Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 29 April 2015 (MN LaTeX style file v2.2)

Mind the Gap: New Tests of Galaxy-Halo Abundance Matching with Galaxy Groups

Andrew P. Hearin[1,2], Andrew R. Zentner[2], Andreas A. Berlind[3], Jeffrey A. Newman[2]
[1] *Fermilab Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, Illinois 60510-0500; aph15@pitt.edu*
[2] *Department of Physics and Astronomy & Pittsburgh Particle physics, Astrophysics and Cosmology Center (PITT PACC), University of Pittsburgh, Pittsburgh, PA 15260; zentner@pitt.edu, janewman@pitt.edu*
[3] *Department of Physics and Astronomy, Vanderbilt University, Nashville, TN; a.berlind@vanderbilt.edu*

29 April 2015

ABSTRACT

MODIFY the abstract. Short statements. Make explicit statements. In particular, we CAN make explicit statements about SHAM assignments (All SHAM models have some tension with the data? vpeak and vzero assignments have tension even with the group abundance!). We CAN make explicit statements about correlations between satellite luminosities or correlations between satellite and central luminosities. We CAN make explicit statements about the CLF assumptions. These may be tentative (and you can use the word tentative), but they can be clear, concise statements. Use the abstract to attract readers! Try to take a stab at this before I go for it. This is an important skill! We employ a mock catalog of galaxy groups constructed via subhalo abundance matching (SHAM) to provide several new tests of the SHAM prescription for the galaxy-dark matter connection. By comparing our mock catalogs to galaxy groups observed in the Sloan Digital Sky Survey (SDSS), we find that abundance matching faithfully reproduces the abundance of galaxy groups as a function of richness ($g(N)$) as well as the relation between group richness, N, and line-of-sight velocity dispersion, σ_v. Additionally, we find that observations of $g(N)$ may be a promising way to constrain models of the stellar mass stripping of satellite galaxies. While the global luminosity function (LF) of galaxies in our mock catalog exactly matches that of the SDSS catalog by construction, we find an intriguing discrepancy between the observed and predicted field (group) galaxy luminosity function, with the SHAM prediction for the LF of field (group) galaxies being systematically too dim (bright). We also test the SHAM prediction for the abundance of galaxy groups a function of magnitude gap, m_{12}, defined as the difference between the r-band absolute magnitude of the two brightest group members. The SHAM prediction for gap abundance is in very good agreement with the data, constituting a new success of the abundance matching prescription. Moreover, our results suggest that the m_{12} abundance is a statistic that is well-suited to constrain the intrinsic scatter in the map between dark matter halo mass and galaxy luminosity, thereby providing complementary constraints on SHAM to $g(N)$. Finally, we generalize existing data randomization techniques to provide a new way to discriminate between competing hypotheses for how galaxies are arranged into groups. We find the hypothesis that the luminosity gap is constructed via random draws from a universal luminosity function to provide a poor description of the data, contradicting recent claims in the literature. We further utilize these data randomization techniques to demonstrate that the observed gap abundance in rich groups is consistent with the hypothesis that the satellite galaxy LF is determined solely by the luminosity of the central galaxy. However, this hypothesis badly fails to describe the gap abundance in groups with only a few members, suggesting the possibility that the satellite LF may require conditioning from a variable in addition to L_{cen} in poor groups, or that draws from $\Phi_{sat}(L)$ and $\Phi_{cen}(L)$ are correlated.

Key words: cosmology: theory – galaxies: structure – galaxies: evolution

© 0000 RAS

1 INTRODUCTION

The centers of dark matter halos are the natural sites for galaxy formation, as these are the locations of the deepest gravitational potential wells in the universe (e.g., **?**). The development of a theory of galaxy formation that encompasses the complex array of physical processes known to contribute to cosmic structure formation is one of the fundamental goals of astrophysics, and ennumerating the connection between galaxies and dark matter halos may help to establish the foundations of any such theory. Additionally, such a connection can serve as an empirical link between large-scale survey data and theoretical predictions. Furthermore, our contemporary theory of cosmology, ΛCDM, makes precise, quantitative predictions for the distribution of dark matter in the universe over a wide range of scales, and so establishing the galaxy-dark matter connection is a key step toward unlocking the predictive power of ΛCDM.

One of the most commonly used techniques for connecting dark matter halos to galaxies is subhalo abundance matching (SHAM). The fundamental tenet of all SHAM models is that there is a monotonic mapping between some elementary property of galaxies (usually luminosity or stellar mass) and an elementary property of halos. SHAM models determine this mapping through the implicit relation defined by matching the predicted abundance of halos with the observed abundance of galaxies. When used in concert with numerical simulations of cosmological structure formation, abundance matching techniques have been shown to predict accurately galaxy clustering statistics (**?????**), the Tully-Fisher relation (**?**), and the conditional stellar mass function (**?**). To date, the vast majority of tests of the SHAM algorithm have been based upon clustering. We will present new tests of the SHAM method.

In our contemporary model of cosmology, ΛCDM, gravitationally self-bound structures form hierarchically, with tiny peaks in the initial cosmic density field collapsing into very small dark matter halos that gradually merge together to form groups and clusters of galaxies. Galaxy groups are thus interesting environments for testing theories of structure formation in general, and the galaxy-halo connection in particular. Indeed, the influence of the group environment on galaxy properties has a long history and has received considerable attention in the recent literature, (e.g., **???????**).

In this paper we investigate SHAM predictions for the assembly of galaxies into groups and test these predictions against galaxy groups observed in SDSS. We briefly describe the SDSS catalog of galaxy groups against which we compare our predictions in § 2. We conduct our study of abundance matching by first constructing a SHAM-based mock catalog of galaxies and then applying a group-finding algorithm on the mock catalog that is identical to that used to find groups in the SDSS galaxy sample. Details of the SHAM algorithm, such as the particular halo property used in the abundance matching, vary in the literature, and we have investigated several choices for the specific implementation of the SHAM algorithm with the aim of determining which methods best reproduce the observed properties of galaxy groups. We describe our methods for constructing our mock catalogs in § 3, and provide a detailed description of some novel features of our SHAM implementation in Appendix A.

By comparing properties of the mock and observed galaxy groups we provide a series of new tests of the abundance matching prescription connecting galaxies to dark matter halos. One of the fundamental properties of any catalog of groups is the *multiplicity function*, the abundance of groups as a function of *richness*, N, the number of members in a group. Previous studies of galaxy group catalogs (**??**) have demonstrated that measurements of the group multiplicity function $g(N)$ may contain valuable information about how galaxies populate dark matter halos. Motivated by this, in § 4.1 we compare the observed multiplicity function to that which is exhibited by our mocks. The line-of-sight velocity dispersion of group members, σ_v, encodes information about the mass of the group (e.g., **?**); in § 4.1 we also investigate the ability of SHAM to predict the correct relationship between group richness and σ_v. While the abundance matching algorithm permits an exact reproduction of the observed galaxy luminosity function by construction, SHAM need not necessarily reproduce the luminosity function of a subsample of galaxies that has been conditioned on some property. In § 4.2 we test the SHAM prediction for the galaxy luminosity function conditioned on whether or not the galaxies are members of a group.

One galaxy group property that has received attention from a rapidly growing body of literature is the luminosity gap, m_{12}, the difference in r-band absolute magnitude between the two most luminous members of a galaxy group. Significant investigation has been focused on a class of systems known as *fossil groups*, usually defined as an X-ray bright ($L_{X,bol} > 10^{42}$erg/s) group of galaxies with $m_{12} \geq 2$. The prevailing theoretical paradigm for fossil group formation is that these systems have evolved quiescently for a significant fraction of a Hubble time, during which dynamical friction has had sufficient time to cause the biggest satellites to merge with the central galaxy, resulting in a massive, bright central galaxy with few bright satellites (**???????????**). For a recent paper on fossil groups that includes an excellent review of the history of their study, we refer the interested reader to **?**. In § 4.3 we study $\Phi(m_{12})$, the abundance of our mock and observed groups as a function of m_{12}, finding that this statistic has the potential to constrain the manner in which galaxies populate dark matter halos. This conclusion is consistent with previous work (**?**), showing that the relative brightness of the central galaxy in a group and its brightest satellite is influenced by parameters governing the conditional luminosity function.

