison will consider galaxies whose spiral arms are detected in both optical and infrared wavelengths.

As a proxy for Hubble types, the responses to the GZ Tasks related to tightness of the spiral arms and dominance of the bulge will be used, since these probe similar features to those that influence T-type classification. The Task related to arm tightness asks, "How tightly wound do the spiral arms appear?", to which a user can choose one of three responses: "tight", "medium", or "loose". For this analysis the fraction of users who answered "tight", $f_{\text{tight arms}}$, will be used to assess the relative appearance of the arms from optical to IR. The task related to bulge prominence asks, "How prominent is the central bulge, compared to the rest of the galaxy?" to which a user can respone "dominant," "obvious," "just noticeable," or "no bulge." For this analysis the sum of vote fractions for the first two responses $f_{\text{obv+dom}}$ will be used to measure the apparent size of the bulge relative to the galaxy.

Figure 4.1 shows the difference in vote fractions for arm tightness and bulge dominance between the GZ2 optical and UKIDSS infrared classifications, as a function of optical classification. The left plot shows that on average, spiral arms have a tighter appearance in optical wavelengths. For galaxies with optically very loose arms ($f_{\text{tight arms}} \sim 0$ or very tight arms ($f_{\text{tight arms}} \sim 1$), the infrared classifications tend

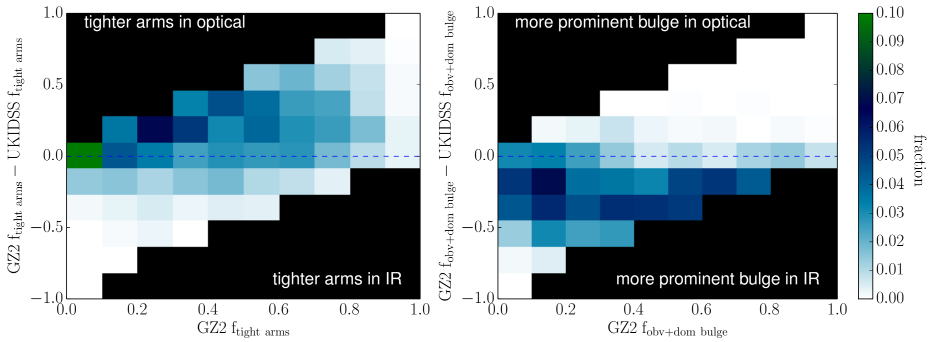

Figure 4.1 boiiiiii

to agree. For intermediately tight optical spiral arms ($0.2 < f_{tight\ arms} < 0.8$), the UKIDSS vote fraction tends to be lower than the optical by ~ 0.3 on average. This supports the work by Eskridge et al. (2002) and Buta et al. (2010) who find slightly earlier IR classifications in intermediate-type spirals. The right panel shows the change in bulge prominance as a function of optical bulge prominence. Also confirming results of previous studies, the bulge tends to be much more prominent in the infrared.

Chapter 5

Using FERENGI to correct f_{features} for redshift-induced classification bias

5.1 Intro

The GZ vote fraction f_{features} plays a crucial role in the majority of science cases that use Galaxy Zoo classifications. It represents the fraction of users who answered "feature or disk" to the first question in the decision tree, and is used to distinguish elliptical/spheroidal galaxies from those with features. Many studies aim to measure the popluation of galaxies exhibiting certain features such as bars (examples), spiral arms (examples), [dump stuff in this list]. In each of these, f_{features} is necessary for creating the sample of galaxies which could potentially contain the feature in question. This is typically acheived by setting a cut, such that all galaxies with f_{features} greater than that threshold are considered to be candidates for that study.

While f_{features} is not a true probability, the measurement is intended to be consistent among all galaxies; that is, two galaxies with similar f_{features} values should have similar likelihoods of being featured (or not featured). This has been shown to be true at low redshift by comparing the f_{features}values to expert classifications (reference Willet et al. 2013); there is a strong correlation between this vote fraction and whether the galaxy was expertly classified as a disk or an elliptical [expand on what W13 actually did.].

For distant galaxies, however, we observe that f_{features} is not consistant with nearby galaxies. As galaxies are observed at higher redshift, the images are inherently less resolved, and smaller features are more difficult to identify. This causes in a decrease in f_{features} than what would be expected if the galaxy had been observed at $z = 0$. Figure [make a figure] shows this effect: [describe some figure, probably a bunch of galaxies that are obviously spirals but with very diff. ffeatures values at diff. redshifts.] Therefore, two identical galaxies, imaged at different redshift, may have small to drastic differences in their f_{features} measurement. In order to keep f_{features} a value correlated with the likelihood of having features that is consistent for *all* galaxies, this bias must be corrected.

A method for correcting redshift bias in the GZ vote fractions was developed and implemented in prior Galaxy Zoo projects (cite GZ1 and GZ2), which contained nearby ($z < 0.2$) galaxies imaged by the SDSS. A correction factor to the classification fractions measured at the higher redshifts was applied by matching the mean vote fractions of those at the lowest redshift. This technique was valid under the assumption that, within this redhisft range, there would be no cosmological evolution of galaxies, and therefore any change in the mean vote fraction for any morphology with redshift was purely due to this observational bias, and not due to a genuine difference in morphological populations.

In GZH, the redshift range is large enough that cosmological evolution of the morphologies of galaxies is expected, and therefore the previous method of correcting redshift-bias will not work. Instead, we have developed a new method of measuring the change in f_{features} as a function of redshift using a set of simulated FERENGI images of galaxies, described in the next section. These images have been classified by volunteers in Galaxy Zoo in the same way as the GZH sample. This chapter will describe how we measure a correction factor for f_{features} using these data as a function of redshift at fixed surface brightness, and apply the correction to the GZH sample.

5.1.1 The FERENGI code

The Full and Efficient Redshifting of Ensembles of Nearby Galaxy Images code (FERENGI, Barden et al. (2008)) is an IDL procedure that generates simulated images of nearby galaxies viewed at higher redshifts, taking into account cosmological effects such as surface brightness dimming and bandpass shifting. Artificially redshifted samples of galaxies, for which the intrinsic morphologies are already known from low-redshift observations, are useful for studying the imact these effects have on observed galaxy morphologies. For Galaxy Zoo, such images are particularly useful for measuring the effects of redshift on the volunteer classifications. Through classifications made on a set of artificially redshifted galaxies, any dependence they might have as a function of redshift can be measured, allowing a correction to be applied to classifications on images of real, high-redshift galaxies. The details of this type of debiasing technique will be described in Section 5.3. This section will first provide a breif summary of how the FERENGI code performs the artificial redshifting.

To create realistic images that mimic the seeing and resolution of HST ACS, the FERENGI redshifting procedure consists of four steps (explained in detail in Barden et al. (2008), but here a simplified outline):

i: Modify angular size and surface brightness

FERENGI first rescales the input image by computing the angular size transformation of the galaxy from its input redshift z_i to output reshift z_o. The angular size a of a distant object is proportional to $a \propto d/(1+z)^2$ (using $\tan(a)$=a for small angles), where d is the luminosity distance to the object. In units of pixels, the transformation from input angular size n_i to output n_o can be expressed as:

$$\frac{n_o}{n_i} = \frac{d_i/(1+z_i)^2}{d_o/(1+z_o)^2}\frac{p_i}{p_o} \tag{5.1}$$

with an input pixel scale p_i (in this thesis $p_i = 0.396''$/pix corresponding to SDSS) and p_o (0.03"/pix, corresponding to ACS). From here a transformation between the observed fluxes is computed, assuming the absolute magnitude is conserved at both redshifts.

