Clustering Constraints on the Relative Sizes of Central and Satellite Galaxies

Andrew Hearin[1], Peter Behroozi[2], Andrey Kravtsov[3], Benjamin Moster[4]

[1] *Argonne National Laboratory, Argonne, IL, USA 60439, USA*

[2] *Department of Physics, University of Arizona, 1118 E 4th St, Tucson, AZ 85721 USA*

[3] *Department of Astronomy & Astrophysics, The University of Chicago, Chicago, IL 60637 USA*

[4] *Universitäts-Sternwarte, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 München, Germany*

4 November 2017

ABSTRACT

We place empirical constraints on the connection between dark matter halos and galaxy half-light radii, $R_{1/2}$. Low-redshift SDSS measurements show that smaller galaxies cluster much more strongly than larger galaxies at fixed stellar mass. Using `Halotools` to forward model the observations, we find that the clustering signal generically requires satellite galaxies to be smaller than central galaxies of the same halo mass. We present a simple empirical model consistent with the clustering results, in which galaxy size is proportional to halo virial radius at the time of peak halo mass. We use this model to predict how galaxy lensing, $\Delta\Sigma$, should depend on $R_{1/2}$ for M_*-complete samples. Other simple empirical models fail the clustering test, such as models in which galaxy size is related to stellar mass alone; these failures persist even when accounting for possible effects from satellite stripping and orphan galaxies. Our results suggest that the relative size of centrals and satellites is predetermined at the time of satellite infall, and that a remarkably simple galaxy–halo scaling relation emerges from the complex physics regulating galaxy size.

1 INTRODUCTION

In the ΛCDM framework of cosmological structure formation, galaxies form at the centers of dark matter halos. Highly complex and nonlinear baryonic processes regulate galaxy formation, and the quest for a fine-grained understanding of these processes is one of the chief goals of theoretical astrophysics today.

Observationally, many properties of observed galaxies exhibit remarkably tight scaling relations. Among the most fundamental of these relations is the strong correlation between galaxy size and stellar mass and luminosity. The scaling of galaxy size with galaxy mass is well-measured in the local Universe (Shen et al. 2003; Guo et al. 2009; Huang et al. 2013; Zhang & Yang 2017) and at high-redshift (Trujillo et al. 2004; van der Wel et al. 2014; Kawamata et al. 2015; Shibuya et al. 2015; Huertas-Company et al. 2013; Lange et al. 2015; Huang et al. 2017).

These well-measured scaling relations are challenging to faithfully recover using ab initio galaxy formation methods such as hydrodynamical simulations and semi-analytic models, and provide useful boundary conditions for the calibration of such modeling efforts (Khochfar & Silk 2006; Dutton et al. 2011; Hopkins et al. 2010; Bottrell et al. 2017). The precision cosmology program also depends critically on accurate modeling of galaxies, so that cosmological parameter inference can be confidently

conducted without undue interference from uncertainty in baryonic physics (LSST Science Collaboration et al. 2009; Robertson et al. 2017).

2 DATA AND SIMULATIONS

AVK comment: Need to be much clearer and more explicit about which galaxy properties come from which dataset, and probably to reiterate this when discussing the Figure 3 plot of the size-mass relation.

Our galaxy sample comes from the catalog of SDSS galaxy profile decompositions provided by Meert et al. (2015). This catalog is based on Data Release 10 of the Sloan Digital Sky Survey (SDSS, Ahn et al. 2014), with improvements to the photometry pipeline and light profile fitting methods (Vikram et al. 2010; Bernardi et al. 2013, 2014; Meert et al. 2013). In the version of this catalog that we use, two-dimensional $r-$band profiles were fit with a two-component de Vaucouleurs + exponential profile to determine the half-light radius $R_{1/2}$. We use stellar mass measurements made available through the MPA-JHU catalog (Kauffmann et al. 2003; Brinchmann et al. 2004), allowing us to define volume-limited samples of galaxies according to the same completeness cuts used in Behroozi et al. (2015) (see Figure 2).

We calculate two-point clustering $w_{\rm p}$ of our SDSS galaxy sample using line-of-sight projection of $\pi_{\max} =$

20Mpc using the `correl` program in `UniverseMachine`. PSB: Please fill in details about `correl` here.

As the bedrock of our modeling, we use the publicly available[1] catalog of `Rockstar` subhalos identified at $z = 0$ in the Bolshoi-Planck simulation (Klypin et al. 2011; Behroozi et al. 2013,?; Riebe et al. 2013; Rodríguez-Puebla et al. 2016). As described in §3.1, we will use traditional abundance matching to connect stellar mass M_* with subhalo peak mass $M_{\rm peak}$. To address issues related to subhalo incompleteness (Guo & White 2014; Campbell et al. 2017), we supplement the Bolshoi-Planck subhalo catalog with *all* subhalos that were ever identified by `Rockstar`, including those that no longer appear in the standard catalog as subhalos that survive to $z = 0$. We describe our treatment of these "orphan" subhalos in the Appendix.

For mock galaxies, to compute galaxy clustering we employ the distant observer approximation by treating the simulation $z-$axis as the line-of-sight. We compute $w_{\rm p}$ using the `mock_observables.wp` function in `Halotools`, which is a python implementation of the algorithm in the `Corrfunc` C library (Sinha & Garrison 2017). We compute $\Delta\Sigma$ using the `mock_observables.delta_sigma` function.

All numerical values of $R_{1/2}$ will be quoted in physical kpc, and all values of M_* and $M_{\rm halo}$ in $M_\odot$, assuming $H_0 = 67.8$ km/s $\equiv 100h$ km/s, the best-fit value from Planck Collaboration et al. (2016). To scale stellar masses to "$h = 1$ units" (Croton 2013), our numerically quoted values for M_* should be multiplied by a factor of h^2, while our halo masses and distances should be multiplied by a factor of h.

2.1   Classifying large vs. small galaxies

Because galaxy clustering has well-known dependence upon M_* that is not the subject of this work, we wish to remove this influence and focus purely on the relationship between $R_{1/2}$ and $w_{\rm p}(r_{\rm p})$. To do so, we determine the value $\langle R_{1/2}|M_*\rangle_{\rm median}$ by computing a sliding median of $R_{1/2}$, calculated using a window of width $N_{\rm gal} = 1000$. Each galaxy is categorized as either "large" or "small" according to whether it is above or below the median value appropriate for its stellar mass. We note that this is directly analogous to the common convention for studying the properties of "red" vs. "blue" galaxies, in which the two subsamples are divided by a M_*- or luminosity-dependent green valley cut (e.g., van den Bosch et al. 2008; Zehavi et al. 2011).

Using this technique, for any M_*-threshold sample, the stellar mass function of the "large" and "small" subsamples are identical. We illustrate this for the particular case of $M_* > 10^{10.25} M_\odot$ in left panel of Figure 1, which shows histograms of stellar mass for the two subsamples. The right panel of Figure 1 compares histograms of their size distributions, which partially overlap due to the variation in $\langle R_{1/2}|M_*\rangle_{\rm median}$ across the M_*- range of the
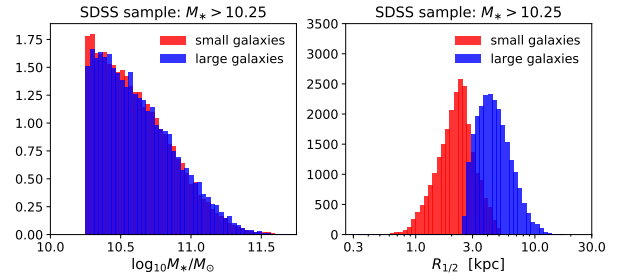
Figure 1. Definition of "small" and "large" galaxies. For a volume-limited SDSS galaxy sample defined by $M_* > 10^{10.25} M_\odot$, we visually demonstrate how we classify galaxies into "small" and "large" subsamples. As described in detail in §2.1, we compute the median value $\langle R_{1/2}|M_*\rangle_{\rm median}$ using a sliding window with a width of 1000 galaxies at each value of M_*. The *left panel* shows a histogram of the stellar masses, confirming that our method yields identical stellar mass functions for the two subsamples. The *right panel* shows histograms of $R_{1/2}$ for the two subsamples, which partially overlap due to the finite range of M_* in the volume-limited sample. AVK comment: left panel should be actual stellar mass function, not just a histogram

threshold sample. AVK comment: left panel should be actual stellar mass function, not just a histogram