In a paper studying the same galaxy group catalog we use here, **?** found that the observed $\Phi(m_{12})$ is consistent with the hypothesis that the brightness distribution of galaxy group members is determined by a set of random draws from a universal luminosity function. This finding implies that, for a galaxy population of a given luminosity function, the gap abundance is uniquely determined by knowledge of the abundance of groups as a function of richness. This result is particularly surprising in light of recent results (**?**) demonstrating that the magnitude gap contains information about group mass that is independent of richness. A possible resolution to this apparent discrepancy was recently pointed out in **?**: the global gap abundance $\Phi(m_{12})$ is a mass function-weighted sum over the mass-conditioned gap abundance, $\Phi(m_{12}|M)$, and so it is possible in principle that the magnitude gap depends on both mass and richness in such a way that the mass function-weighting washes out any statistically-significant mass dependence in the global $\Phi(m_{12})$.

In § 5 we show that the global gap abundance exhibited by galaxy groups in SDSS is inconsistent with the random draw hypothesis, contradicting the conclusions in **?**. As discussed in Appendix B, we find that a subtle systematic error in the **?** measurement of $\Phi(m_{12})$ is responsible for the difference in our conclusions. Additionally, we generalize these data randomization techniques and test an alternative hypothesis of galaxy group formation, namely that the satellite luminosity function need only be conditioned on the brightness of the central galaxy in order to account for the observed gap abundance. We discuss our results and compare them to those in the existing literature in § 6, and conclude with a brief overview of our primary findings in § 7.

2 DATA

We study galaxy group properties in a volume-limited catalog of groups identified in Sloan Digital Sky Survey (SDSS) Data Release 7 (**?**, DR7 hereafter) using the algorithm described in **?**. This catalog is an update of the **?** group catalog (which was based on SDSS Data Release 3). The galaxies in this sample are all members of the Main Galaxy Sample of SDSS DR7. Groups in this galaxy catalog are identified via a redshift-space friends-of-friends algorithm that takes no account of member galaxy properties beyond their redshifts, positions on the sky, and relative velocities. Our groups are constructed from galaxies in a volume-limited spectroscopic sample ($V_{\rm eff} \simeq 5.8\times10^6 (h^{-1}{\rm Mpc})^3$) in the redshift range $0.02 \leq z \leq 0.068$ with r-band absolute magnitude $M_r - 5\log h < -19$. We refer to this catalog as the "Mr19" group catalog. Each of the 6439 groups in the Mr19 catalog contains $N \geq 3$ members. We refer the reader to **?** for further details on the group finding algorithm.

Fiber collisions occur when the angular separation between two or more galaxies is closer than the minimum separation permitted by the finite width of the optical fibers used to measure galaxy spectra (see **?**, and references therein). We briefly note here that fiber collisions in DR7 are treated differently than the catalog based on DR3 data. As we will see in § 5, this different treatment has important consequences for the measurement of magnitude gaps. In Appendix B we discuss these differences in detail and argue that the DR3 treatment induces systematic errors in magnitude gap measurements that can be avoided if fiber collisions are instead modeled as they are in DR7.

3 MOCK CATALOGS

We compare the SDSS DR7 group data to a mock catalog of galaxy groups based on the Bolshoi N-body simulation (**?**). The Bolshoi simulation models the cosmological growth of structure in a cubic volume $250\,h^{-1}$Mpc on a side within a standard ΛCDM cosmology with total matter density $\Omega_M = 0.27$, Hubble constant $h = 0.7$, power spectrum tilt $n_s = 0.95$, and power spectrum normalization $\sigma_8 = 0.82$. The Bolshoi data are available at `http://www.multidark.org` and we refer the reader to **?** for additional information. Our analysis requires reliable identification of self-bound subhalos within the virial radii of distinct halos. We utilize the `ROCKSTAR` (**?**) halo finder in order to identify halos and subhalos within Bolshoi.

We utilize the subhalo abundance matching (SHAM) technique to associate galaxies with dark matter halos. Although abundance matching is widely used to construct mock galaxy catalogs (e.g., **??????**), our particular implementation of SHAM is novel and so we describe it in detail in Appendix A. In this section we provide a brief sketch of our SHAM prescription and review the primary advantages of our implementation.

SHAM models assume a monotonic relationship between the stellar masses of galaxies and the maximum circular speeds of test particles within their host dark matter halos, $V_{\rm max} \equiv \max\left[\sqrt{GM(<r)/r}\right]$, where r is the distance from the halo center and $M(<r)$ is the halo mass contained within a distance r of the halo center. In practice, stellar masses must be inferred from imaging and spectroscopic data. Inferring galaxy stellar masses is non-trivial, so in practice galaxy luminosities are often used to associate galaxies with halos using SHAM, though this may introduce important biases (e.g., **?**). It is most common to construct SHAM assignments for SDSS data using the r-band absolute luminosities of the galaxies as rough proxies for stellar mass. We follow this approach in this paper.

Assuming a monotonic relationship between $V_{\rm max}$ and r-band absolute magnitude M_r, the SHAM galaxy-halo assignment follows by requiring the number density of galaxies of a given M_r to be equal to the number density of halos of the $V_{\rm max}$ to which they are assigned. As a further complication, subhalos within host dark matter halos evolve significantly due to interactions within the dense environments of the larger host halos. As a result, the present values of $V_{\rm max}$, which we denote $V_{\rm max}^{\rm z=0}$, may be a poor proxy for stellar masses or r-band luminosities. It is now common practice to assign luminosities to subhalos based on their values of $V_{\rm max}$ *evaluated at the time at which they merged into their distinct host halos*, $V_{\rm max}^{\rm acc}$. Often, a halo can be significantly affected by interactions prior to entering the virialized region of a distinct host halo, so it is also interesting to explore using the maximum value of $V_{\rm max}$ ever attained by a subhalo as the stellar mass/luminosity proxy, $V_{\rm max}^{\rm peak}$.

The SHAM assignment of r-band luminosities to halos and subhalos occurs through the implicit relation

$$n_g(<M_r) = n_h(>V_{\rm L}), \qquad (1)$$

where $n_g(<M_{\rm r})$ is the number density of observed galax-

ies with r-band magnitudes brighter than $M_{\rm r}$ (**?**), and $n_h(>V_{\rm L})$ is the predicted number density of dark matter halos and subhalos with assigned circular speeds $>V_{\rm L}$. As we mentioned in the previous paragraph, the circular speeds assigned to subhalos are often not their circular speeds at the time of observation. The quantity $V_{\rm L}$ is evaluated as

$$\begin{aligned} V_{\rm L} &= V_{\rm max}^{\rm z=0} \quad \text{(host halos)} \\ &= V_{\rm sub} \quad \text{(subhalos)} \end{aligned}$$

where $V_{\rm sub} = V_{\rm max}^{\rm z=0}$ if one chooses to use $V_{\rm max}^{\rm z=0}$ to describe the luminosities of subhalos, $V_{\rm sub} = V_{\rm max}^{\rm acc}$ if one chooses to use the maximum circular velocity at accretion for subhalos, and $V_{\rm sub} = V_{\rm max}^{\rm peak}$ if on chooses to use the maximum value of $V_{\rm max}$ ever attained by the subhalo. There is significant empirical support for SHAM models premised on $V_{\rm max}^{\rm acc}$ (e.g., **?**) as well as $V_{\rm max}^{\rm peak}$ (e.g., **?**); however, we explore each of these three choices. Throughout this paper, we refer to mock catalogs constructed using $V_{\rm max}^{\rm z=0}$ as "SHAM0" catalogs, those built with $V_{\rm max}^{\rm acc}$ as "SHAMacc", and catalogs constructed using $V_{\rm max}^{\rm peak}$ as "SHAMpeak".

The SHAM construction summarized in Eq. (1) enables luminosities to be assigned to dark matter halos in a manner that can match *any* observed galaxy luminosity function. Tests of SHAM consist of comparing observational data with independent predictions. The advantages of SHAM-like models is that they are simple, they embody the fundamental theoretical prejudice that dark matter halos that represent deeper gravitational wells should host larger (more luminous) galaxies, and such models describe galaxy clustering over a range of redshifts remarkably well (e.g., **????**).

In practice, some scatter between $V_{\rm max}$ and M_r is often introduced into the basic SHAM assignments. The scatter accounts for the fact that galaxy formation is a complex process, so a single halo parameter cannot specify a stellar mass (or luminosity). Perhaps more importantly, scatter brings SHAM predictions into better agreement with galaxy clustering statistics (e.g., **???**). We investigate the influence of scatter on our results using three different models of scatter between circular velocity and absolute magnitude. Our fiducial model, whose construction is described in detail in Appendix A, is designed to be similar to the model explored in **?** and has 0.2dex of scatter at the faint end (**ARZ: perhaps a parenthetical definition of "faint end" would be useful here.**) and 0.15dex of scatter at the bright end (**ARZ: same for bright end**). [**ARZ: I thought the following statement might help here, please evaluate it as you see fit.**] Our model cannot be precisely the same as that of **?** because we use a different SHAM algorithm to incorporate scatter. Second, we explore a model which has a constant scatter of 0.1dex. We will refer to as our "alternate" scatter model. Third, we consider models with no scatter between M_r and $V_{\rm max}$.