FERENGI also offers an option to apply an evolutionary correction to the absolute magnitude, which is helpful for a fair comparison of real and artificial high redshift morphologies. Artificially redshifted galaxies will appear much dimmer than their low redshift counterparts if absolute magnitude is conserved. Since galaxies intrinsically tend to be brighter at high redshift, visual classification of real galaxies cannot be compared as accurately to dimmer, simulated galaxies. To brighten galaxies in a similar way to real galaxies, a magnitude correction e can be input using a linear function:

$$M_{evo} = e \times z + M \tag{5.2}$$

where e represents the magnitude difference between two redshifts separated by $\Delta z = 1$.

ii: Account for bandpass shifting

As a consequence of cosmological expansion, the flux from a source measured using a broadband filter will not, in general, perfectly correspond to the rest-frame flux emitted at the target wavelength range of the filter. Rather, since observed wavelengths are redder than emitted wavelengths as a function of redshift ($\lambda_{obs} = \lambda_{rest-frame}(1 + z)$), filters will tend to pick up light that is bluer (in the galaxy's rest-frame) than its target wavelength; this effect is known as *bandpass shifting.* In order to produce fluxes that mimic those measured by ACS at high redshifts, FERENGI simulates the bandpass shifting effects by applying a correction to the output flux calculated via the IDL routine KCORRECT, which incorporates spectral template models from Bruzual & Charlot (2003), to measure the expected shifts in flux for a given output filter.

iii: Point Spread Function and noise

In order to best mimic the HST ACS resolution, the image is then convolved with a PSF created to be as close as possible in shape and width to the ACS PSF. This is done by deconvolving a typical ACS PSF with the input SDSS PSF for each galaxy. This technique works well in general but has limitations - mainly, the widths of the in- and output PSFs must be sufficiently different. If they are comparable, the convolving function can become too narrow. In these cases, the image will be introduced to noise which results in ringing patterns and other oddities (examples of images with this effect are shown in Section 5.4.1). Since the difference in PSF widths increases with redshift, this imposes a minimum redshift at which FERENGI can successfuly create images for any given galaxy (discussed more in Section 5.2). Last, Poissonian noise is added to each pixel.

5.2 The FERENGI sample

To generate an artificially redshifted sample of galaxies to be used in debiasing the Galaxy Zoo: Hubble catalog, a source sample was generated [1] consisting of 288 galaxies from SDSS, all of which were previously classified in GZ2. These galaxies were chosen to span a wide range of morphologies, surface brightnesses, and redshifts. Seven morphological classes were considered: spiral galaxies, edge-on disks without a bulge, edge-on disks with a bulge, face-on disks with a bulge, galaxies with any features, galaxies undergoing mergers, and barred galaxies. For each of these categories, galaxies were chosen with from three "strength" bins, defined using the GZ2 vote fractions. Weak strengths were defined as having $f_{class} < 0.2$, intermediate as $0.2 < f_{class} < 0.8$, and strong as $f_{class} > 0.8$. In each strength bin, galaxies were also chosen to represent three different surface brightnesses: $\mu_r > 21.5$, $20.5 < \mu_r < 21.5$, and $\mu_r < 20.5$. Finally, from each morphological class, strength, and surface brightness bin, one galaxy was chosen for four redshift bins: $z < 0.013$, $0.013 < z < 0.02$, $0.02 < z < 0.025$, and $z > 0.025$, with the exception of the bar class, in which two galaxies were chosen for each redshift bin, doubling the sample size for that class.

The 288 SDSS galaxies were processed with the FERENGI code to mimic *HST* imaging parameters [2] , in order to ultimately measure and correct any redshift-dependant biases in the classifications of the real *HST* images. I-814 and V-606 images, chosen to match the *HST* ACS AEGIS imaging, were output for each subject at a range of redshifts and with a range of applied evolution factors. The range of simulated redshifts possible for any galaxy is dependent on the intrinsic redshift and size of the source galaxy, since the simulated images cannot be resampled at better angular resolution than the original SDSS data. This imposes a minimum simulated "target" redshift that can be achieved for each galaxy. For the lowest redshift bin in the source sample ($z < 0.013$), galaxies could be redshifted the full range of $0.3 < z < 1.0$, in increments of $dz = 0.1$. For the second lowest redshift bin, galaxies could only be redshifted in the range $0.5 < z < 1.0$, for the third, galaxies could be redshifted in the range $0.8 < z < 1.0$, and for the highest redshift bin, galaxies were only redshifted in FERENGI to $z = 1.0$. Only galaxies which were redshifted the full range were considered in the

[1] The source sample for FERENGI was created by the Galaxy Zoo science team in 2012.
[2] This work was done by Edmond Cheung, a Galaxy Zoo science team member.

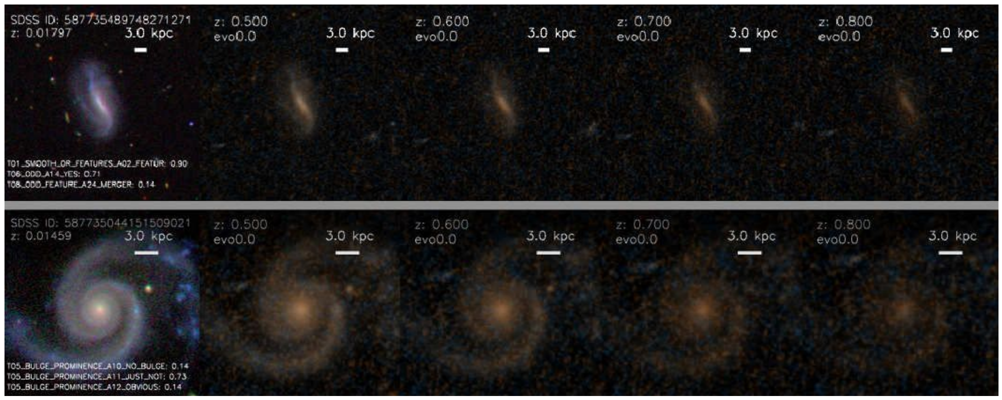

Figure 5.1 Examples of two SDSS galaxies which have been run through the FERENGI code to produce simulated *HST* images. The measured value of f_{features} from GZH for the images in each panel are (1) Top row: f_{features} = (0.900, 0.625, 0.350, 0.350, 0.225) and (2) Bottom row: f_{features} = (1.000, 0.875, 0.875, 0.625, 0.375).

debiasing procedure outlined in the next section (5.3.1), because the method calibrates galaxies to a low redshift of $z = 0.3$, data for which is not available for galaxies in the remaining three redshift bins. Last, for each simulated redshift, a range of evolution factors was applied from $0 < e < 3$ in increments of $de = 0.5$.