3   GALAXY-HALO MODEL

3.1   Abundance Matching

We map M_* onto subhalos using deconvolution abundance matching (SHAM). See section 3.3 of **?** or the appendix of Kravtsov et al. (2014) for an account of the mathematics underlying this technique. To perform the abundance matching, we use the publicly available code[2] developed by Yao-Yuan Mao (Lehmann et al. 2017) that provides a python wrapper around the C kernel developed by Peter Behroozi (**?**). For the subhalo property used in SHAM, we use $M_{\rm peak}$, the largest value of $M_{\rm vir}$ ever attained by the subhalo throughout the entire history of the main progenitor halo. Using the stellar mass function provided in Moustakas et al. (2013), we model M_* as a log-normal distribution with 0.2dex of scatter about the median relation $\langle M_*|M_{\rm peak}\rangle_{\rm median}$ determined by deconvolution abundance matching. As described in the appendix, we include a prescription for supplementing the ordinary `Rockstar` subhalo catalog with disrupted subhalos, so that some of our model galaxies inhabit subhalos that are no longer resolved and do not appear in the standard publicly available catalog.

3.2   Galaxy size models

In §4, we calculate predictions for the $R_{1/2}-$dependence of galaxy clustering for several different kinds of empirical models, described in turn below.

---

[1] `http://www.slac.stanford.edu/~behroozi/BPlanck_Hlists`

[2] `https://bitbucket.org/yymao/abundancematching/overview`
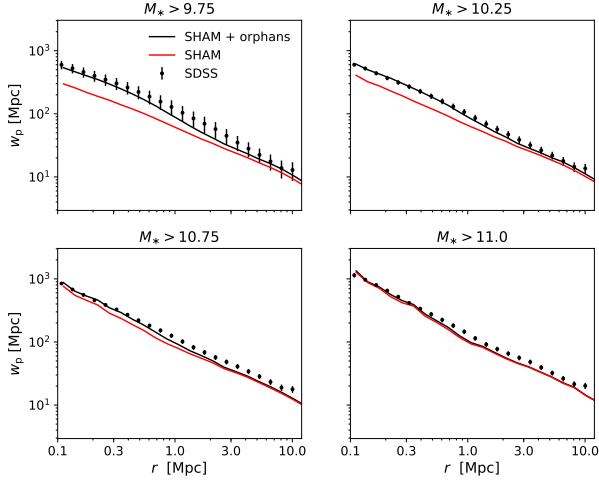
Figure 2. SHAM + orphan clustering predictions. Using the abundance matching methods described in §**??**, we compare the projected clustering of mock vs. SDSS galaxies. Each panel shows the comparison for a volume-limited sample defined by a different M_*-threshold. Black points with error bars show SDSS measurements; solid red (black) curves show the abundance matching prediction including (excluding) the effect of orphan subhalos (see Appendix A). Including orphans mitigates the discrepancy in the clustering predicted by traditional, $M_{\rm peak}-$based SHAM based, though mild tension remains for $M_* \gtrsim 10^{10.75} M_\odot$.

3.2.1 M_*-based model

In the first class of models we explore, we suppose that stellar mass M_* is the statistical regulator of $R_{1/2}$, so that galaxy sizes are drawn from a log-normal distribution with 0.2 dex of scatter centered at $\langle R_{1/2}|M_*\rangle_{\rm median}$. To implement this model, for simplicity we directly tabulate $\langle R_{1/2}|M_*\rangle_{\rm median}$ directly from the data, rather than pursue a parametric form (see, e.g., Zhang & Yang 2017).

3.2.2 $R_{\rm vir}-$based model

AVK comment: The choice of nomenclature of $R_{\rm vir}-$based model and M_*-based model is not clear at all. APH: The big picture being compared/tested here are "pre-infall" vs. "post-infall" processes, respectively. Those don't seem like great names either though.

Motivated by Kravtsov (2013), we explore a model in which $R_{1/2}$ is linearly proportional to halo virial radius:

$$R_{1/2} = 0.01 R_{\rm vir} \tag{1}$$

For the virial radius of halos and subhalos, we use $R_{\rm M_{peak}}$, the value of $R_{\rm vir}$ in physical units of kpc measured at the time of peak subhalo mass, defined by

$$M_{\rm peak} \equiv \frac{4\pi}{3} R_{\rm M_{peak}}^3 \Delta_{\rm vir}(z_{M_{\rm peak}})\rho_{\rm m}(z_{M_{\rm peak}}), \tag{2}$$

where for $\Delta_{\rm vir}(z_{M_{\rm peak}})$ we use the fitting function to the "virial" definition used in Bryan & Norman (1998). For the model we refer to as the "$R_{\rm vir}-$based model", we add uncorrelated log-normal scatter of $\sigma_{\rm R_{1/2}} = 0.2$ dex to generate a Monto Carlo realization of the model population.

3.2.3 M_*+stripping model

As we will show in §4, the chief ingredient needed to recover the observed clustering properties of galaxies is that satellites need to be smaller than centrals of comparable halo mass. Thus it is natural to consider a class of models in which stellar mass is stripped from satellite galaxies after infall.

The basis of this class of models is the fitting function presented in Smith et al. (2016), which was calibrated by studying stellar mass loss in a suite of high-resolution hydrodynamical simulations. In this model, f_* quantifies the fraction of stellar mass lost as a function of $f_{\rm DM}$, the amount of dark matter that has been stripped since infall:

$$f_* = 1 - \exp(-14.2 f_{\rm DM}) \tag{3}$$

For $f_{\rm DM}$ we use the ratio of present-day subhalo mass divided by the peak mass, $M_{\rm vir}/M_{\rm peak}$. If we denote the post-stripping stellar mass as M'_*, then we have $M'_* \equiv f_* M_*$, where M_* is given by the initial application of abundance matching. We then calculate the post-stripping radius by interpolating $\langle R'_{1/2}|M'_*\rangle$ directly from SDSS data.

4 RESULTS

4.1 Size-Mass Scaling Relation

In Figure 3 we show the scaling of galaxy size $R_{1/2}$ with M_*. The black curve enveloped by the gray bands shows the scaling relation for SDSS galaxies, while the blue curve shows the median relation $\langle R_{1/2}|M_*\rangle_{\rm median}$ implied by the $R_{\rm vir}-$based model described in §3.2.2. This figure shows that models in which $R_{1/2} \propto R_{\rm vir}$ can naturally give rise to the characteristic curvature in the $R_{1/2} - M_*$ relation, confirming the results in Kravtsov (2013) in a forward modeling context.

4.2 Size-Dependent Clustering

In Figure 4, we present new measurements of the $R_{1/2}-$dependence of projected galaxy clustering, $w_{\rm p}(r_{\rm p})$. We measure $w_{\rm p}(r_{\rm p})$ separately for large and small subsamples for four different M_* thresholds, $M_* > 10^{9.75} M_\odot$, $M_* > 10^{10.25} M_\odot$, $M_* > 10^{10.75} M_\odot$, and $M_* > 10^{11} M_\odot$. For each threshold, we split the galaxies into "large" and "small" subsamples; as described in §2.1 and illustrated in Figure 1, our size-based selection is defined so that the subsamples have identical stellar mass functions. Red points with jackknife-estimated error bars show SDSS measurements of $w_{\rm p}(r_{\rm p})$ for small galaxy samples, blue points show the same for large galaxies. Solid curves show $w_{\rm p}$ as predicted by the $R_{\rm vir}-$based model described in §3.2.2.