A detail of SHAM implementation is that a specific choice must be made for the galaxy luminosity function to use for the SHAM assignments. A common and convenient choice is to use a fit to observed luminosity functions, such as that provided by **?**. However, in order to ensure that our results are not affected by the residuals of any such fit, the mock catalogs that we construct match the global luminosity function of the Mr19 galaxies *exactly*. Enforcing the requirement that the SHAM galaxy catalog have a luminosity function that matches the observed luminosity function exactly complicates the introduction of scatter into the SHAM prescription. To enforce the observed galaxy luminosity function on our SHAM assignments with scatter in the $V_{\rm L}$-M_r relation, we use a novel implementation of SHAM, the details of which are given in Appendix A. The important features of this implementation are as follows.

1. Our mock galaxy catalog has a luminosity function that matches the observed Mr19 luminosity function *exactly*, by construction.
2. The amount of scatter in the $V_{\rm L}$-M_r relation can be specified simply, so that implementing SHAM assignments with differing amounts of scatter is straightforward.
3. Even when the above two requirements are met, the algorithm is very fast, lending itself to applications that require the construction of a large number of mock catalogs. This advantage is significant compared to, for example, the algorithm of **?**.

Once galaxies with r-band luminosities have been assigned to dark matter halos and subhalos, we proceed to find groups. The galaxies inherit the positions and velocities of their host halos and we have identified groups using the same algorithm that was applied to the observational data. This guarantees that our mock groups are subject to the same redshift-space projection effects as the Mr19 catalog.

Once groups have been found, we introduce fiber collisions to our mock galaxies as follows. For each mock group of richness N we randomly select a Mr19 group with a similar richness. If the number of fiber-collided members of the randomly selected group is $N_{\rm fc}$, then we randomly choose $N_{\rm fc}$ of the members of the mock group and assign fiber collisions to these members. This procedure ensures that the fraction of fiber-collided galaxies in our mock groups *scales with richness in the same way as it does in the data*. Fiber collisions play an important role in our definition of the magnitude gap, as discussed in § 4.3. [**ARZ: Perhaps we should mention this caveat simply so that we do not appear naive.**] Of course, this correction does not account for the spatial biases of fiber collisions, but we do not consider spatial clustering in the current study.

4 NEW TESTS OF ABUNDANCE MATCHING

In this section, we scrutinize the abundance matching prescription for the mapping between galaxies and dark matter halos in several novel ways. In § 4.1, we consider the number density of groups as a function group richness, in § 4.2 we study the group and field luminosity functions of galaxies, and in § 4.3 we investigate the distribution of galaxy luminosities within groups, concentrating significant attention on the magnitude gap statistic. As discussed in § 3, we explore three distinct assignments for the effective maximum circular velocities to be used in the luminosity assignments of subhalos, namely "SHAMacc", "SHAMpeak", and "SHAM0." We also analyze SHAM predictions based upon three distinct models for the amount of scatter between M_r and circular velocity. Throughout this paper we refer to the SHAMacc model (in which the maximum circular velocity at the time of accretion is used for subhalos) with 0.2dex of scatter at the faint end and 0.15dex of scatter at the bright end, following **?**, as our default model. We refer explicitly to results from the other models where relevant.

4.1 Multiplicity Function

In this section, we compare the predictions of abundance matching for the average number density of groups as a function of the number of group members, N. We refer to this abundance as the group *multiplicity function* and represent it as $g(N)$. The multiplicity function of the observed SDSS Mr19 sample is plotted in blue diamonds in the top panel of Fig. 1. The points represent the mean number densities of groups, while the errors were computed by bootstrap resampling of the group catalogs. For all of the results in this paper, our bootstrap errors are computed as the variance over 10^4 bootstrap realizations, where each realization is constructed by randomly selecting (with replacement) the same number of objects in the sample. In all but the smallest richness bin, the observed multiplicity function is consistent with a power-law of $g(N) \propto N^{-2.5}$.

In addition to depicting the SDSS Mr19 group multiplicity function, Fig. 1 shows how these data may be used to scrutinize empirical methods to assign galaxies to halos, such as SHAM. To illustrate this, we plot with red triangles the multiplicity function prediction of our fiducial SHAMacc mock catalog. The upper (lower) dashed curve in the top panel of Fig. 1 shows the SHAMpeak (SHAM0) prediction for the multiplicity function; error bars for these two models have been omitted as they are similar to those from our fiducial SHAMacc model. To further facilitate the illustration of the potential of $g(N)$ measurements to discriminate between different abundance matching prescriptions, in the bottom panel of Fig. 1 we plot the fractional difference between the predicted and observed multiplicity functions. All the points appearing in Fig. 1 trace models with our fiducial amount of scatter between M_r and $V_{\rm max}$.

Our fiducial SHAMacc model, is consistent with the observed multiplicity function well within 2σ for any richness cut $N \geq 2$. On the other hand, the SHAM0 assignment with $V_{\rm sub} = V_{\rm max}^{\rm z=0}$ is a significantly poorer description of the data. SHAM0 underestimates the abundances of all groups with $N \geq 5$. The significance of the difference between the SDSS group data and the predicted SHAM0 multiplicity function is $\simeq 4.9\sigma$ for groups with $N \geq 5$, and $\simeq 4.0\sigma$ for $N \geq 10$ groups. These discrepancies hold true at similar levels in the alternate scatter models we studied.

This discrepancy may not be surprising because SHAM assignments based upon $V_{\rm max}^{\rm z=0}$ have already been shown to be less effective at describing independent data than SHAM assignments based upon $V_{\rm max}^{\rm acc}$ (e.g. **????**). More interestingly, our SHAMpeak assignments of galaxies to halos with $V_{\rm sub} = V_{\rm max}^{\rm peak}$ do not describe the data as well as SHAMacc either. SHAMpeak models significantly *overestimate* the abundances of rich groups. The statistical significance of this discrepancy is $\simeq 3.9\sigma$ for groups with $N \geq 5$ members, and $\simeq 2.6\sigma$ for groups with $N \geq 10$.

These results are interesting as a demonstration that the multiplicity functions of groups can be used as valuable statistics with which to constrain the connection between dark matter and galaxies. This group multiplicity function test clearly indicates that the SHAMacc assignment with $V_{\rm sub} = V_{\rm max}^{\rm acc}$ is consistent with SDSS group data while the alternative SHAM0 and SHAMpeak assignments cannot describe the SDSS group data adequately. In particular, note that the SHAMacc results are straddled by the SHAMpeak and SHAM0 results, suggesting that it may be possible to use $g(N)$ measurements to constrain models of the mass stripped from satellite galaxies (e.g. **?**). The success of our SHAM models with $V_{\rm sub} = V_{\rm max}^{\rm acc}$ motivates the choice of SHAMacc as our fiducial model.

We have also investigated the role that scatter plays in the abundance matching prediction for $g(N)$. Each of the models depicted in Fig. 1 pertain to our fiducial model of scatter, which has 0.2dex of scatter at the faint end and 0.15dex at the bright end. However, we find no qualitative change to our results when using our alternate scatter model (which has a constant 0.1dex of scatter) or models with no scatter. In particular, we find that for all the scatter models we explored, SHAMpeak models significantly overestimate the abundances of rich groups and SHAM0 models significantly underestimate the abundances of rich groups. Accordingly, we do not explicitly show the results for the alternative scatter models for the sake of brevity.

In this subsection *only*, we study an alternate SDSS catalog that is complete to $M_{\rm r} \leq -20$, spans the redshift range $0.02 < z < 0.106$, and has an effective volume of $V_{\rm eff} \simeq 2.3 \times 10^7$ Mpc$^3/h^3$ to test the sensitivity of our conclusions to the sample selection. We constructed mock catalogs for this Mr20 catalog in exactly the same fashion for Mr19, except that we imposed a brightness cut of $M_{\rm r} \leq -20$ for the mock galaxies. Our findings for the Mr20 catalog are the same as for Mr19: the observed $g(N)$ is well-fit by a power law with exponent -2.5 in all but the smallest richness bin; the multiplicity function predicted by our fiducial model of the Mr20 groups exhibits less than 1σ discrepancy with the data; SHAMpeak (SHAM0) significantly over-(under-)predicts the abundances of rich groups.