The final FERENGI sample totals 6,624 simulated images which were classified as part of GZ4, using the same decision tree as used in GZH. The debiasing technique described next (Section 5.3) used only the 4,446 images corresponding to the 72 galaxies which were redshifted the full $0.3 < z < 1.0$ range. Because the debiasing method takes into account surface brightness as a parameter, photometry was measured for all images using SEXTRACTOR [3] . The mean surface brighness μ within effective radius (R_e) was calculated as:

$$\mu = m + 2.5 * \log_{10}(2 \times (b/a) \times \pi R_e^2) \tag{5.3}$$

where 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% `FLUX_RADIUS` converted into arcsec.

[3] SEXTRACTOR measurements for the original FERENGI sample were done by Tom Melvin, a former Galaxy Zoo science team member.

5.3 Measuring the drop in f_{features} as a function of z and μ using FERENGI data

5.3.1 Identifying "correctable" and "lower limit" samples.

The objective is to use the simulated data from FERENGI to predict, for a galaxy imaged at a redshift z, and with a measured $f_{\text{features},z}$ value, what its f_{features} value *would have been* if it had been viewed at $z = 0.3$. This predicted value is defined as the "debiased' vote fraction $f_{\text{features,debiased}}$, and is calculated by applying a correction to the measured value of f_{features}.

The amount that a galaxy's f_{features} vote fraction must be corrected is assumed to primarily depend on the apparent size and brightness of the galaxy. As described in 5.1, these factors will affect the overall clarity of the image viewed by the GZ volunteers, which in turn affects the likelihood of being able to identify distinct feature. The apparent size and brightness are controlled by both instrinsic parameters (absolute size and luminosity), and extrinsic (distance to the galaxy). The change in f_{features} then is measured as a function of redshift (z, an extrinsic freature, measuring distance to the galaxy), and surface brightness (μ, an intrinsic feature, taking into account both brightness and size).

Figure 5.2 shows the change in f_{features} for FERENGI galaxies in bins of redshift and surface brightness. Points in each z, μ represent individual FERENGI galaxies. On the x-axis of each bin is the value of f_{features} measured in that galaxy's $z = 0.3$ image (the lowest redshift of the simulated images). On the y-axis of each bin is the value of f_{features} measured in that galaxy's $z = z$ image, where z corresponds to the redshift associated with that bin. As predicted, the value of f_{features} measured at a higher redshift, z, is, in general, *lower* than the value measured at lower redshift, $z = 0.3$, *for the same galaxy.* This effect is strongest as redshift increases (to the right in Figure 5.2) and as surface brightness decreases (upwards in Figure 5.2).

A reliable predicted value can be obtained so long as the relationship between $f_{\text{features},z}$ and $f_{\text{features},z=0.3}$ is single-valued; that is, for a given $f_{\text{features},z}$, there is exactly one corresponding value of f_{features} at $z = 0.3$. Unfortunately, this is *not* always the case. Figure 5.3 shows f_{features} measured at $z = 1$ vs f_{features} measured at $z = 0.3$ for FERENGI galaxies with average surface brightnesses $< \mu >= 20.8$ (a zoomed-in version

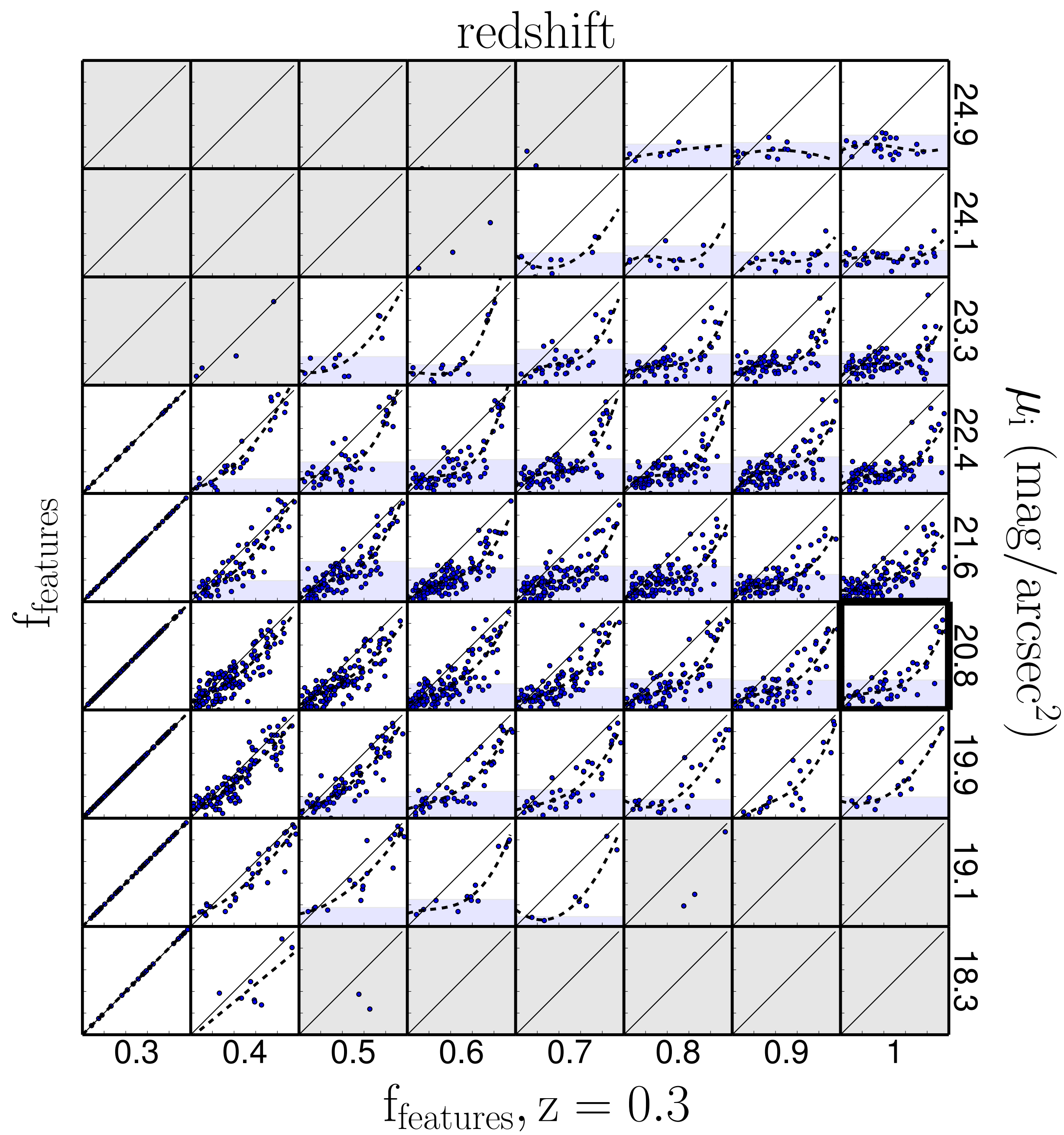

Figure 5.2 Effects of redshift bias in 3,449 images in the FERENGI sample. Each point *in a given redshift and surface brightness bin* represents a unique galaxy. On the y-axis in each bin is the f_{features} value of the image of that galaxy redshifted to the value corresponding to that redshift bin. 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-fit polynomials to the data in each square. The solid black line represents $f_{\mathrm{features},z}$=$f_{\mathrm{features},z=0.3}$. 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 Figure 5.3.