The salient feature of these clustering measurements is that small galaxies cluster more strongly than large galaxies of the same stellar mass. This feature also holds true for galaxies predicted by the $R_{\rm vir}-$based model. This result may be surprising, since $R_{1/2} \propto R_{\rm vir}$, halo mass $R_{\rm vir} \propto M_{\rm halo}^{1/3}$, and clustering strength increases with
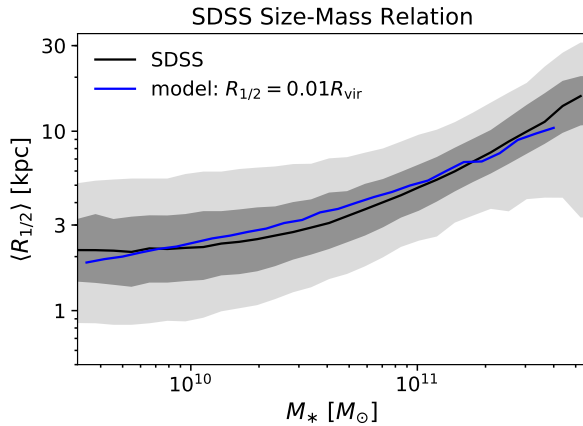
Figure 3. The black curve shows the median size-mass relation of SDSS galaxies as measured in Meert et al. (2015). The two gray bands enveloping the black curve show the 50% and 90% percentile regions. The blue curve shows $R_{\rm vir}-$based model in which $\langle R_{1/2}|R_{\rm vir}\rangle_{\rm median} = 0.01R_{\rm vir}$. This figure confirms that a linear relationship between $R_{\rm vir}$ and $R_{1/2}$, convolved against the nonlinear relationships between $R_{\rm vir}$, $M_{\rm halo}$ and M_*, predicts the characteristic curvature in the relation $\langle R_{1/2}|M_*\rangle_{\rm median}$ over a wide range in mass.

$M_{\rm vir}$. Based on this simple argument, one would expect that large galaxies would be the more strongly clustered. We provide a resolution to this conundrum in §4.3; before doing so, we first examine the clustering test of the $R_{\rm vir}-$based model in more detail.

As shown in Figure 2, the abundance matching prediction for $w_{\rm p}(r_{\rm p})$ exhibits tension with SDSS observations at the $10-20\%$ level, particularly for $M_* \gtrsim 10^{10.75}M_\odot$. This tension is inherited by our $R_{\rm vir}-$based model for size, which is the subject of this work, and so we wish to compare our size models to data in such a way that minimizes the role played by the underlying stellar-to-halo-mass relation. We accomplish this using *the $R_{1/2}$ clustering ratios,* described below.

For each volume-limited M_*-threshold sample, we additionally measure $w_{\rm p}(r_{\rm p})$ *without* splitting on size, giving us measurements $w_{\rm p}^{\rm all}, w_{\rm p}^{\rm large}$, and $w_{\rm p}^{\rm small}$ for each threshold sample. This allows us to compute the ratio $(w_{\rm p}^{\rm large} - w_{\rm p}^{\rm small})/w_{\rm p}^{\rm all}$, which we refer to as *the $R_{1/2}$ clustering ratio,* denoted as $\delta_{R_{1/2}}(w_{\rm p})$. For example, a clustering ratio of -0.5 corresponds to small galaxies being 50% more strongly clustered than large galaxies of the same stellar mass. These ratios are the measurements appearing on the y-axis in each panel of Figure 5.

The points and curves in Figure 5 are all negative: small galaxies cluster more strongly relative to large. This presentation of the measurement makes plain that the underlying signal of $R_{1/2}-$dependent clustering is strongest for samples with smaller stellar mass; as M_* increases, the signal weakens and nearly vanishes for $M_* \gtrsim 10^{11}M_\odot$. Strikingly, the $R_{\rm vir}-$based model exhibits this M_*-dependent behavior, as well as the scale-dependence of the observed clustering signal at each M_*. As described in §4.3 below, we attribute the success of this prediction to the relative sizes of central vs. satellite galaxies.

4.3   Central vs. Satellite Sizes

The clustering measurements shown in the previous section present a puzzle. The modeling assumption $R_{1/2} \propto R_{\rm vir}$ predicts size-dependent galaxy clustering that is in quite good agreement with observations, particularly considering the simplicity of the model. Yet, small galaxies cluster more strongly relative to large, which seems at odds with naive expectations based on halo mass, since $R_{\rm vir} \propto M_{\rm vir}$. A straightforward resolution to this puzzle is shown in Figure 6, which compares the $R_{1/2}$ distributions of central, satellite, and splashback galaxies with the same halo mass $M_{\rm halo} \equiv M_{\rm peak} \approx 10^{12}M_\odot$. A "splashback central" is defined as a present-day central that used to be a satellite, i.e., its main progenitor halo passed inside the virial radius of a larger halo at some point in its past history. On the other hand, we define a "true central" as a galaxy that has never been a satellite.

In the $R_{\rm vir}-$based model, satellite and splashback galaxies are smaller than centrals of the same halo mass due to the physical size of their halo being smaller at earlier times $z_{M_{\rm peak}}$. There are two distinct reasons why this feature results in small galaxies being more strongly clustered relative to larger galaxies of the same mass. First and foremost, satellite galaxies statistically occupy higher mass host halos that are more strongly clustered. So in models where satellites are smaller than centrals, at fixed M_* the "small" subsample will naturally have a higher satellite fraction, boosting the clustering of small galaxies relative to large of the same stellar mass. Second, at fixed mass, halos of L_* galaxies that form earlier are more strongly clustered, a phenomenon commonly known as *halo assembly bias.* Splashback halos are typically earlier-forming than true centrals, and so models where splashback halos host smaller-than-average galaxies will naturally predict smaller galaxies being the more strongly clustered. We highlight this point in section 4.4 below by testing predictions for M_*-based models that do not possess central vs. satellite size differences.

4.4   Failure of M_*-based Models

The green curves in Figure 7 test the clustering predictions of the M_*-based model described in §3.2.1. In this model, $R_{1/2}$ is a log-normal distribution centered at $\langle R_{1/2}|M_*\rangle_{\rm median}$. By construction, this model predicts no dependence at all of galaxy clustering upon $R_{1/2}$ at fixed M_*, even though the observed one-point distributions are recovered exactly.

The reason for the null signal of the purely M_*-based model is simply that the scatter of $R_{1/2}$ about $\langle R_{1/2}|M_*\rangle_{\rm median}$ is uncorrelated with any other property. This is in contrast to the $R_{\rm vir}-$based model, which correlates the scatter with host halo mass and as well as subhalo assembly. The purely M_*-based model gives us a useful starting point to test simple alternative hypotheses for how the scatter may be correlated with large-scale structure. This is the motivation behind the M_*+stripping model described in §3.2.3. In this alternative model, satellites have smaller sizes than centrals of the same M_* due to post-infall stripping, where the amount of stripping has been calibrated to the satellite

Figure 4. $R_{1/2}-$**dependence of galaxy clustering.** Red and blue points with error bars show our SDSS measurements of the clustering of small and large galaxies, respectively. For each volume-limited sample of M_*-complete galaxies, the small and large subsamples have identical stellar mass functions, as shown in Figure 1. Small galaxies cluster much more strongly relative to large galaxies of the same stellar mass. Solid curves show the clustering predictions of the $R_{\rm vir}-$based model described in §3.2.2. The $R_{\rm vir}-$based model inherits the shortcoming of ordinary abundance matching at $M_* \gtrsim 10^{10.75} M_\odot$, although the model faithfully captures the *relative* clustering of small vs. large galaxies, as shown in Figure 5.

mass loss seen in high-resolution hydrodynamical simulations.