In addition to group abundance as a function of richness, we also present results concerning the map between group richness and line-of-sight velocity dispersion in Figure 2. The agreement is quite good except for possibly in the largest σ_v bin. Plots based on $V_{\rm max}^{\rm peak}$ or $V_{\rm max}^{\rm z=0}$ results look similar, demonstrating that this mapping is a generic success of the SHAM paradigm. For reference, we also fit the Mr19 data to a power law $N \propto \sigma_v^3$, and plot the result with the black solid line.

© 0000 RAS, MNRAS **000**, 000–000

Figure 1. Group abundance as a function of the number of group members. In the top panel we illustrate a comparison of the group multiplicity function $g(N)$ seen in theMr19 SDSS catalog (blue diamonds) and that in our fiducial mock catalog (red triangles). Our fiducial mock was made with abundance matching on $V_{\rm max}^{\rm acc}$. We also display results when the abundance matching is done on $V_{\rm max}^{\rm peak}$ (top dashed curve) as well as $V_{\rm max}^{\rm z=0}$ (bottom dashed curve). Error bars for these alternate SHAM models have

Figure 2. Group richness N, as a function of line-of-sight velocity dispersion σ_v. The SDSS Mr19 measurement of this mapping appears in blue, the prediction of our fiducial abundance matching model appears in red. The black line represents a fit of the Mr19 data to a power law $N \propto \sigma_v^3$, as this is the dependence that would be expected if all the galaxy groups are virialized and group mass $M \propto N$. [**ARZ: labels bigger! How exactly did you do this? What do the horizontal error bars signify?**]

4.2 Field & Group Galaxy Luminosity Functions

The SHAM method for assigning galaxies to halos results in mock luminosity functions that match observed luminosity functions by construction. However, this procedure does not guarantee agreement between luminosity functions that are conditioned on a specific galaxy property or environment. In this section, we consider a simple distinction in galaxy environment based directly on our group catalogs. Specifically, we explore the luminosity functions of galaxies conditioned upon whether or not the galaxy is identified as a member of a group. We refer to the luminosity function constructed from all galaxies residing in groups as $\Phi_{\mathrm{group}}(L)$. Likewise, we refer to all galaxies that we do not identify as members of a group as "field" galaxies and the luminosity function conditioned on the galaxy being a member of the field as $\Phi_{\mathrm{field}}(L)$. SHAM predictions for $\Phi_{\mathrm{group}}(L)$ and $\Phi_{\mathrm{field}}(L)$ are *not* guaranteed to match observational determinations of these quantities, so this is a test of the allocation galaxies to group and field environments by SHAM methods.

Figure 3 shows comparisons between our predicted, SHAM luminosity functions for group galaxies and field galaxies and the corresponding observed luminosity functions. To highlight differences, the quantity shown on the vertical axis of all panels in Fig. 3 is the fractional difference $\Delta\Phi/\Phi_{\mathrm{SDSS}}$ between the predictions of the SHAM mocks and the observations, where $\Delta\Phi(L) \equiv \Phi_{\mathrm{mock}}(L) - \Phi_{\mathrm{SDSS}}(L)$, so that points in Fig. 3 with positive vertical axis values correspond to luminosities where SHAM over-predicts the abundance of galaxies of that brightness, and conversely for negative values of $\Delta\Phi(L)$. Blue diamonds show the error on $\Phi(L)$ for field galaxies, red triangles show the same for group galaxies. Group galaxies defined by a richness cut of $N_{group} \geq 3$ appear in the left columns, $N_{group} \geq 10$ appear in the right columns. Our fiducial model, SHAMacc which abundance matches on $V_{\mathrm{max}}^{\mathrm{acc}}$, appears in the top panels while the alternate, SHAMpeak model which uses $V_{\mathrm{sub}} = V_{\mathrm{max}}^{\mathrm{peak}}$, appears in the bottom panels; all abundance matching prescriptions in Figure 3 were constructed with our fiducial scatter. The error bars on $\Phi_{\mathrm{SDSS}}(L)$ and $\Phi_{\mathrm{mock}}(L)$ have each been estimated by bootstrap resampling as in § 4.1. We reiterate that the differences shown in this plot are strictly due to the separation of field galaxies from group galaxies because our SHAM implementation *guarantees* that the overall luminosity function matches the data *exactly.*

As Fig. 3 shows, our fiducial SHAMacc model, the baseline SHAM procedure which is known to describe galaxy clustering correctly and which provided an adequate description of the group multiplicity function in § 4.1, systematically *overestimates* the abundance of dim, field galaxies and *overestimates* the abundance of bright, group galaxies. Notice that this qualitative conclusion holds irrespective of the choice for the minimum number of member galaxies required in order to have the group members be included in the group luminosity function. In fact, the discrepancies grow larger as the number of group members grows, emphasizing that the SHAM excesses of bright, group galaxies grow more egregious for larger groups (particularly for SHAMpeak). These results highlight at least one weakness of the SHAM procedure for exploring the connection between galaxies and dark matter halos. The widely-used SHAM procedure that adequately describes low-redshift galaxy clustering *does not allocate luminosities to group and field halos in a manner consistent with observations.*

[**ARZ: Naively, one might expect that the sum of the field and group luminosity functions must be the global luminosity function. This isn't true given your definitions. However, this may lead to a minor confusion interpreting Fig. 3. In particular, one might expect that both the diamonds and triangles cannot be positive (or negative) simultaneously. However, this does occur at some points. I assume that this occurs because the intervening multiplicities (say objects with $N_{group} = 2$ in the left panels) make up the difference so that the global luminosity function is always fixed? If this is true, you should add a sentence or two somewhere to assure the reader of this.**]

We also explore the influence that scatter between absolute magnitude and halo circular velocity has on the SHAM predictions for the group and field luminosity functions. In Figure 4, we again plot the fractional difference between the observed and predicted $\Phi_{\mathrm{group}}(L)$ (top left panel) and $\Phi_{\mathrm{field}}(L)$ (top right panel), this time comparing results between mocks made with different amounts of scatter. The SHAMpeak abundance match-

Figure 3. Fractional differences between the field and group galaxy luminosity functions predicted by abundance matching and observed in SDSS. The fractional differences for field (group) galaxies appear as blue diamonds (red triangles). For this purpose, field galaxies are defined to be those galaxies that do not reside in groups, and group galaxies are defined by the richness cut given in the legend of each panel. Results pertaining to our fiducial SHAM model, SHAMacc based on abundance matching with $V_{\mathrm{sub}} = V_{\mathrm{max}}^{\mathrm{acc}}$, appear in the top panels. Results pertaining to SHAM using $V_{\mathrm{max}}^{\mathrm{peak}}$, SHAMpeak, appear in the bottom panels. The left panels show results for groups with multiplicity $N_{group} > 2$ while the right panels show results for $N_{group} > 9$. The group (field) galaxy luminosity functions in all of our SHAM mocks is systematically too bright (dim).

ing procedure was used for all models depicted in Figure 4. Blue diamonds in Fig. 4, give results pertaining to our fiducial scatter model, which has 0.2dex of scatter at the faint end and 0.15dex of scatter at the bright end, magenta squares designate our alternate model with a constant 0.1dex of scatter, and red triangles depict our SHAMacc mock without scatter. In all relevant panels of Fig. 4, group galaxies have been defined by the richness cut $N_{group} > 4$.

Evidently, the amount of scatter in SHAM significantly influences the predictions for $\Phi_{\mathrm{group}}(L)$ and $\Phi_{\mathrm{field}}(L)$ at the bright end, but appears to have little role in the group and field luminosity functions at the faint end. A more exhaustive exploration of different prescriptions for the scatter may yield a model that correctly predicts the abundance of *bright* group and field galaxies. Any such model would require significantly larger scatter than the fiducial model we have used in this work, which is, in turn, based on the success of **?** in describing galaxy clustering and a variety of other properties using SHAM. [**ARZ: Can we guess/estimate what scatter this would have to be? 0.3dex? 0.4dex? It would be**

useful to put an approximate number here because I think scatter as large as those numbers is too big to be consistent with other papers?] However, the robustness of our results to differences at the faint end of the luminosity functions indicates that this is a generic weakness of SHAM that cannot be overcome by adding scatter alone. We discuss this point further in § 6.