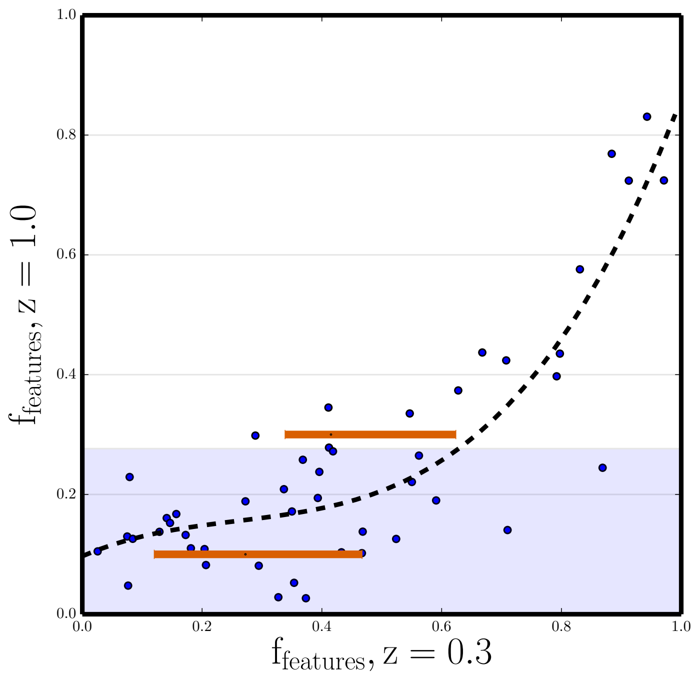

Figure 5.3 A larger version of the dark-outlined square in Figure 5.2, containing FERENGI galaxies that have been artificially redshifted to $z = 1.0$ and have surface brightnesses between $20.3 < \mu < 21.0$ (mag/arcsec^2). The orange bars represent the inner 68% (1σ) of the uncorrectable f_{features} quantiles, which are used to compute the limits on the range of debiased values.

of the dark outlined bin in Figure 5.2). This figure shows that if the value of f_{features} measured for a galaxy at $z = 1$ is particularly low, there is a wide range that f_{features} could have been if measured at $z = 0.3$. Therefore, a low measured value of f_{features} at high redshift could represent two morphological types of galaxies: 1) The galaxy has no distinguishable features and may be classified as a smooth elliptical, or 2) the galaxy *does* have features, but these have become blurred and too difficult to detect at high redshift.

It is important to identify such regions of surface brightness/redshift/f_{features} space since vote fractions cannot be confidentely corrected to a single value for galaxies in these regions. The criteria for determining whether a region of this space is single-valued, and therefore correctable, is as follows: In each surface brightness and redshift bin, the relationship between $f_{\mathrm{features},z}$ and $f_{\mathrm{features},z=0.3}$ is modelled by fitting the data with polynomials of degress 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 Figures 5.2 and 5.3. Flat regions of the bins are areas in which there is *not* a clear single-valued relationship between $f_{\mathrm{features},z}$ and $f_{\mathrm{features},z=0.3}$. This is quantified by measuring the slope of the best-fit polynomial to the vote fractions; regions of the bins with a slope less than 0.4 are considered *not* one-to-one, and therefore $f_{\mathrm{features},z}$ cannot be boosted to its $f_{\mathrm{features},z=0.3}$ value. These are colored blue in Figure 5.2 and 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.

The unshaded regions of Figure5.3 define descrete ranges of redshift, surface brightness, and f_{features} within which a galaxy must lie in order for the debiased vote fraction to be confidentely 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_{features}. To define this smooth space, a convex hull was calculated to enclose the correctable and lower-limit FERENGI galaxies in the $z - \mu - f_{\mathrm{features}}$ space (see Figure 5.4). The space defined by this hull was used to ultimately separate the GZH galaxies into correctable samples (those for which a correction to f_{features} can confidentely be applied, see next section) and lower-limit samples (those for which a single-valued correction cannot be applied). The final categorization of the GZH sample, split by imaging survey, is shown in Table 5.1.

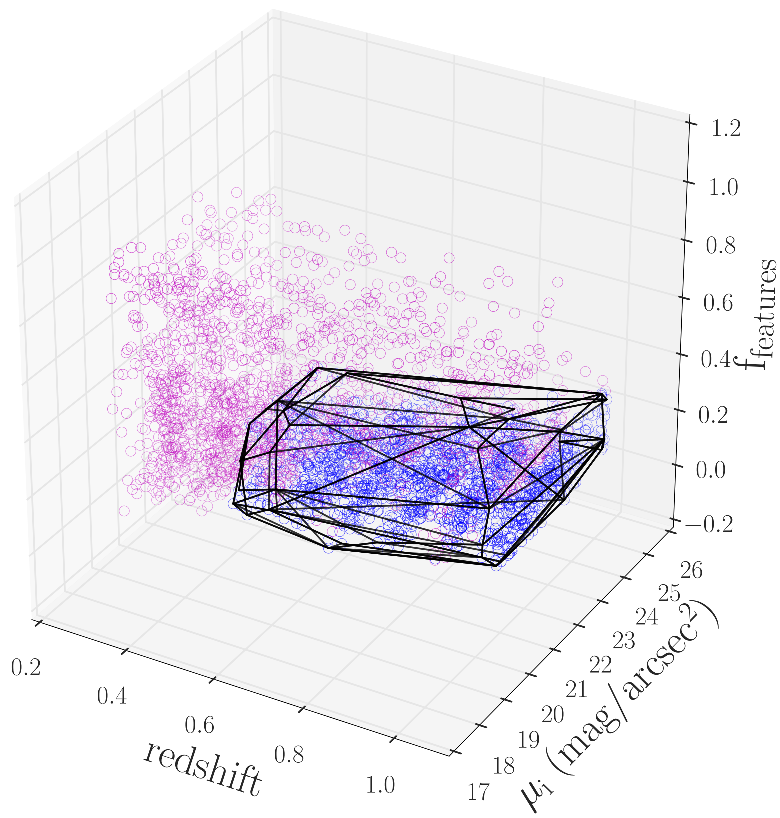

Figure 5.4 The final separation of the correctable and lower-limit samples in redshift/surface brightness/f_{features} space.**Pink** points are all FERENGI galaxies in the **unshaded** regions of Figure 5.2. **Blue** points are all FERENGI galaxies in the **blue shaded** regions of Figure 5.2. The solid black line is the convex hull which encloses the uncorrectable points and defines the region of the lower-limit sample.

Table 5.1 Number of correctable galaxies for the top-level task in GZH, split by *HST* survey.