The treatment of stellar mass loss in this model naturally results in an enhancement of size differences between satellites and centrals. As shown by the purple curves in Figure 7, these central/satellite differences impact $R_{1/2}-$dependent clustering in the expected manner: small galaxies cluster more strongly relative to large galaxies of the same stellar mass. Since the satellite fraction increases with decreasing stellar mass, the clustering ratios are more negative for samples with smaller M_* thresholds. This confirms the notion that central vs. satellite differences drive the $R_{1/2}-$dependence of clustering, although the magnitude of the effect is evidently is not strong enough to produce clustering predictions that are consistent with SDSS. It is difficult to strip enough mass from satellites so that the $R_{1/2}-$dependent clustering is in agreement with observations. Taken together with the results for the $R_{\rm vir}-$based model presented in §4.2 and shown in Figure 5, this supports the conclusion that the relative difference between central and satellite size is at least partially in place at the time of satellite infall. We discuss the physical implications of this result in §5.2.

4.5 The Role of Morphology and Color

In this section we take a preliminary look at how galaxy clustering exhibits simultaneous dependence upon broadband color, morphology, and size. Beginning from the $M_* > 10^{10.25} M_\odot$ sample used above, we first divide the sample into "red" and "blue" subsamples according to a $g - r = 0.65$ cut, the rough location of the trough of the green valley for this stellar mass. For each color-selected subsample, we separately measure $\langle R_{1/2}|M_*; {\rm red}\rangle_{\rm median}$ and $\langle R_{1/2}|M_*; {\rm blue}\rangle_{\rm median}$, and use these median size values to split each subsample into two.

In the top left panel of Figure 8 we show the clustering of large vs. small red galaxies; in the top right panel we show the same for blue galaxies. For each color-selected sample, large galaxies are slightly more strongly clustered relative to small galaxies. Comparing the top panels of Figure 8 to the the upper right panel of Figure 4 shows the dramatic impact of galaxy color selection upon $R_{1/2}-$dependent clustering. For M_*-complete samples, small galaxies cluster much stronger than large galaxies; for color-selected samples, the reverse is true, and the magnitude of the effect weakens considerably.

The bottom panels of Figure 8 show results that are analogous to the top panels, but instead using of $g - r$ color, we first divide the $M_* > 10^{10.25} M_\odot$ sample into "bulge-dominated" and "disk-dominated" sequences according to B/T, the fraction of $r-$band flux coming from the bulge component of the 2d light profile measurements
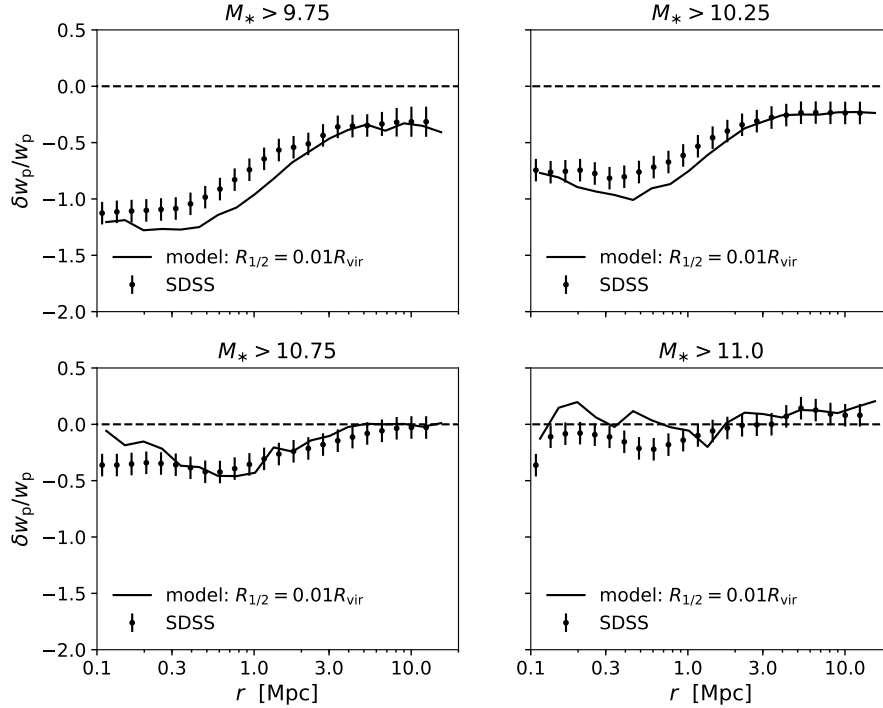
Figure 5. $R_{1/2}-$**dependence of galaxy clustering: clustering ratios.** Closely related to Figure 4, the y-axes show *clustering strength ratios,* defined as $(w_{\rm p}^{\rm large} - w_{\rm p}^{\rm small})/w_{\rm p}^{\rm all}$. Thus a y-axis value of -0.5 corresponds to small galaxies being 50% more strongly clustered than large galaxies of the same stellar mass. Solid curves show the clustering ratio predictions of the $R_{\rm vir}-$based model described in §3.2.2. Normalizing the measurements and predictions by $w_{\rm p}^{\rm all}$ scales away the shortcoming of ordinary abundance matching at high stellar mass (see Figure 2), highlighting the successful prediction of the $R_{\rm vir}-$based model for the $R_{1/2}-$dependence of galaxy clustering.
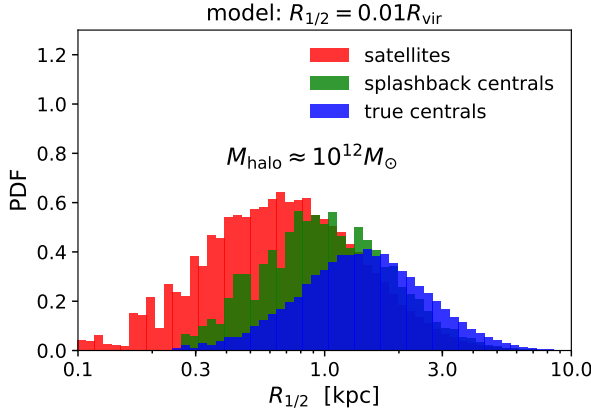
Figure 6. Relative sizes of centrals and satellites. In a narrow bin of halo mass $M_{\rm halo} \equiv M_{\rm peak} \approx 10^{12} M_\odot$, we show the distribution of model galaxy sizes for different subpopulations galaxies, as predicted by the $R_{\rm vir}-$based model. The red histogram shows the sizes of satellites; the blue histogram shows host halos that have never passed inside the virial radius of a larger halo ("true centrals"); the green histogram host halos that were subhalos inside a larger at some point in their past history ("splashback centrals"). In the $R_{\rm vir}-$based model, galaxy size is set by the *physical* size of the virial radius at the time the halo attains its peak mass, naturally resulting in smaller sizes for satellites and splashback centrals relative to true centrals of the same $M_{\rm halo}$.

in the Meert et al. (2015) catalog. We define a galaxy as bulge-dominated if $B/T < 0.25$, and disk-dominated if $B/T > 0.75$. We separately split the disk-dominated and bulge-dominated subsample into two based on the median size $\langle R_{1/2}|M_*; B/T\rangle_{\rm median}$ appropriate for each subsample. The bottom left panel of Figure 8 compares the clustering of small vs. large disk-dominated galaxies; the bottom right panel shows the same comparison for bulge-dominated galaxies.

The contrast between the two bottom panels is stark. The clustering of large vs. small bulge-dominated galaxies has almost no dependence upon $R_{1/2}$. On the other hand, for disk-dominated galaxies, small galaxies are much more strongly clustered, and the effect is strong. In fact, comparing the bottom left panel of Figure 8 to the top right panel of Figure 4, we see that $R_{1/2}-$dependent clustering is quite similar between disk-dominated galaxies and M_*-threshold samples. We discuss a halo model interpretation of these trends in §5.3.