In the bottom panels of Fig. 4 we illustrate how the errors in the SHAM prediction for $\Phi_{\mathrm{group}}(L)$ translate into errors in the luminosities of the brightest group members (bottom left panel) and next-brightest group members (bottom right panel). Qualitatively, the trends in the bottom panels reflect the sense of the errors on $\Phi_{\mathrm{group}}(L)$ shown in the upper left panel. SHAM predicts brightest group galaxies and next-brightest group galaxies that are significantly too bright on average, compared to observations. Moreover, this conclusion is insensitive to scatter, so it is, again, a *generic weakness of the SHAM assignments*.

Before proceeding, we note that we include the bottom panels in Fig. 4 in large part because our results in the following sections focus on the relative brightnesses of the brightest and next-brightest group members, and so the errors illustrated in the bottom panels are germane to all of our results pertaining to the SHAM prediction for this relative brightness.

4.3 Magnitude Gap Abundance

[ARZ: There is a lot of interpretive work here, which makes this a strong paper. However, to some degree, simply ennumerating various quantities like the fossil fraction and the magnitude gap abundance function could be very interesting to many people. For each statistic, you should state why your presentation is unique. My understanding is that you fossil fraction and magnitude gap abundance functions supersede what has been done. You should state this explicitly where appropriate in this section.]

In § 3, we compared the luminosity functions of groups predicted by SHAM to SDSS group data. In this section, we explore a group statistic that is related to the group luminosity function known as the "magnitude gap." We define the magnitude gap to be the difference in r-band absolute magnitude between the two brightest non-fiber collided members of any group, $m_{12} = M_{r,2} - M_{r,1}$, where $M_{r,\mathrm{i}}$ is the r-band absolute magnitude of the ith brightest group member. In particular, we will be interested in $\Phi(m_{12})$, the statistical distribution of magnitude gaps, defined so that $\Phi(m_{12})\mathrm{d}m_{12}$ represents the number density of galaxy groups with magnitude gap m_{12} in a bin of width $\mathrm{d}m_{12}$.

As we discussed in § 1, the magnitude gap abundance statistic has received significant attention in recent literature. Of course, it is possible to enumerate the absolute magnitudes of all group members, but the magnitude gap has received particular attention for a number of reasons. These include: (1) the simplicity of a single statistic; (2) dynamical friction timescales vary in inverse proportion to galaxy mass, so the largest satellites merge the most quickly, making magnitude gap a possible indicator of the dynamical age of a group; and (3) the magnitude gap may help to refine mass estimates for optically-identified clusters (**??**).

In this section we focus on the magnitude gap as a simple diagnostic of the partitioning of galaxies among groups. Magnitude gaps provide a test of the relationship between galaxies and dark matter halos that is distinct from the tests provided by measurements of group abundance (§ 4.1) and the group luminosity function (§ 4.2). Galaxy luminosities can be partitioned among groups and group members in an infinite variety of ways that all lead to the same average luminosity function. An empirical model such as SHAM that produces the correct average group luminosity function at a particular richness (or range of richnesses) may nevertheless fail to produce the correct distribution of magnitude gaps.

The abundance of groups by magnitude gap is a rapidly declining function of the gap, with approximately 90% of all groups with $N \geq 3$ members in the Mr19 sample having a magnitude gap smaller than $m_{12} = 1.5$. However, gap abundance depends sensitively on group richness, as demonstrated in the top left panel of Figure 5, where we show a histogram (normalized to have unit area) of $\Phi(m_{12})$ exhibited by galaxy groups in two different richness ranges. The blue histogram traces $\Phi(m_{12}|N = 3)$, the gap abundance of groups with $N = 3$ members, while the red histogram traces $\Phi(m_{12}|9 < N < 16)$. A comparison of the two histograms provides a demonstration of the richness-dependence of m_{12} abundances: richer groups tend to have smaller magnitude gaps. This trend has been demonstrated previously in the literature (e.g., **???**).

A simple way to gain insight into the sense of this trend is to consider a toy model universe in which galaxy groups are assembled by randomly drawing galaxy luminosities from a global luminosity function $\Phi_{\mathrm{global}}(L)$ (for example, a Schechter function). As the richness N of the toy groups increases, the number of random draws from Φ_{global} increases, and the probability that a very bright member is drawn increases. Denoting the luminosity of the i^{th} brightest members as L_i, the expectation value of L_i becomes brighter with increasing N. As the number of random draws increases, the expectation value of L_1 is the first to become brighter than L_*, the exponential cutoff of Φ_{global}, and when this occurs there is a rapid decrease in the rate at which the expectation value of L_1 brightens with increasing N. The reason for this rapid decrease is because of the exponential suppression in the luminosity function for galaxies with luminosties exceeding L_*; most draws correspond to lower luminosities. A relatively larger number of draws is required for the expectation value of L_2 to exceed L_*, and so as N continues to increase the expectation value of the ratio L_1/L_2 decreases.[1]

[1] Note that this intuitive explanation also demonstrates that the root cause of the shrinking of the magnitude gap with N random draws is that the slope of Φ_{global} steepens as the brightness increases. If Φ_{global} did not have this property, for

Figure 4. Fractional difference between a variety of conditioned luminosity functions seen in SDSS and the predictions for SHAMacc with our fiducial scatter model (blue diamonds), our alternate scatter model (magenta squares), and no scatter (red triangles). For this purpose, field galaxies are those galaxies that do not reside in groups, and the group galaxy sample is defined by requiring that each galaxy in the sample reside in a group with $N > 4$ members. The top two panels show the differences in the group (left) and field (right) luminosity functions, similar to Fig. 3, in order to address the influence of scatter in the SHAMacc predictions. After rank-ordering the members of each group by their brightnesses, we have measured the luminosity function of the brightest groups members as well as that of the next-brightest group members. Fractional differences between the SHAMacc predictions for the luminosity function brightest galaxies in $N > 4$ groups and the observed brightest galaxy luminosity function appear in the bottom left panel. Analogous differences for the next brightest group galaxies in the bottom right panel.

To illustrate this point explicitly, the blue diamonds and red triangles in Figure 5 show histograms of $\Phi(m_{12})$ in Monte Carlo (MC) realizations of this toy universe. For each observed Mr19 group of richness N, we have constructed 1000 realizations of the group by randomly drawing N times from the *observed* $\Phi(L|N \geq 3)$ (note that the observed luminosity function is used and not a Schechter function approximation). Thus, for group samples in each richness bin plotted in the top, left panel of Fig. 5, the multiplicity functions of the observed and MC groups match *exactly*. The difference between the r-band magnitudes of the brightest two draws gives the

example if the global luminosity function were a simple power law, $\Phi_{\mathrm{global}}(L) \propto L^{\alpha}$, then the expectation value of the magnitude gap in the random draw universe would be entirely independent of N since the expectation value of L_1 and L_2 would each brighten with increasing N at the same rate.

Figure 5. [ARZ: Be a bit careful here. We may want to have a figure that prints clearly in Black & White just in case. Also, you should use the notation $F_{\rm fos}$ since you gone through the trouble of defining it in the text.] Fossil group statistics. In the top left panel the blue (red) histogram traces the relative abundance of groups as a function of magnitude gap, $\Phi(m_{12})$ (normalized to have unit area), exhibited by groups with richness $N = 3$ ($9 < N < 16$). The blue diamonds and red triangles trace $\Phi(m_{12})$ of the corresponding Monte Carlo randomizations of the groups. **[ARZ: I think the x-axis label should probably just be N rather than $> N$, do you agree?]** In the top right panel we plot the *fossil fraction* $F_{\rm fos}(> N)$, defined as fraction of systems in a group sample with magnitude gap $m_{12} \geq 2$. The richness-threshold defining the samples used in the fossil fraction measurements appears on the horizontal axis. Blue diamonds (red triangles) show the fossil fraction of the observed (Bolshoi with SHAMacc) group samples for a range of richness cuts. In the bottom left panel we plot the magnitude gap abundance $\Phi(m_{12})$, exhibited by groups with $N > 4$ members. The observed $\Phi(m_{12})$ is illustrated with blue diamonds; we have chosen a random subsample of our mock groups with a multiplicity function matching that of the observed groups and plotted the gap abundance exhibited by this random subsample with red triangles. The purpose of the multiplicity function matching is to decouple the dependence of the predicted gap abundance from the richness-dependence illustrated in the top panels. In the bottom right panel we show the fractional difference between the observed gap abundance and that predicted by SHAM models for different amounts of scatter between $M_{\rm r}$ and $V_{\rm L}$, demonstrating the potential to use $\Phi(m_{12})$ measurements to constrain the scatter in SHAM models.

m_{12} value of the MC group. Blue diamonds give the $\Phi(m_{12}|N=3)$ that results from this exercise, red triangles represent $\Phi(m_{12}|9<N<16)$. Of course, this toy model ignores any evolution of, or interactions between, group members during the process of group formation, although the broad similarity between $\Phi(m_{12}|N)$ in the MC and in Mr19 data demonstrate that this model nonetheless provides a reasonable approximation of the relationship between richness and magnitude gap.