	Correction type	AEGIS	COSMOS	GEMS	GOODS-N 5-epoch	GOODS 5-epo
correctable	0	2,908	21,169	2,802	1,459	1,1
lower-limit	1	833	5,169	1,021	1,377	1,2
no correction needed ($z \leq 0.3$)	2	955	10,870	1,175	415	4
not enough information (NEI)	3	2,677	43,058	3,559	2,077	2,1
no redshift information	4	1,134	4,688	530	687	1
total		8,507	84,954	9,087	6,015	5,1

For the "lower limit" galaxies, since a single debiased f_{features} value cannot be confidentely assigned, a *range* of debiased values is estimated. In each z, μ bin in Figure 5.2, the spread of intrinsic values of $f_{\text{features},z=0.3}$ for five quantiles of observed f_{features} is computed - these are denoted by the gray lines in the close-up Figure 5.3. The range of intrinsic values of f_{features} is defined by the upper and lower 1 σ limits, enclosing the inner 68% of the data; this is represented by the orange bars in Figure 5.3. For any galaxy which cannot be directly debiased, these ranges are used to denote the upper and lower limits on the expected values $f_{\text{features},z=0.3}$ as a function of the observed f_{features}.

5.3.2 Computing debiased f_{features} for the "correctable" sample using the ζ equation

For the "correctable" sample of simulated FERENGI galaxies, an equation is derived to model the dropoff in f_{features} with redshift for each galaxy. Such a model is assumed to have the following criteria: (1) For a given galaxy, f_{features} should decrease relative to its $f_{\text{features},z=0.3}$ as redshift increases. (2) The corrected f_{features} value must be contained within 0 and 1, since it is a fraction. (3) The degree of dropoff may depend on the surface brightness of the galaxy. Given these three assumptions, a simple exponential function was derived:

$$f_{\mu,z} = 1 - (1 - f_{\mu,z=0.3})e^{\frac{z-z_0}{\zeta}} \tag{5.4}$$

where $f_{\mu,z=0.3}$ is the vote fraction at the lowest redshift in the artificially-redshifted FERENGI sample ($z_0 = 0.3$). ζ is a parameter that controls the rate at which f_{features}decreases with redshift.

Equation 5.4 is then fit to each galaxy in the "correctable" FERENGI sample, and ζ is measured for each. Figure 5.5 shows the best fit equations for 16 galaxies, and the ζ corresponding to the best fit is displayed with each galaxy. As it was assumed that surface brightness likely plays a role in the level of dropoff in f_{features}, and hence the value of ζ which controls this dropoff, it is assumed that ζ follows a simple linear dependence with surface brightness:

$$\log_{10}(\hat{\zeta}) = \zeta_0 + (\zeta_1 \times \mu), \tag{5.5}$$

where $\hat{\zeta}$ is the correction factor applied to each galaxy. Figure 5.6 shows the relationship between the derived ζ values and the surface brightness μ of the FERENGI galaxies, which is fit with equation 5.5. The best-fit parameters to this linear fit from least-squares optimization are $\zeta_0 = 0.50$, $\zeta_1 = -0.03$. Interestingly, only a very weak surface brightness dependence is detected. It is difficult to determine from these data whether the weak detection is due to a true lack of dependence, or insufficient data (only 28 galaxies had sufficient data to accurately measure ζ).

Using the ζ parameters measured in the FERENGI sample, a final debiased correction equation is derived to correct the f_{features} vote fractions in the HST data:

$$f_{\mathrm{features,debiased}} = 1 - (1 - f_{\mathrm{features,weighted}})e^{\frac{-(z-z_0)}{\hat{\zeta}}} \tag{5.6}$$

where $f_{\mathrm{features,weighted}}$ is the weighted vote fraction, and $f_{\mathrm{features,debiased}}$ is bounded between $f_{\mathrm{features,weighted}}$ and 1.

5.3.3 Results and the NEI sample: limitations of the ferengi simulated data

- show table of HST sample breakdown 5.1
- show how above method only works for HST galaxies with z/mu corresponding to ferengi space 5.8
- discuss why ferengi can't completely replicate mu distribution of HST (cite Karen response to referee)

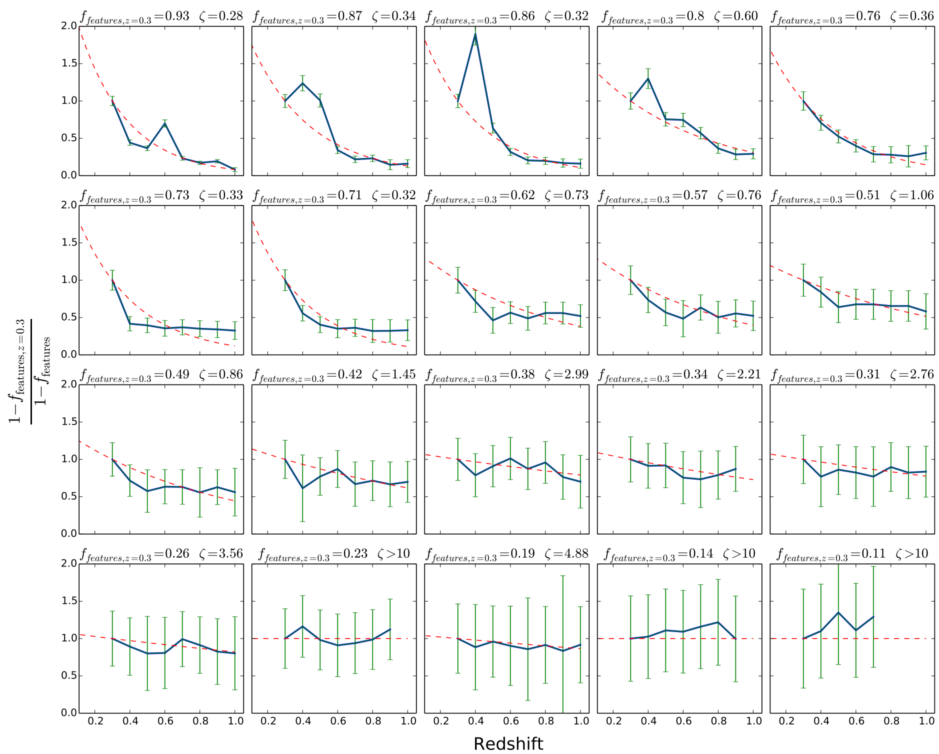

Figure 5.5 Behaviour of the normalised, weighted vote fractions of features visible in a galaxy (f_{features}) as a function of redshift in the artificial FERENGI images. Galaxies in this plot were randomly selected from a distribution with evolutionary correction $e = 0$ and at least three detectable images in redshift bins of $z \geq 0.3$. The displayed bins are sorted by $f_{\mathrm{features},z=0.3}$, labeled above each plot. Measured vote fractions (blue solid line) are fit with an exponential function (red dashed line; Equation 5.4); the best-fit parameter for ζ is given above each plot.