5   DISCUSSION

5.1   Relation to Previous Work

Numerous previous analyses have employed the Yang et al. (2005) group catalog to "directly" study the relationship between galaxy mass, size, and environment. For example, by treating observed groups as genuine
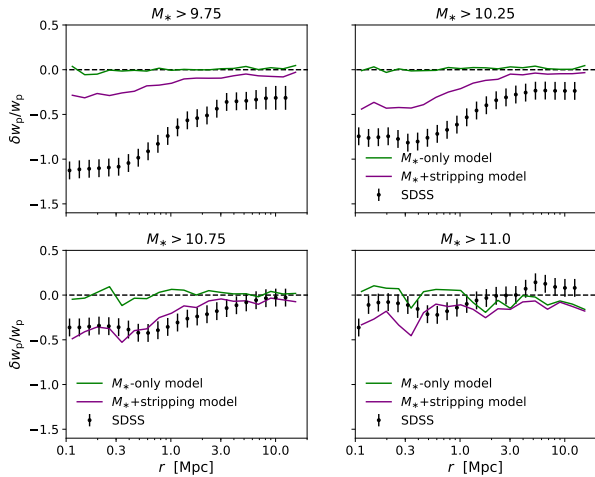
Figure 7. Impact of tidal stripping. In all panels, the axes and points with error bars are the same as in Figure 5. The solid green curves shows the prediction of the M_*-based model, which is in gross tension with the data due to satellites having the same size as centrals of the same mass. The solid purple curves show results for a model in which satellites lose mass after infall in a manner similar to what is seen in high-resolution hydrodynamical simulations, as described in §3.2.3. This produces satellites that are smaller than centrals, but the effect is too mild to correctly capture the observed clustering. Evidently, satellite-specific mass stripping plays a sub-dominant role in setting the relative size of centrals and satellites.

dark matter halos, several analyses have found that the $M_* - R_{1/2}$ relation of early-type galaxies exhibits weak, if any, environmental dependence (**?**Huertas-Company et al. 2013; Shankar et al. 2014). Meanwhile, **?** additionally find that for late-type galaxies, there is a strong correlation between disk size and central/satellite designation. Qualitatively, our clustering measurements shown in Figure 8 are in good agreement with these findings: once galaxies have been morphology-selected, only the clustering of disk shows significant dependence upon $R_{1/2}$. Our clustering results are also consistent with the disk model presented in Dutton et al. (2007, 2011), which predicts that at fixed stellar mass, there is weak, positive correlation between disk size and halo mass.

As discussed in Spindler & Wake (2017), it is entirely possible that size differences between centrals and satellites of the same mass can be accounted for by mutual covariance with an additional variable such as star formation rate or morphological type (see also Lilly & Carollo 2016, for an explicit demonstration of this scenario). In particular, suppose that centrals and satellites of the same mass have different early-type fractions as indicated in Weinmann et al. (2006), and that early- and late-type galaxies exhibit universal, but distinct, $M_* - R_{1/2}$ relations. In such a case, centrals would be larger than satellites of the same mass, even though *early-type centrals* would have the same sizes as *early-type satellites*. This scenario could also explain the sign-change and weakening of $R_{1/2}-$dependent clustering measurements in the present work.

The $R_{1/2}-$dependence of galaxy lensing has recently

been measured using CFHTLenS observations (Heymans et al. 2012; Erben et al. 2013). For both red- and blue-sequence galaxies, it was found in Charlton et al. (2017) that the lensing signal, $\Delta\Sigma$, of large galaxies is slightly stronger relative to smaller galaxies. This result is in good agreement with the clustering measurements in the top panels of Figure 8, which show the same trend. To date, the $R_{1/2}-$dependence of $\Delta\Sigma$ not yet been measured for M_*-complete samples. We show predictions of the $R_{\rm vir}-$based model for future measurements of this signal in Figure 9. The halo model explanation for this trend is the same as for $w_{\rm p}$: satellites occupy higher mass host halos relative to centrals of the same stellar mass, boosting $\Delta\Sigma$ for small relative to large galaxy samples.

As discussed in §5.3, we are currently pursuing follow-up work in which we jointly model morphological type together with galaxy size. We consider forward modeling methods to be a requisite for meaningful progress on this issue, not only to properly handle the multi-dimensional nature of the problem, but also to rigorously treat systematic errors that plague inference based on galaxy group catalogs (see Campbell et al. 2015, for a thorough discussion).

Our methodology is closely aligned with Somerville et al. (2017), who studied the empirical modeling features that are necessary to recover the tight scatter in the observed $\langle R_{1/2}|M_*\rangle$ relation. By building models where $R_{1/2}$ is set by halo spin $\lambda_{\rm halo}$, the authors in Somerville et al. (2017) found that the level of intrinsic scatter about $\langle\lambda_{\rm halo}|M_{\rm halo}\rangle$ in dark matter halos is at least as large as the scatter about $\langle R_{1/2}|M_*\rangle$ seen in observed galaxies. In our approach, the level of scatter is simply a modeling parameter whose fiducial value was motivated by Somerville et al. (2017). In ongoing follow-up work discussed in §5.3, we will systematically test the large-scale structure implications of the assumption that $\lambda_{\rm halo} \propto R_{1/2}^{\rm disk}$ using forward-modeling methods analogous to those advocated for in Somerville et al. (2017) and employed here.

5.2 Implications for Satellite Evolution

Physical explanations for differences between the sizes of centrals and satellites fall into two categories, depending on whether or not the processes involved are post-infall and satellite-specific. The success of the $R_{\rm vir}-$based model, together with the failure of the M_*+stripping model, suggests that the relative size of central and satellite galaxies is largely driven by the pre-infall evolution of the satellite. The clustering ratios in Figure 7 show that the impact of physically motivated levels of satellite mass loss likely cannot explain the $R_{1/2}-$dependent clustering seen in SDSS. As an additional test, we reach the same conclusion when implementing an even more extreme M_*-stripping model in which stellar mass loss is linearly proportional to halo mass loss (Watson et al. 2012, Model 1).