Our chief goal in this section is to demonstrate the utility of the observed gap abundance $\Phi(m_{12})$ for constraining the galaxy-dark matter connection, with a particular focus on SHAM-based models. A basic consequence of the relationship between gap and richness is that the multiplicity function $g(N)$ plays a critical role in $\phi(m_{12})$. As shown in § 4.1, SHAMpeak and SHAM0 models over- and under-predict $g(N)$ at moderate and large values of group richness, respectively. Therefore, we should expect that these models will not predict the correct gap abundance function, $\Phi(m_{12})$. However, it is still useful to explore the distribution of magnitude gaps, given a common richness or distribution of richness. Such a statistic eliminates the multiplicity function $g(N)$ as a possible cause of discrepancy, so it can serve as a pure test of the ability of the SHAM formalism to allocate the brightest galaxies into physically associated, group-sized systems. We proceed to undertake such a comparison by randomly selecting a subsample of mock groups from our SHAM catalogs with a multiplicity function that matches that of the observed, SDSS Mr19 sample. Specifically, we compare the observed number density of groups as a function of magnitude gap, $\Phi_{\mathrm{SDSS}}(m_{12})$ to that in a subsample of our SHAM mocks restricted to have an identical group multiplicity function, $\Phi_{\mathrm{mock}}(m_{12}|g=g_{\mathrm{SDSS}})$.

The top right panel of Figure 5 shows one result of such a test. Plotted as blue diamonds in Fig. 5 is the *richness-threshold conditioned fossil fraction* $F_{\mathrm{fos}}(>N)$, defined as the fractional abundance of SDSS Mr19 galaxy groups with more than N members that have $m_{12}\geq 2$,:

$$F_{\mathrm{fos}}(>N)\equiv\frac{\int_2^\infty \mathrm{d}m_{12}\Phi(m_{12}|>N)}{\int_0^\infty \mathrm{d}m_{12}\Phi(m_{12}|>N)}. \qquad (2)$$

The fossil fraction is a measure of the size of the large-gap tail of the magnitude gap distribution; galaxy groups with $m_{12}\geq 2$ whose X-ray brightnesses are greater than some threshold value (commonly $L_{X,bol}>10^{42}$erg/s) are often referred to as *fossil groups*, and are conventionally thought to be galaxy systems that assembled most of their mass at high redshift, representing the end products of galaxy group evolution. We address the consistency of this interpretation with our findings in § 5 and § 6. The decrease of $F_{\mathrm{fos}}(>N)$ with increasing N reflects the relationship between richness and gap discussed above: large gap systems are rarer in systems of larger richness.

Plotted in red triangles in the top right panel of Figure 5 is the SHAMacc prediction for $F_{\mathrm{fos}}(>N)$ from our fiducial mock catalog after selecting a random subsample with a $g(N)$ distribution that matches the observed Mr19 multiplicity function. For both the mock and observed data points the error bars come from bootstrap resampling. In the case of the SHAMacc groups, each bootstrap realization corresponds to a new, random selection of a subsample of the mock groups with a matched $g(N)$ distribution. Regardless of the richness threshold, the difference between the observed fossil fraction and that predicted our fiducial SHAMacc model is less than 2σ. [**ARZ: This isn't entirely obvious from the figure: do you mean less than 2-sigma at any given point, N, or less than 2-sigma even after summing over all points.**]

[**ARZ: Is this the most complete exposition of the fossil group abundance and the magnitude gap distribution? If so, you should say that somewhere and make this result known.**] In the bottom left panel of Fig. 5, we show the gap abundance $\Phi(m_{12})$ predicted by our fiducial mock catalog and compare it to the observed Mr19 abundance. As in the upper right panel of Fig. 5, the subsample of mock groups has been chosen to match the observed multiplicity function. Thus this figure illustrates the results of a direct test of the abundance matching prediction for the relative brightnesses of the brightest group galaxies, independent the observed group multiplicity. There is less than a 1σ difference between the observed $\Phi(m_{12})$ and our fiducial prediction, which constitutes a new success of SHAM-based models for the galaxy-dark matter connection. With the same model of scatter, the matched $g(N)$ prediction for $\Phi(m_{12})$ when abundance matching with either $V_{\mathrm{max}}^{\mathrm{peak}}$ or $V_{\mathrm{max}}^{\mathrm{z=0}}$ results in less than a 2σ discrepancy with the data. The magnitude gap is not an effective statistic with which to discriminate between the different SHAM algorithms in common use. [2]

The success with which SHAM describes the distribution of magnitude gaps may seem less interesting because we have already shown that SHAM does not accurately predict the luminosities of our group galaxies (§ 4.2, Figs. 3 & 4). However, we first note that although these results are related, the gap abundance prediction does not follow directly from the group galaxy luminosity function. For example, one could have imagined the average luminosities of group galaxies to have been *correctly* predicted, while the distribution of magnitude gaps was incorrect due to a failure of SHAM to correctly arrange bright galaxies within groups. In this way, one can see that magnitude gaps test not just the mean conditional luminosity function (CLF) of group galaxies, but also correlations between the luminosities of group members.

Additionally, magnitude gap abundance observations provide constraints on the SHAM models that are *complementary* to measurements of group multiplicity. We illustrate this in the bottom right panel of Fig. 5. In that panel, we show the fractional difference between the observed abundance of groups as a function of magnitude gap and that predicted by SHAM. The difference, $\Delta\Phi(m_{12})=\Phi_{\mathrm{mock}}(m_{12})-\Phi_{\mathrm{SDSS}}(m_{12})$, analogous to the

[2] Note that if this comparison is done without first matching the predicted $g(N)$ to the observed group multiplicity then SHAMpeak and SHAM0 predictions for $\Phi(m_{12})$ are in stark disagreement with the data, demonstrating the importance of multiplicity matching when one is interested solely in the relative luminosities of the brightest group galaxies.

group and field luminosity functions defined in § 4.2. The red triangles show results from our fiducial scatter model (0.2dex of scatter at the faint end, 0.15dex at the bright end), the magenta squares depict our alternate scatter model (constant scatter of 0.1dex), and the blue crosses show SHAMacc with no scatter. It is evident that $\Phi(m_{12})$ is a relatively sensitive probe of the underlying scatter between luminosity and $V_{\rm L}$. The alternate scatter model is discrepant with the data at 2.7σ, the no scatter model at 4.1σ, strongly suggesting that $\Phi(m_{12})$ observations can be exploited to constrain the scatter between luminosity and halo circular velocity. We reiterate that the gap abundance prediction has been decoupled from the group multiplicity prediction for each model, so the complementarity of the constraints on the SHAM model provided by $\Phi(m_{12})$ and $g(N)$ can be realized as these statistics can be used concurrently. Magnitude gap or related statistics may thus prove to be incisive statistics with which to constrain scatter in galaxy-halo assignments, but we relegate a detailed study of this possibility to future work.

We repeated this entire exercise for mock groups identified in real space (as opposed to redshift space), and found that the observed $\Phi(m_{12})$ and that predicted by our fiducial mock are different at a level of $\simeq 4.5\sigma$. This demonstrates the importance of redshift-space group-finding in making the prediction for the gap abundance. To our knowledge, we have performed this analysis for the first time; all previous studies relying on numerical simulations to predict gap abundances have used "halo-level" abundances as the prediction, in which halo membership is used to define group membership. However, real-space predictions systematically under-estimate the abundance of low-gap systems, a fact that we find to hold true regardless of the SHAM prescription. This is sensible since interlopers occur more often in redshift-space groups, and interlopers can only reduce the gap, so it is natural to expect that including interlopers by finding the groups in redshift-space should boost the low-gap abundance. Any theoretical study of the magnitude gap abundance must properly account for interlopers due to redshift-space projection effects in order to predict $\Phi(m_{12})$ correctly.