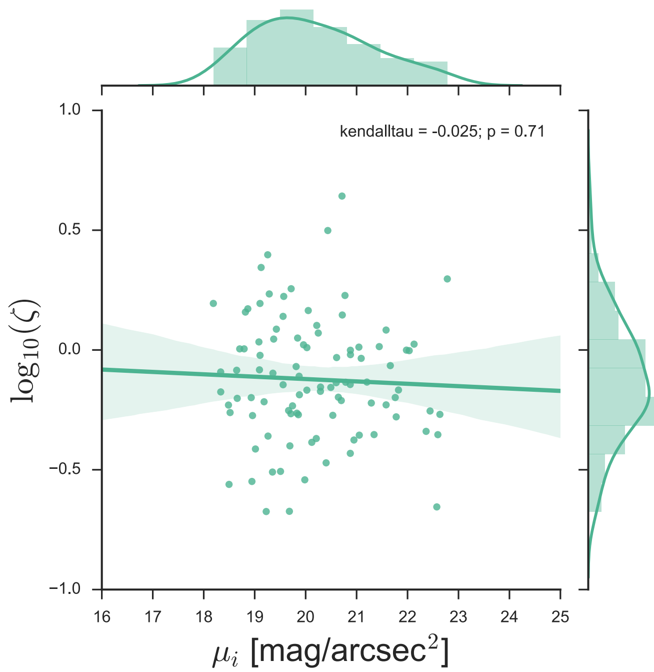

Figure 5.6 All fits for the FERENGI galaxies of the vote fraction dropoff parameter ζ for f_{features} as a function of surface brightness. This includes only the simulated galaxies with a bounded range on the dropoff ($-10 < \zeta < 10$) and sufficient points to fit each function (28 original galaxies, each with varying images artificially redshifted in one to eight bins over a range from $0.3 \lesssim z_{\mathrm{sim}} \lesssim 1.0$).

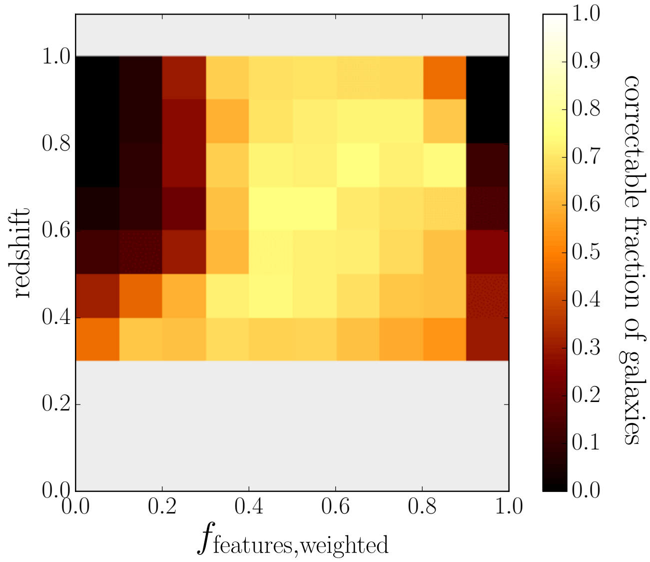

Figure 5.7 Histogram showing the fraction of galaxies that have a finite correction for the debiased vote fractions $f_{\mathrm{features,debiased}}$ as a function of f_{features} and redshift. The parameter space for corrections is limited to $0.3 \leq z \leq 1.0$ due to the sampling of the parent SDSS galaxies and detectability in the FERENGI images.

- discuss limits shown in 5.7

5.4 Ferengi 2: using simulated images to measure incompleteness in disk fraction as a function of redshift and surface brightness

In the previous section I described how we used the simulted FERENGI images to measure the redshift/surface brightness dependence on f_{features}, and applied the measurements towards a correction factor to the vote fractions directly. In this section I will describe the motivation, selection, and application of a second set of FERENGI images used to measure and correct for the incompleteness in *number* of disks detected as a function of redshift and surface brightness, used in the work described in Chapter 6.

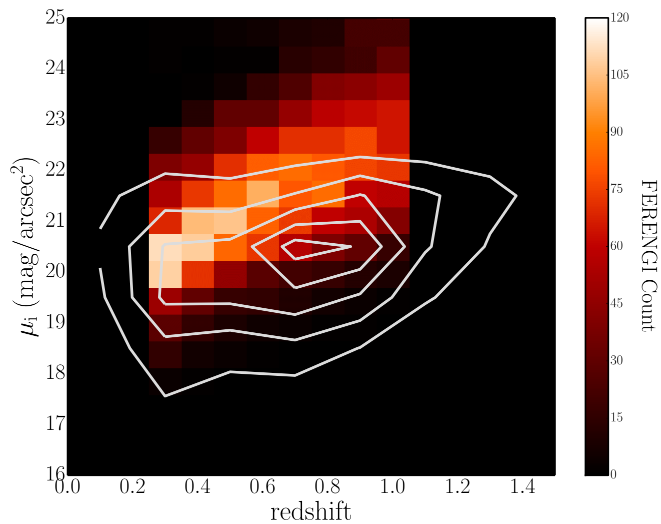

Figure 5.8 Surface brightness as a function of redshift for 3,449 FERENGI images and the 102,548 `main` galaxies with measured μ and z values. The colour histogram shows the number of FERENGI images as a function of μ and $z_{\rm sim}$. White contours show counts for the galaxies in the `main` sample, with the outermost contour starting at $N = 1500$ and separated by intervals of 1500.

5.4.1 The Ferengi 2 Sample

The creation of a second set of FERENGI images was motivated by the scientific goal of measuring the redshift evolution of the fraction of red disk galaxies using the Galaxy Zoo: Hubble dataset. This project is described in full in Chapter 6, but the reasons requiring a new set of simulated images will be described briefly here. First, as described in the previous section, the analysis of the first FERENGI set revealed that, for a large area of z-μ parameter space, galaxies with low measured values of f_{features} could not be corrected to a point that could clearly distingush them as disks with washed-out features or ellipticals. Due to this limitation, any measurement of the number of disk galaxies in a given redshift interval can only be reported as a *lower-limit* to the true value. The difference in the measured lower-limit and the true number of disks is what we will refer to as the *incompleteness* in number of disks detected.

It is possible, then, to use the FERENGI images to measure this incompleteness by measuring the number of disks detected at a given redshift, and comparing to the number of disks detected out of the same galaxies at the lowest redshift (this would be considred the true, or intrinsic, number of disks.) The details of this approach will be described in the next section. A complication specific to this project is that the number of disks will be ultimitely used to compute the *red disk fraction*, that is, the ratio of the number of red disks to all disks, as a function of redshift. It is then necessary to measure the level of incompleteness for both red and blue galaxies separately, to calculate this fraction most accurately.

The color separation method for the *HST* galaxies in Chapter 6 uses NUV, r, and J magnitudes. To separate the FERENGI sample of galaxies into red and blue samples in the same way, these magnitudes are required. In the first set, however, only 44 of the 288 galaxies had these data available, which were not enough to properly measure any incompleteness, especially after binning the data further in surface brightness and redshift. So, a larger set of galaxies to be artificially redshifted, all which had the aforementioned data necessary to separate by color, was required.