While these clustering measurements give supporting evidence that satellite mass stripping plays a sub-dominant role in shaping galaxy size, we point out that our treatment of satellite mass loss is only approximate: we have modified the size of satellites without self-
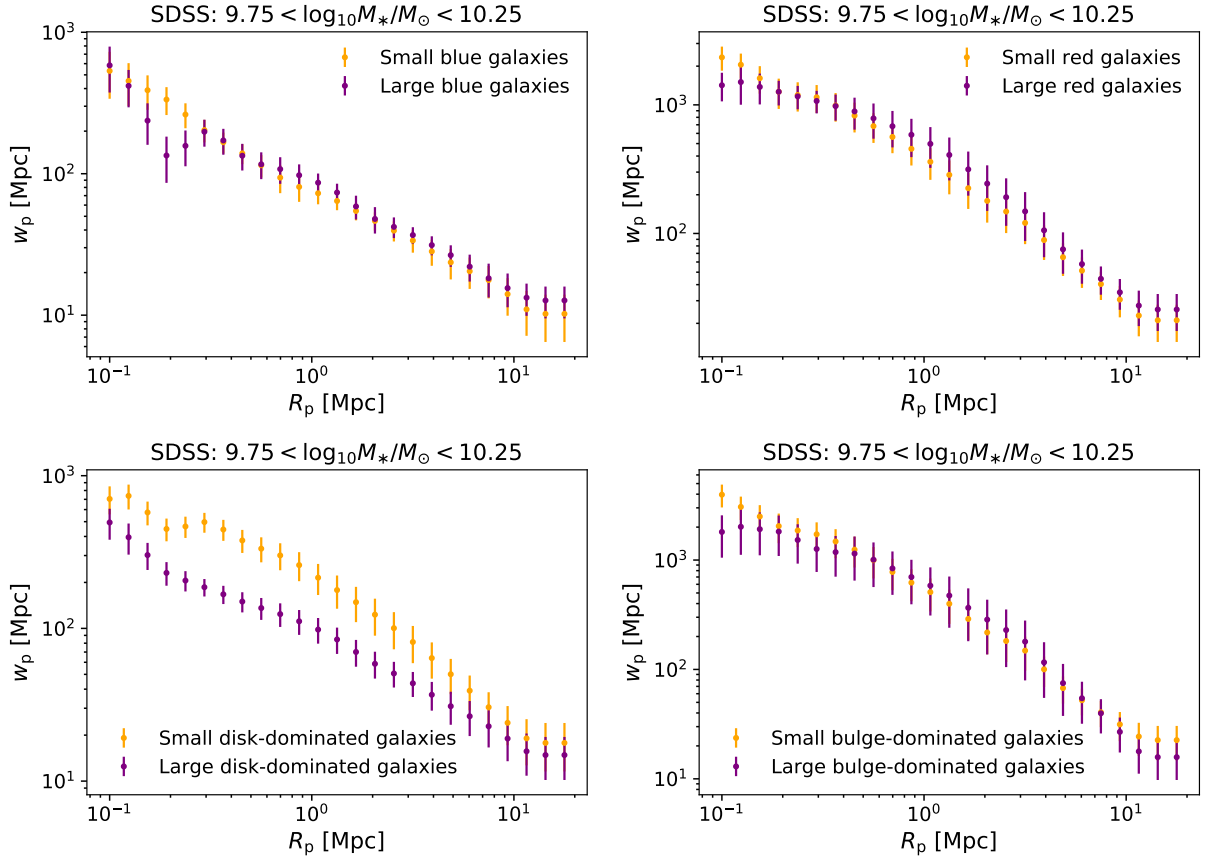
Figure 8. **Distinct $R_{1/2}$−dependence of clustering for color- or morphology-selected galaxy samples.** All panels show the clustering of galaxies in the same bin of stellar mass: $10^{9.75} < M_*/M_\odot < 10^{10.25}$. In the *upper left* panel, we first select blue galaxies based on $g - r < 0.6$, and subsequently compute $\langle R_{1/2}|M_*\rangle_{\text{median}}$ *of the color-selected sample*. This results in identical stellar mass function for large and small blue galaxy samples, in analogy to the left panel of Figure 1. The *upper right panel* shows results of the same procedure for "red" galaxies, defined by $g - r > 0.6$. The top two panels show that for color-selected samples, the magnitude of the $R_{1/2}$−dependence of clustering is dramatically reduced, and changes sign, relative to M_*−complete samples. Using the Meert et al. (2015) measurements of the disk/bulge decomposition of the 2d r−band luminosity profile L_{r}, the bottom left panel shows analogous results for disk-dominated galaxies defined by $L_{\text{r}}^{\text{bulge}} < L_{\text{r}}^{\text{tot}}/4$; the bottom right panel shows results for bulge-dominated galaxies defined by $L_{\text{r}}^{\text{disk}} < L_{\text{r}}^{\text{tot}}/4$.

consistently modifying total stellar mass. In the present work, we have opted not to develop and fine-tune a complex, semi-analytic model; instead, we have made a simple "first-order" estimate for the level at which satellite-specific stripping impacts the two-point function. Our findings support the idea that the small sizes of satellites arise primarily from the satellite's distinctive evolution within a high density region and a steep tidal field *prior* to becoming a present-day satellite.

5.3 Future Directions for Empirical Modeling of Galaxy Size

#### 5.3.1 *Jointly modeling $\langle M_*|M_{\text{halo}}\rangle$*

The scope of this work is to serve as a pilot study in which we identify the chief ingredients that can influence the $R_{1/2}$−dependence of galaxy clustering. While our results suggest that satellite evolution *prior* to infall drives the correct relative size of centrals vs. satellites, our implementation cannot be correct in quantitative detail: the

clustering ratios of this model shown in Figure 5 only roughly agree with SDSS. For present purposes, we have chosen to focus on exploring bracketing cases of qualitative ingredients, rather than fine-tuning the model.

Since the clustering signal is strongly influenced by differences between centrals and satellites, then the satellite fraction $F_{\text{sat}}(M_*)$ plays an important role in $R_{1/2}$−dependent clustering. For *fixed* scaling relations $\langle R_{1/2}^{\text{cens}}|M_{\text{halo}}\rangle_{\text{median}}$ and $\langle R_{1/2}^{\text{sats}}|M_{\text{halo}}\rangle_{\text{median}}$, models with different satellite fractions will exhibit different $R_{1/2}$−clustering ratios because $F_{\text{sat}}(M_*)$ controls the relative weighting of the two populations. An explicit demonstration of this point appears in the Appendix, which repeats the analyses in the main body of the text, but for a subhalo catalog with no orphan correction.

On the one hand, this degeneracy with the satellite fraction is unfortunate, because it means galaxy size does not leave a pure and unique signature on $R_{1/2}$−clustering ratios. However, this can also be viewed as an opportunity to extract tighter constraints on $F_{\text{sat}}(M_*)$, which are sorely needed to discriminate between competing models
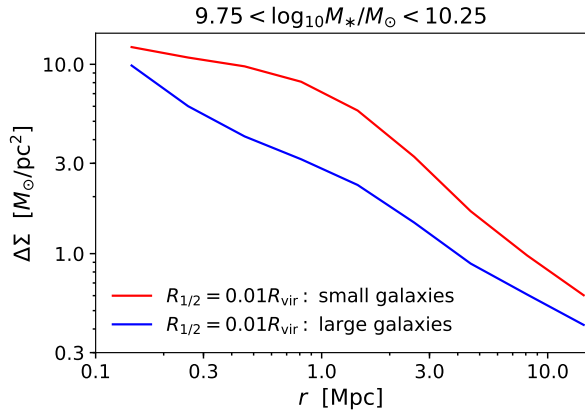
Figure 9. Prediction for $R_{1/2}-$dependence of galaxy lensing. Using the $R_{\rm vir}-$based model described in §3.2.2, we make predictions for as-yet-unseen measurements of the $R_{1/2}-$dependence of galaxy lensing of M_*-complete samples. To date, the $R_{1/2}-$dependence of $\Delta\Sigma$ has only been measured for color-selected samples (Charlton et al. 2017), in which the reverse is true: for both blue and red samples, $\Delta\Sigma$ of small galaxies is (weakly) suppressed relative to large galaxies. As a generic consequence of satellites being smaller than centrals of the same mass, we predict that the analogous measurement for M_*-complete samples will show $\Delta\Sigma$ of small galaxies to be much stronger relative to large galaxies of the same stellar mass.

(Watson & Conroy 2013). Traditional galaxy clustering is already being used to validate and/or fit models of the stellar-to-halo-mass relation (e.g., Leauthaud et al. 2012; Moster et al. 2013; Behroozi et al. 2013; Lehmann et al. 2017). Based on our results, we advocate that empirical models for $\langle M_*|M_{\rm halo}\rangle$ be supplemented with additional model ingredients for $\langle R_{1/2}|M_*\rangle_{\rm median}$, and that the parameters of the composite model be *jointly* constrained by measurements of the stellar mass function, M_*-dependent clustering, and $R_{1/2}-$dependent clustering.

5.3.2 *Jointly modeling morphology*

A proper interpretation of the results clustering results shown in Figure 8 requires forward-modeling the joint dependence of the galaxy-halo connection upon $M_*, R_{1/2}, g-r$, and B/T, which is beyond the scope of this work. Nonetheless, the broad features of these $w_{\rm p}(r_{\rm p})$ measurements give insight into the characteristics that any model for these properties should exhibit.