5 DATA RANDOMIZATION COMPARISONS

In the previous section, we explored a variety of properties of galaxy groups including group multiplicity and magnitude gap. As we pointed out in our discussion of Figure 5 at the beginning in § 4.3, it is natural to expect the number of group members to be correlated with, for example, the luminosity of the brightest group galaxy or the magnitude gap. The reason is simple. Consider a toy model in which the luminosities of group members are consistent with being random draws from a universal group luminosity function. In this case, the typical luminosity of the brightest group galaxy will increase with the number of group members. Likewise, the magnitude gap will decrease with the number of group members. The fidelity with which such a toy model represents the real universe has been explored previously by **?**, and it is indeed interesting to determine whether observed galaxies are consistent with such a simple hypothesis.

We revisit this investigation here and proceed as follows. For a given set of groups, let $\Phi(m_{12}| > N)$ be the abundance of the groups of a given magnitude gap subject to the condition that the group has greater than N members (of course, other conditions could be placed on this distribution as well, see below for further discussion). Assume some luminosity function of galaxies within such groups, $\Phi(L| > N)$. If we assume that galaxy luminosities are consistent with random draws from $\Phi(L| > N)$, there is a definite prediction for the distribution of magnitude gaps in this random-draw hypothesis, $\Phi_{\rm rand}(m_{12}| > N)$. Of course, real data need not be consistent with this hypothesis, so it is interesting to examine deviations from the random draw hypothesis. We do this by defining the fractional deviations from the random-draw prediction,

$$\Psi(m_{12}| > N) \equiv \frac{\Phi(m_{12}| > N) - \Phi_{\rm rand}(m_{12}| > N)}{\Phi_{\rm rand}(m_{12}| > N)}. \quad (3)$$

Of course, one can make different assumptions about the luminosity function from which the galaxies are being drawn in order to test different hypotheses. For example, one could draw the luminosities from the all-galaxy luminosity function $\Phi(L)$, rather than a richness threshold-conditioned LF. We intend to explore these and related details in a future follow-up paper.

[ARZ: Isn't "$>N=3$" a goofy notation? Shouldn't it be ">3" to follow the prototype given in the paragraph above? After all, N is a random variable and 3 is a particular value of that variable that may or may not be realized. I think this notation should be changed a bit to reflect this or something similar to this. Also, the description here seems to have slightly different numbers from the figure labels. Please check this for consistent notation.] With the blue diamonds in Figure 6, we show $\Psi(m_{12}| > N = 3)$ for SDSS Mr19 groups in the top panel and $\Psi(m_{12}| > N = 10)$ in the bottom panel. This is similar to the case that was tested in **?**, except those authors used an analytical fit (**?**) to $\Phi(L| > N)$, and restricted attention to $N > 10$. The $\Psi(m_{12}| > N)$ in Fig. 6 are clearly inconsistent with zero. The statistical significance of this difference is $\simeq 4.6\sigma$ for $N > 3$ groups and $\simeq 2.7\sigma$ for $N > 10$ groups. Alternatively, the p-values from two-sided KS tests are $\lesssim 10^{-4}$ in both cases. This test yields the clear conclusion that the luminosities of galaxies within SDSS groups are not consistent with random draws from the global luminosity function. This is a direct contradiction of the conclusions drawn in **?**. The source of the discrepancy between these two conclusions lies in the treatment of fiber collisions, which we provide a detailed account of in Appendix B.

We have also measured the $\Psi(m_{12}| > N)$ that results from an alternative data randomization procedure, which we describe as follows. First, we divide our SDSS Mr19 sample of galaxies into "centrals" and "satellites"; the set of centrals is defined to be those galaxies that are the brightest among the galaxies in the group of which they are a member; the set of satellites is the complement to the set of centrals. For each group of richness N in the sample being randomized, we construct 1000 realizations of the group. For each realization, we fix L_1, the luminosity of the "first" randomized group member, to be equal to L_{cen}, the luminosity of the central galaxy in the group that is being randomized. The luminosities of the remaining $N-1$ members are drawn at random from $\Phi_{sat}(L|L_{cen})$, the luminosity function of all satellite galaxies that are found in groups whose central galaxy has a luminosity within 0.2dex of L_{cen}.

This randomization scheme may appear unfamiliar at first glance, but in fact it is well-motivated by and quite similar to the method by which mock catalogs of galaxies are constructed from N-body simulations in the standard Conditional Luminosity function (CLF) formalism. For CLF-based mocks, the value L_{cen} is chosen at random from a log-normal luminosity function whose mean and spread are governed by the mass of the dark matter halo, and the luminosities of the remaining galaxies associated with the halo are chosen from a modified Schechter function whose form is governed by L_{cen}. Our approach to the construction of our randomized groups is very similar except that we know the richnesses of the groups rather than their masses. By keeping fixed the luminosity of the central galaxy in each randomized group, this randomization preserves the relationship between L_{cen} and N exhibited by the observed groups. Rather than using an analytical fit to Φ_{sat}, we use the data itself determine this distribution; our choice to condition the luminosity function from which the brightness of each randomized group's satellites are drawn mirrors the convention adopted by the standard CLF formalism.

A non-negligible fraction of the randomly-drawn satellites are brighter than the central galaxy in the randomized group. However, in calculating the magnitude gap of each randomized group we proceed exactly as we do with the data and define m_{12} to be $M_{r,1}-M_{r,2}$, the magnitude difference between the brightest two members. The resulting $\Psi(m_{12}| > N)$ is plotted with red squares in both panels of Fig. 6. The second data randomization performs better than the first at faithfully describing the relative brightness of the brightest group galaxies in rich groups. For this randomization, the observed $\Psi(m_{12}| > N = 9)$ differs from zero at a level of just $\simeq 1.7\sigma$, although $\Psi(m_{12}| > N = 2)$ is still $\simeq 4.1\sigma$ discrepant with zero. Because of the similarity of the second data randomization to the standard CLF formalism, these results suggest that the CLF adequately describes the arrangement of the brightest galaxies within rich groups, but may be an inadequate description of the luminosities of satellite galaxies in very poor groups.

There are several possibilities that may explain the difference of $\Psi(m_{12}| > N = 2)$ from zero. First, the satellite luminosity function in poor groups may require conditioning from an additional variable besides L_{cen}. Second, correlated draws from $\Phi_{cen}(L)$ and $\Phi_{sat}(L)$ may be required to describe accurately the observed magnitude gap abundance in low-richness groups. Third, this could be a manifestation of the difficulty of group identification for systems with just three or four group members. We will attempt to discern between these possibilities in a follow-up paper.

[**ARZ: I think it still remains an obvious omission in this context not to have some treatment of the randomization tests for the brightest galaxy distribution. I think you should consider seriously including that test in this section to make is feel more complete and well thought out. Get back to me on your thoughts in this regard.**] While we have focused entirely on magnitude gaps m_{12} in this section, it is also interesting to construct analogous statistics from alternative observables such as the luminosity function of the brightest group galaxy. Studying these and other such statistics constructed from a variety of data randomizations may yield new insight into the physics of galaxy group formation. We explore this promising technique more fully in a follow-up paper.

6 DISCUSSION

[**ARZ: I think you are a bit too cautious throughout this entire discussion section. If we should make specific statements, even if we refer to them as "tentative" conclusions. Otherwise, the results won't be noticed. We have specific, but perhaps tentative, statements that can be made about SHAM models and CLF models. Please try to rework this, along with my comments below, in order to have the discussion make specific statements.**]

We have used a volume-limited catalog of galaxy groups observed in SDSS DR7 to provide a number of new tests of the abundance matching prescription for connecting galaxies to dark matter halos. In § 4.1 we demonstrated that our fiducial mock galaxy catalog, constructed by abundance matching using $V_{\rm max}^{\rm acc}$, as the luminosity proxy for subhalos accurately reproduces the observed group multiplicity function $g(N)$, that is, the abundance of groups as a function of group richness. This constitutes a new success of SHAM that is distinct from previous tests that rely on measurements of galaxy clustering. Additionally, we showed that mock catalogs using $V_{\rm max}^{\rm peak}$ or $V_{\rm max}^{\rm z=0}$ as the abundance matching parameters incorrectly predict group multiplicity measurements and straddle the $g(N)$ predicted by $V_{\rm max}^{\rm acc}-$based catalogs (Figure 1). We have checked that this qualitative behavior holds true for models with very different (or without) scatter between luminosity and $V_{\rm max}$, as well as for volume-limited samples with different brightness thresholds, indicating that this a generic conclusion. This is particularly interesting in the context of a recent study by **?**, who showed that incorporating stellar mass loss into SHAM-based models of galaxy formation improves the predictions for galaxy clustering. Since a subhalo's $V_{\rm max}$ value at the time of accretion represents an intermediary stage of dark matter mass loss between $V_{\rm max}^{\rm peak}$ and $V_{\rm max}^{\rm z=0}$, our results point towards the possibility that $g(N)$ measurements may have the potential to constrain evolution of stellar mass (both new star formation and stellar mass loss) in satellite galaxies within group and cluster halos.