This set of new galaxies to be put through the FERENGI code, hereafter refered to as the FERENGI 2 sample, was selected as follows: All candidates were pulled from a parent sample of all SDSS galaxies which had previously been classified in GZ2. As discussed in Section 5.2, only galaxies with redshifts below $z < 0.013$ were able to be redshifted the full simulated redshift range $0.3 < z < 1.0$, so a redshift cut was implemented of $z < 0.013$. These galaxies were cross-matched with catalogs from GALEX (Martin et al., 2005) for NUV magnitudes and 2MASS (Skrutskie et al., 2006) for J magnitudes. 1,435 galaxies fit these criteria.

Bulk SDSS u, g, r, i, and z-band fits images were then downloaded for all 1,435 galaxy candidates [4] . Cutouts were made for each galaxy, using the 90% r-band petrosian radius to set the size of the cutout (PETROR90_R). The default prescription used was to define the edges as 2.5*PETROR90_R, measured from the galaxy as the center. If the galaxy was within this distance from the edge of the bulk fits image, 2.0*PETROR90_R was used. Cutouts were not made for galaxies within this distance from the edge, both to ensure the full galaxy was visible in all cutouts in the sample, and to avoid over-zooming the image. 187 galaxies were thus removed from FERENGI2; an example of such a galaxy "too close" to the edge of the edge is shown in Figure 5.9.

While all 78 $z < 0.013$ galaxies from the original FERENGI sample were successfuly simulated to a minimum redshift of $z_{sim} = 0.3$, this was not always true for the FERENGI2 candidates. Redshift of the source galaxy is the largest factor in determining the minimum possible simulated redshift, but other factors including the size of the psf and physical size of the source galaxy also come into play. All 1,248 candidates were then put through FERENGI at only the lowest redshift $z_{sim} = 0.3$ to begin, and each image was visualy inspected to determine whether the code succeeded. 312 "failures" were detected; two examples are shown in Figure 5.10. The remaining 936 "successes" were then artificially redshift the full range of $0.3 < z < 1.0$ in increments of $dz = 0.1$; these make up the final FERENGI2 sample of 7,488 images of 936 galaxies redshifted 8 times. A single evolution factor, rather than a range, of $e = 1$ was applied to all images. This value was chosen by analyzing the spectra template models of Brinchmann et al. (2004), which showed that the most typical galaxies evolve in brightness by one magnitude per redshift. Example images are shown in Figure 5.11

[4] http://data.sdss3.org/bulkFields

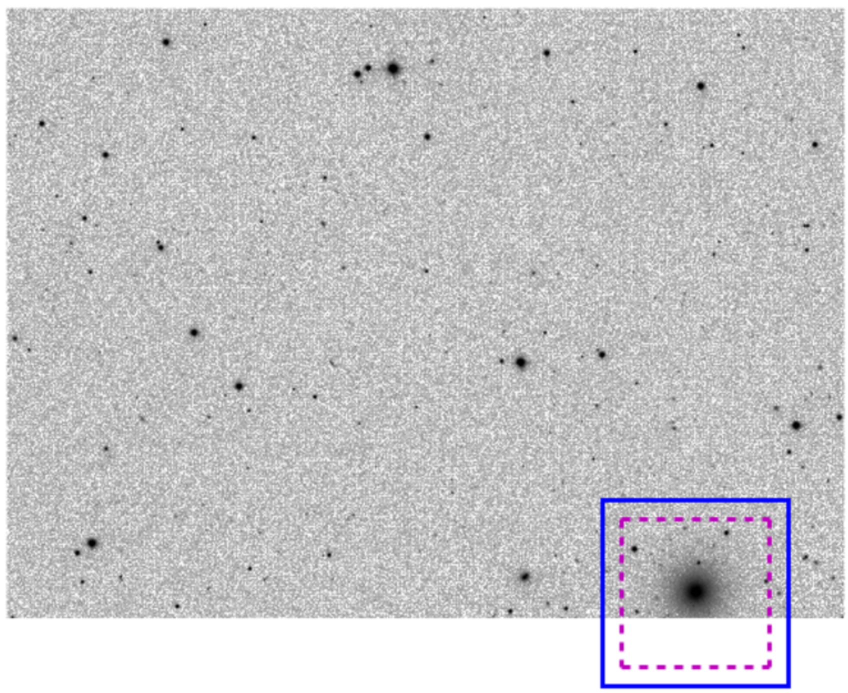

Figure 5.9 Example of a galaxy overlapping the edge of the SDSS frame. Shown is the bulk r-band fits image for SDSS DR12 run 3903, camcol 6, and field 60. The boxed-in galaxy (SDSS DR12 objid 1237662239079268544) is too close to the edge of the image to create a cutout that encloses the entire galaxy. The pink dashed box indicates a cutout size of 2*PETROR90_R, the blue solid line indicates a cutout size of 2.5*PETROR90_R.

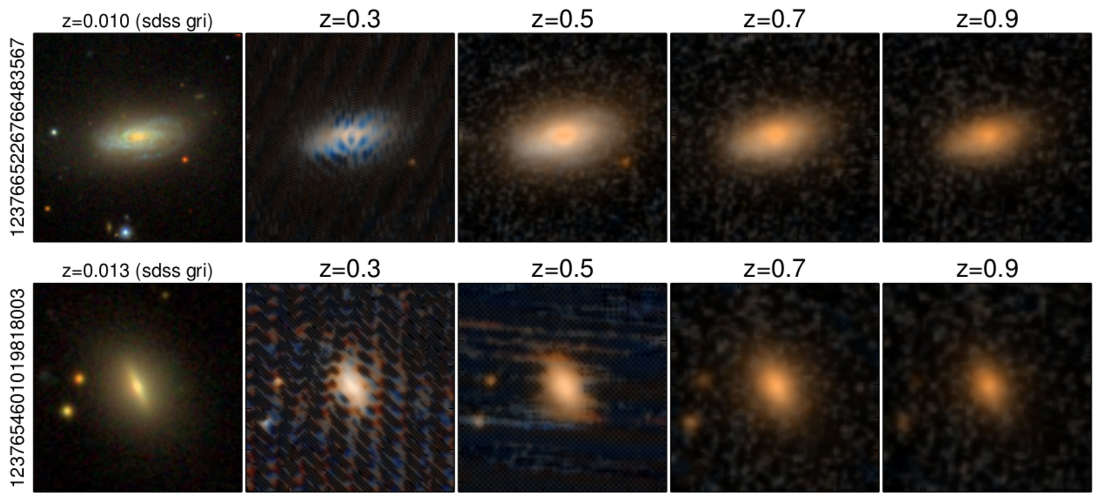

Figure 5.10 Examples of two galaxies whose minimum simulated redshifts in FERENGI were larger than $z_{sim} = 0.3$. These were detected via visual inspection and removed from the final FERENGI2 sample.