First, the bottom right panel of Figure 8 indicates that the size of the bulge has only a tenuous connection to the properties of its parent dark matter halo.[3] Second, for disk-dominated galaxies, the observed $R_{1/2}-$dependent clustering suggests that at fixed M_*, the size of the disk is strongly correlated with whether or not the galaxy

---

[3] Strictly speaking, the size of the bulge could in principle be closely connected to any number of dark matter halo properties, so long as those halo properties do not significantly impact two-point clustering.

is a satellite. These findings imply that bulge prominence is intimately connected with the quenching of star-formation, but that there is wide variation in star-formation histories amongst disk-dominated galaxies, in agreement with numerous previous findings (see, e.g., Woo et al. 2015; Font et al. 2017).

Previous studies of large-scale structure measured by SDSS indicate that morphological type has a significant influence on galaxy clustering (Skibba et al. 2009). Because morphology is covariant with star-formation rate (SFR) and $R_{1/2}$, accounting for this dependency will require simultaneously modeling each of these variables. Historically, such a task has been beyond the purview of empirical models. However, recent progress has substantially improved the richness of the predictions that galaxy-halo models are capable of (Becker 2015; Cohn 2017; Moster et al. 2017). This new class of models has accomplished this by empirically modeling star-formation rate, and then integrating SFR along each halo's history to differentially build up stellar mass. We are currently extending this new class of models to differentially build up the disk and bulge across cosmic time using computationally efficient empirical modeling methods. In addition to treating the multi-dimensional nature of the problem, this will allow us to use data at higher redshift in a self-consistent manner.

6 CONCLUSIONS

We have presented new measurements of the dependence of galaxy clustering upon galaxy size, and used `Halotools` to identify the basic ingredients that influence the signal. We conclude with a brief summary of our primary findings:

(i) Small galaxies cluster more strongly than large galaxies of the same stellar mass. Differences between the clustering of small and large galaxies increase on small scales $R \lesssim 1$Mpc, and decrease with stellar mass.

(ii) The most important ingredient influencing this signal is the relative size of central and satellite galaxies. The magnitude, scale-dependence, and M_*-dependence of $R_{1/2}-$dependent clustering provides strong evidence that satellite galaxies are smaller than central galaxies of the same halo mass.

(iii) A simple empirical model in which $R_{1/2}$ is set by halo $R_{\rm vir}$ at the time of peak halo mass exhibits a clustering signal that is strikingly similar to that seen in SDSS.

(iv) Models in which $R_{1/2}$ is regulated by M_*, rather than $M_{\rm halo}$, are grossly discrepant with the observed clustering signal, even when accounting for satellite mass stripping.

(v) Taken together, our findings indicate that satellite-specific processes play a sub-dominant role in setting the relative size of centrals and satellites, which instead appears to be largely predetermined at the time of satellite infall.

Our results can be treated as a boundary condition for more complex and fine-grained models of galaxy size, such as semi-analytic models and hydrodynamical simulations. We view the present work as a pilot study that

motivates a Bayesian inference program to tightly constrain the galaxy size-halo connection with forward modeling techniques, in direct analogy to the literature on the stellar-to-halo-mass relation. Our publicly available python code provides a simple means for cosmological surveys to generate synthetic galaxy populations with realistic sizes across the cosmic web.

REFERENCES

Ahn C. P., Alexandroff R., Allende Prieto C., Anders F., Anderson S. F., Anderton T., Andrews B. H., Aubourg É., Bailey S., Bastien F. A., et al. 2014, ApJS , 211, 17

Astropy Collaboration Robitaille T. P., Tollerud E. J., Greenfield P., Droettboom M., Bray E., Aldcroft T., et al., 2013, AAP, 558, A33

Becker M. R., 2015, ArXiv:1507.03605

Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ , 770, 57

Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ , 762, 109

Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013, ApJ , 763, 18

Behroozi P. S., Zhu G., Ferguson H. C., Hearin A. P., Lotz J., Silk J., Kassin S., Lu Y., Croton D., Somerville R. S., Watson D. F., 2015, MNRAS , 450, 1546

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MN-RAS , 436, 697

Bernardi M., Meert A., Vikram V., Huertas-Company M., Mei S., Shankar F., Sheth R. K., 2014, MNRAS , 443, 874

Bottrell C., Torrey P., Simard L., Ellison S. L., 2017, MNRAS , 467, 2879

Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS , 351, 1151

Bryan G. L., Norman M. L., 1998, ApJ , 495, 80

Campbell D., van den Bosch F. C., Hearin A., Padmanabhan N., Berlind A., Mo H. J., Tinker J., Yang X., 2015, MNRAS , 452, 444

Campbell D., van den Bosch F. C., Padmanabhan N., Mao Y.-Y., Zentner A. R., Lange J. U., Jiang F., Villarreal A., 2017, ArXiv:1705.06347

Charlton P. J. L., Hudson M. J., Balogh M. L., Khatri S., 2017, MNRAS , 472, 2367

Cohn J. D., 2017, MNRAS , 466, 2718

Croton D. J., 2013, PASA, 30, e052

Dutton A. A., van den Bosch F. C., Dekel A., Courteau S., 2007, ApJ , 654, 27

Dutton A. A., van den Bosch F. C., Faber S. M., Simard L., Kassin S. A., Koo D. C., Bundy K., Huang J., Weiner B. J., Cooper M. C., Newman J. A., Mozena M., Koekemoer A. M., 2011, MNRAS , 410, 1660

Erben T., Hildebrandt H., Miller L., et al., 2013, MN-RAS , 433, 2545

Font A. S., McCarthy I. G., Le Brun A. M. C., Crain R. A., Kelvin L. S., 2017, ArXiv:1710.00415

Guo Q., White S., 2014, MNRAS , 437, 3228

Guo Y., McIntosh D. H., Mo H. J., Katz N., Van Den Bosch F. C., Weinberg M., Weinmann S. M., Pasquali A., Yang X., 2009, MNRAS , 398, 1129

Heymans C., Van Waerbeke L., Miller L., et al., 2012, MNRAS , 427, 146

Hopkins P. F., Bundy K., Croton D., Hernquist L., Keres D., Khochfar S., Stewart K., Wetzel A., Younger J. D., 2010, ApJ , 715, 202

Huang K.-H., Fall S. M., Ferguson H. C., van der Wel A., Grogin N., Koekemoer A., Lee S.-K., Pérez-González P. G., Wuyts S., 2017, ApJ , 838, 6

Huang S., Ho L. C., Peng C. Y., Li Z.-Y., Barth A. J., 2013, ApJ , 766, 47

Huertas-Company M., Mei S., Shankar F., Delaye L., Raichoor A., Covone G., Finoguenov A., Kneib J. P., Le F. O., Povic M., 2013, MNRAS , 428, 1715

Huertas-Company M., Shankar F., Mei S., Bernardi M., Aguerri J. A. L., Meert A., Vikram V., 2013, ApJ , 779, 29

Jiang F., van den Bosch F. C., 2014, ArXiv e-prints

Jones E., Oliphant T., Peterson P., et al., 2001-2016, http://www.scipy.org

Kauffmann G., Heckman T. M., White S. D. M., et al., 2003, MNRAS , 341, 33

Kawamata R., Ishigaki M., Shimasaku K., Oguri M., Ouchi M., 2015, ApJ , 804, 103

Khochfar S., Silk J., 2006, ApJL , 648, L21

Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ , 740, 102

Kravtsov A., Vikhlinin A., Meshscheryakov A., 2014, ArXiv:1401.7329

Kravtsov A. V., 2013, ApJL , 764, L31

Lange R., Driver S. P., Robotham A. S. G., et al., 2015, MNRAS , 447, 2603

Leauthaud A., Tinker J., Bundy K., et al., 2012, ApJ , 744, 159

Lehmann B. V., Mao Y.-Y., Becker M. R., Skillman S. W., Wechsler R. H., 2017, ApJ , 834, 37

Lilly S. J., Carollo C. M., 2016, ApJ , 833, 1

LSST Science Collaboration Abell P. A., Allison J., Anderson S. F., Andrew J. R., Angel J. R. P., Armus L., Arnett D., Asztalos S. J., Axelrod T. S., et al. 2009, ArXiv e-prints

Meert A., Vikram V., Bernardi M., 2013, MNRAS , 433,

1344

Meert A., Vikram V., Bernardi M., 2015, MNRAS , 446, 3943

Moster B. P., Naab T., White S. D. M., 2013, MNRAS , 428, 3121

Moster B. P., Naab T., White S. D. M., 2017, ArXiv e-prints

Moustakas J., Coil A. L., Aird J., Blanton M. R., Cool R. J., Eisenstein D. J., Mendez A. J., Wong K. C., Zhu G., Arnouts S., 2013, ApJ , 767, 50

Planck Collaboration Ade P. A. R., Aghanim N., Arnaud M., Ashdown M., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., et al. 2016, AAP, 594, A13

Riebe K., Partl A. M., Enke H., Forero-Romero J., Gottlöber S., Klypin A., Lemson G., Prada F., Primack J. R., Steinmetz M., Turchaninov V., 2013, Astronomische Nachrichten, 334, 691

Robertson B. E., Banerji M., Cooper M. C., Members of the LSST Galaxies Science Collaboration 2017, ArXiv e-prints

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS , 462, 893

Shankar F., Mei S., Huertas-Company M., Moreno J., Fontanot F., Monaco P., Bernardi M., Cattaneo A., Sheth R., Licitra R., Delaye L., Raichoor A., 2014, MNRAS , 439, 3189

Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, MNRAS , 343, 978

Shibuya T., Ouchi M., Harikane Y., 2015, ApJS , 219, 15

Sinha M., Garrison L., , 2017, Corrfunc: Blazing fast correlation functions on the CPU, Astrophysics Source Code Library

Skibba R. A., Bamford S. P., Nichol R. C., Lintott C. J., Andreescu D., Edmondson E. M., Murray P., Raddick M. J., Schawinski K., Slosar A., Szalay A. S., Thomas D., Vandenberg J., 2009, MNRAS , 399, 966

Smith R., Choi H., Lee J., Rhee J., Sanchez-Janssen R., Yi S. K., 2016, ApJ , 833, 109

Somerville R. S., Behroozi P., Pandya V., et al., 2017, ArXiv e-prints

Spindler A., Wake D., 2017, MNRAS , 468, 333

Trujillo I., Rudnick G., Rix H.-W., Labbé I., Franx M., Daddi E., van Dokkum P. G., Förster Schreiber N. M., Kuijken K., Moorwood A., Röttgering H., van der Wel A., van der Werf P., van Starkenburg L., 2004, ApJ , 604, 521

van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, MNRAS , 387, 79

Van Der Walt S., Colbert S. C., Varoquaux G., 2011, ArXiv:1102.1523

van der Wel A., Franx M., van Dokkum P. G., et al., 2014, ApJ , 788, 28

Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, MNRAS , 409, 1379

Watson D. F., Berlind A. A., Zentner A. R., 2012, ApJ , 754, 90

Watson D. F., Conroy C., 2013, ApJ , 772, 139

Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS , 366, 2

Woo J., Dekel A., Faber S. M., Koo D. C., 2015, MNRAS , 448, 237

Yang X., Mo H. J., Jing Y. P., van den Bosch F. C., 2005, MNRAS , 358, 217

Zehavi I., Zheng Z., Weinberg D. H., et al., 2011, ApJ , 736, 59

Zhang Y., Yang X., 2017, ArXiv e-prints

APPENDIX: TREATMENT OF DISRUPTED SUBHALOS

We use an extension of `Consistent Trees` that models the evolution of subhalos after disruption. The phase space evolution of disrupted subhalos is approximated by following a point mass evolving in the host halo potential according to the orbital parameters of the subhalo at the time of disruption; the evolution of subhalo mass and circular velocity is approximated using the semi-analytic model presented in Jiang & van den Bosch (2014). We then use the `orphans` program in `UniverseMachine` to walk through all the Bolshoi-Planck `hlist` files, yielding the main progenitor information of every subhalo that was ever identified by `Consistent Trees`.

Since it is likely that some portion of these disrupted subhalos should be populated with model galaxies (Guo & White 2014; Campbell et al. 2017), in our initial application of deconvolution abundance matching we derive the $M_* - M_{\rm peak}$ relation using *all* subhalos, including those that may be disrupted. We then apply a selection function to the disrupted subhalos, so that a fraction of these objects will host galaxies in our mock universe. We refer to this as the *orphan selection function*, $\mathcal{F}_{\rm orphan}$, which we consider to be an integral component of our application of abundance matching.

Since a rigorous calibration of $\mathcal{F}_{\rm orphan}$ is beyond the scope of the present work, we instead opt for a simple parameterization that yields reasonably accurate recovery of the galaxy clustering signal observed in SDSS. We model $\mathcal{F}_{\rm orphan} = \mathcal{F}_{\rm orphan}(M_{\rm peak}, M_{\rm host})$, where $M_{\rm peak}$ is the peak mass of the disrupted subhalo, and $M_{\rm host}$ is the present-day virial mass of its $z = 0$ host halo. For the $M_{\rm peak}-$dependence, we select 50% of disrupted subhalos with $M_{\rm peak} = 10^{11} M_\odot$, 0% of subhalos with $M_{\rm peak} = 10^{13} M_\odot$, linearly interpolating in log $M_{\rm peak}$ for intermediate values of $M_{\rm peak}$. At each $M_{\rm peak}$, the selection of disrupted halos is not random; instead, we preferentially select the subhalos with larger $M_{\rm host}$, which we intend to offset the increased difficulty of subhalo-finding algorithms to identify subhalos with especially small values of $\mu \equiv M_{\rm peak}/M_{\rm host}$.

The impact of our orphan treatment on M_*-threshold clustering is shown in Figure 2 in the main body of the text. We show the impact on $R_{1/2}-$dependent clustering in Figure 10. The character of the effect of orphan subhalos in Figure 10 is as expected. The orphan catalog has a higher satellite fraction than the standard subhalo catalog; since satellites have preferentially smaller sizes and stronger clustering, the black curves are more negative relative to the red.
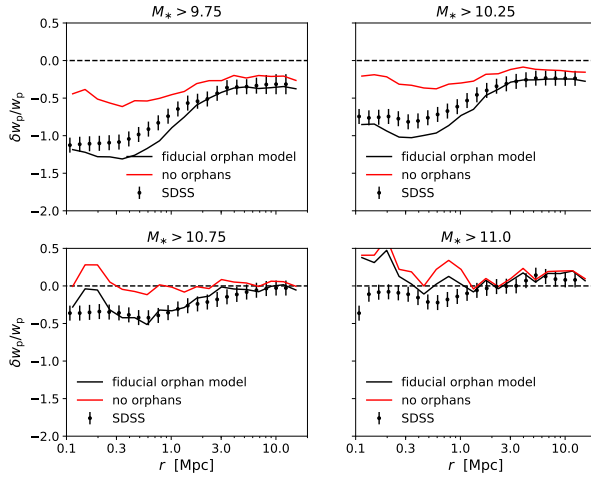
Figure 10. Points with error bars as well as the black curves are the same as in Figure 5. Red curves show predictions for the $R_{1/2}-$clustering ratios of the $R_{\mathrm{vir}}-$based model, but excluding orphan subhalos.

Orphan halos also typically have earlier-than-average values of $z_{M_{\mathrm{peak}}}$ relative to ordinary subhalos, which also enhances differences between central and satellite sizes.

Encouragingly, when tuning $\mathcal{F}_{\mathrm{orphan}}$ so that the overall M_*-threshold clustering is roughly correct (as in Figure 2), the same level of tuning is required for a successful prediction of $R_{1/2}-$dependent clustering. This figure gives quantitative motivation for the program described in §5.3 to jointly constrain the satellite fraction $F_{\mathrm{sat}}(M_*)$ together with models for galaxy size.