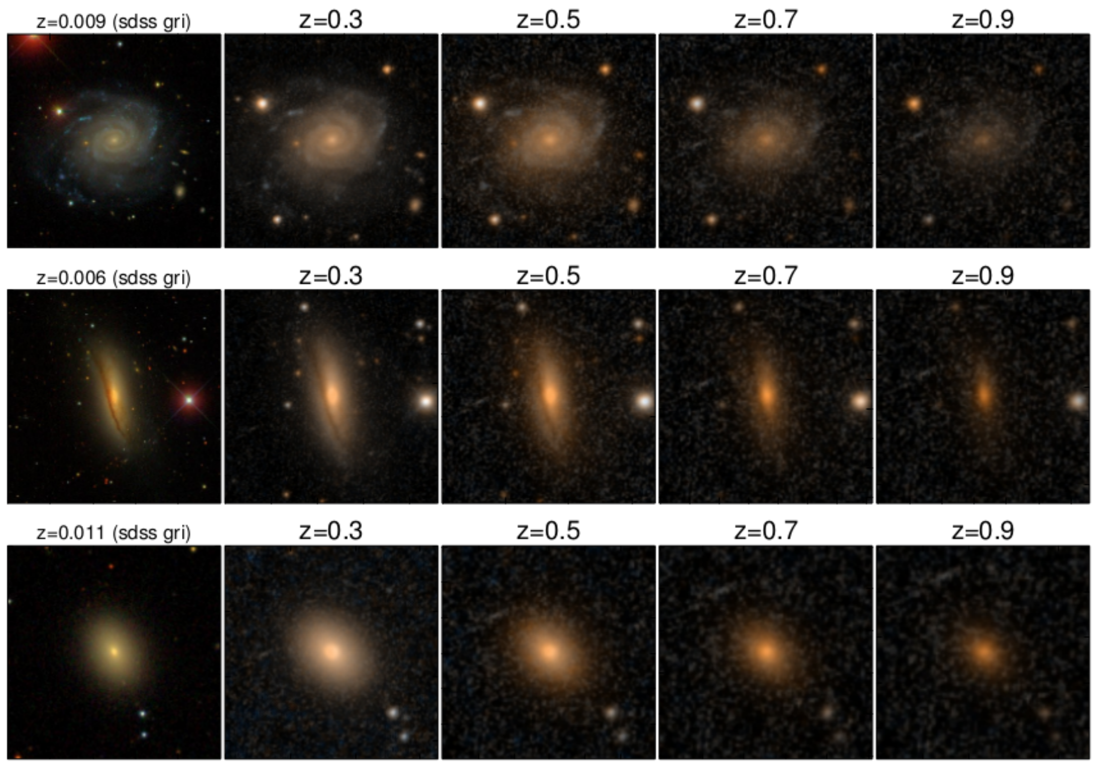

Figure 5.11 Examples of FERENGI2 galaxies. The left is the original gri-composite image of the source galaxy. Images on the right are simulated output from the FERENGI code. Only four of the eight simulated redshifts are shown in the interest of space.

Chapter 6

GZH red disk fraction

red disk fraction project

Chapter 7

Summary & Future Work

References

Anderson, D. P. 2002, Commun. ACM, 45, 56

Barden, M., Jahnke, K., & Häußler, B. 2008, The Astrophysical Journal Supplement Series, 175, 105

Block, D. L., Buta, R., Knapen, J. H., et al. 2004, The Astronomical Journal, 128, 183

Block, D. L., & Puerari, I. 1999, Astronomy & Astrophysics, 342, 627

Block et al., D. 1994, Astronomy & Astrophysics, 288, 365

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, Monthly Notices of the Royal Astronomical Society, 351, 1151

Bruzual & Charlot. 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000

Buta, R. J., Sheth, K., Regan, M., et al. 2010, The Astrophysical Journal Supplement Series, 190, 147

Conselice, C. J. 2003, The Astrophysical Journal Supplement Series, 147, 1

D. Block, R. W. 1991, Nature, 353, 48

de Vaucouleurs, G. 1963, The Astrophysical Journal Supplement Series, 8, 31

—. 1991, Third Reference Catalogue of Bright Galaxies. (New York: Springer)

Dieleman, S., Willett, K. W., & Dambre, J. 2015, Monthly Notices of the Royal Astronomical Society, 450, 1441

Eskridge, P. B., Frogel, J. A., Pogge, R. W., et al. 2000, The Astronomical Journal, 119, 536

—. 2002, The Astrophysical Journal Supplement Series, 143, 73

Fortson, L., Masters, K., Nichol, R., et al. 2011, 11

Fukugita, M., Nakamura, O., Okamura, S., et al. 2007, The Astronomical Journal, 134, 579

Galloway, M. A., Willett, K. W., Fortson, L. F., et al. 2015, Monthly Notices of the Royal Astronomical Society, 448, 3442

Hackwell, J., & Schweizer, F. 1983, The Astrophysical Journal, 265, 643

Hart, R. E., Bamford, S. P., Willett, K. W., et al. 2016, Monthly Notices of the Royal Astronomical Society, 461, 3663

Hubble, E. 1926, The Astrophysical Journal, 64, 321

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, The Astrophysical Journal Supplement Series, 221, 8

Knapen, J. H., Shlosman, I., & Peletier, R. F. 2000, The Astrophysical Journal, 529, 93

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, Monthly Notices of the Royal Astronomical Society, 379, 1599

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, Monthly Notices of the Royal Astronomical Society, 389, 1179

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, The Astrophysical Journal, 619, L1

Masters, K. L., Mosleh, M., Romer, A. K., et al. 2010, Monthly Notices of the Royal Astronomical Society, 405, 783

MenendezDelmestre, K., Sheth, K., Schinnerer, E., Jarrett, T. H., & Scoville, N. Z. 2007, The Astrophysical Journal, 657, 790

Mulchaey, J. S., & Regan, M. W. 1997, The Astrophysical Journal, 482, L135

Nair, P. B., & Abraham, R. G. 2010, The Astrophysical Journal Supplement Series, 186, 427

Odewahn, S. C., Cohen, S. H., Windhorst, R. A., & Philip, N. S. 2002, The Astrophysical Journal, 568, 539

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, The Astronomical Journal, 124, 266

Sandage, A. 1961, The Hubble Atlas of Galaxies (Washington: Carnegie Institution)

Scarlata, C., Carollo, C. M., Lilly, S., et al. 2007, The Astrophysical Journal Supplement Series, 172, 406

Schawinski, K., Thomas, D., Sarzi, M., et al. 2007, Monthly Notices of the Royal Astronomical Society, 382, 1415

Scoville et al., N. Z. 1988, The Astrophysical Journal Letters, 327, L61

Seigar, M. S., & James, P. A. 1998, 11

Sheth, K., Regan, M. W., Scoville, N. Z., & Strubbe, L. E. 2003, The Astrophysical Journal, 592, L13

Sheth, K., Elmegreen, D. M., Elmegreen, B. G., et al. 2008, The Astrophysical Journal, 675, 1141

Sheth, K., Regan, M., Hinz, J. L., et al. 2010, Publications of the Astronomical Society of the Pacific, 122, 1397

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, The Astronomical Journal, 131, 1163

Thronson et al., H. 1989, The Astrophysical Journal, 343, 158

Warren, S. J., Cross, N. J. G., Dye, S., et al. 2007, arXiv:0703037

Whyte, L. F., Abraham, R. G., Merrifield, M. R., et al. 2002, Monthly Notices of the Royal Astronomical Society, 336, 1281

Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, Monthly Notices of the Royal Astronomical Society, 435, 2835