We have developed a novel implementation of SHAM that allows for the construction of mock galaxy catalogs with a luminosity function $\Phi(L)$ that exactly matches the $\Phi(L)$ exhibited by an observed galaxy sample, even when scatter between halo circular velocity and galaxy luminosity is included. We have exploited this implementation to test the ability of SHAM to predict the galaxy luminosity function conditioned on whether the galaxies in the sample are members of groups.

[**ARZ: The discussion in the next two paragraphs needs some work. It is too vague right now and I'm not sure what you want to get at. There are differences which we have become aware of only due to a private communication. One thing that we need to know (and perhaps to state here) is how closely Reddick et al. get the global LF. Could it be that their SHAM method doesn't match the global LF correctly and that this induces other problems? There are only a small set of possible sources for any differences. One could be the SHAM algorithm. Our method guarantees the correct, global LF. Theirs doesn't. Alternatively, group identification or some other such difference could be relevant.**] We find that field (group) galaxies in SHAM-based catalogs are systematically too dim (bright), and that this behavior holds true in all of the SHAM models we studied. This indicates that none of the popular SHAM-based models allocate galaxies to field and group environments correctly.

NEW PARAGRAPH HERE. In a study closely related to ours, **?** construct SHAM-based mock catalogs of galaxy groups and produced exhaustive constraints on SHAM-based models of the galaxy-halo connection, so it is useful to compare our results with theirs. We find somewhat different trends in $\Delta\Phi_{group}$ and $\Delta\Phi_{field}$ with luminosity than we report here. **?** find **... whatever it is they find. they also find this other thing (private communication, Rachel Reddick).** Furthermore, there exists a $\sim 4\sigma$ discrepancy between **?** best-fit prediction for $\Phi_{field}(L)$ and the observed field luminosity function using $V_{\max}^{\rm peak}$ (private communication, Rachel Reddick). It is possible that this discrepancy could, in part, induce the trends observed above.

There are few additional possibilities that may account for the differences between our results and those of **?** because we analyze the same N-body simulation, using the same halo-finder, and compare to the same galaxy sample. The primary differences between our methodologies lie in the details of our SHAM models and in our group-finding algorithms. If either of these two differences accounts for the different luminosity trends, it would have important consequences for the construction of mock galaxy catalogs that faithfully represent the properties of observed galaxies. This emphasizes a need for systematic and detailed examinations of the influences of mock catalog construction, group finding, and other methodological issues in order to understand the potential systematic differences induced by these choices.

One of the most common statistics used to quantify magnitude gaps is the *fossil fraction*, defined as the fraction of galaxy groups in a given sample with a magnitude gap $m_{12} \geq 2$. This is the statistic we present in the upper right panel of Figure 5. We have not adopted an X-ray brightness threshold criterion on our groups, nor have we imposed a maximum distance from the group centroid used to restrict the set of group members used in the measurement of the magnitude gap, and so a direct comparison between the fossil fraction we measure and the $8-20\%$ reported in **?** would not be meaningful. Among the existing results in the literature, the approach in **?** is most similar to ours: the authors employ their group-finding algorithm on a volume-limited, optical sample of galaxies and simply define m_{12} to be the difference in r-band magnitude between the brightest two group members. Thus their definition of a fossil is very similar to ours, and they quote fossil fractions for several different ranges of group mass, ranging from 0.5% for groups of mass $\sim 10^{14.5}M_{\odot}$ to $18-60\%$ for groups of mass $\sim 10^{13}M_{\odot}$. Unfortunately, the mass-binning of these values makes a direct, quantitative comparison impossible because, unlike the group-finding algorithm in **?**, our algorithm does not enforce the same assumptions about dark matter halo properties, and so groups found with our algorithm are not inextricably connected with a unique prediction for group mass. Nonetheless, the fossil fractions quoted in **?** do appear to be significantly higher than those we report in Fig. 5, which may be another sign of the sensitivity of group properties to the algorithms with which they are found.

Recently, **?** claimed that the low richness of the ten fossil groups they studied indicates a problem for the standard scenario of fossil group formation. In particular, **?** argued that the fact that their fossil groups are under-rich at all observed luminosities is difficult to understand within the standard ΛCDM theory of structure formation. Our results are a direct refutation of this claim. The magnitude gap abundance and fossil fraction exhibited by our fiducial mock catalog *constitutes* the simplest ΛCDM prediction for these statistics, and we have shown that these predictions are in good agreement with the data. In particular, we find, both in our mock catalogs and in the observed SDSS DR7 groups, that groups with large magnitude gaps are under-rich at all luminosities. We thus find no evidence that the abundance or properties of fossil systems presents a problem for the ΛCDM picture of structure formation.

Data randomization techniques similar to the ones we use in § 5 have been used previously in the literature. For example, **?**, **?**, and **?** all addressed the connection between richness and magnitude gap with Monte Carlo (MC) realizations of a group sample. These authors constructed a population of 10^4-10^6 Monte Carlo groups by drawing a fixed number of times from a global Schechter luminosity function to populate each MC group with a set of galaxies. In finding that the fraction of their MC groups with $m_{12} \geq 2$ was lower than that of the groups in their sample, they each concluded that fossil groups do not have a "statistical origin." The authors interpreted these exercises as evidence that fossil groups do not arise as extreme realizations of a Poisson process based on the global galaxy luminosity function, but through a dynamical process that preferentially eliminates satellite galaxies with large luminosities, namely mergers driven by dynamical friction.

More recently, **?** generalized this technique to construct a Monte Carlo realization of a group sample with the same multiplicity function as the observed group sample, finding that $\Phi(m_{12})$, the abundance of groups as a function of magnitude gap, is well-described by their Monte Carlo population. As discussed in § 5 and in Appendix B, we repeat the **?** analysis with an improved treatment of fiber collisions and reach the opposite conclusion.[3] [**ARZ: Do we actually disagree that such comparisons can be used as evidence of a dynamical origin? Part of the problem with making such a statement is that 'dynamical origin' can mean so many things. It could even mean simply that different things happen in groups from in the field. I would say that if we falsify the random-draw hypothesis, then it is false. Galaxy groups do not get their galaxies assigned luminosities as a result of random draws from the global luminosity function. I have commented out a few sentences here in order to rephrase this the way I would say it.**]

We have extended such comparisons by constructing the predictions for group properties in a simple, but specific, theory and tested the predictions of the theory. The cosmological simulation on which our mock catalog is based traces the evolution of a ΛCDM universe from the initial seeds of structure formation through to the present day, including the dynamical processes conventionally thought to determine the magnitude gap. The successful prediction for $\Phi(m_{12})$ of our fiducial mock catalog thus provides strong supporting evidence that the magnitude gap exhibited by galaxy groups is influenced by dynamical processes. On the other hand, tests based on the Monte Carlo realizations described above can only determine whether or not knowledge of the richness of groups together with a universal galaxy luminosity function provides sufficient information to predict the magnitude gap distribution. In § 5 we established that such information is insufficient. However, from this we can only conclude that knowledge of some other group property besides richness is required to predict the observed $\Phi(m_{12})$. This insufficiency does not reveal the origin of systems with a large gap.

[**ARZ: The use of this as a "test" of the CLF formalism here is not sufficiently well developed for most readers to understand it and needs to be reworked to be clear, and explicit. I think you have to state this very explicitly, for example... The CLF formalism assumes x, y, z. We construct data randomizations in which the central galaxy luminosity is fixed and the satellite luminosity distribution is determined only by the central luminosity. This design mimics the CLF formalism. However, notice that it is more general. We have not specified the form of the CLF. We are allowing the data to determine whether or not the underlying premises of the standard CLF formalism, namely that the satellite luminosities can be drawn from a distribution conditioned on the central galaxy luminosity, can be supported in detail by existing data. ... After some sort of exposition along those lines, you can continue. Play with this to make some clear, explicit statements.**] We have generalized these data randomization techniques to test an alternative hypothesis for the arrangement of the brightest galaxies into groups, testing the premises behind the CLF. These test demonstrate through explicit examples that randomization techniques can be quite useful despite our words of caution in the preceding paragraph. We have tested for the possibility that draws from $\Phi_{sat}(L)$ and $\Phi_{cen}(L)$ are correlated (Fig. **??**. We find no evidence for such correlations when restricting our group sample to those which have $N \geq 10$ members. However, when we include all groups with $N \geq 3$, members the data is poorly described by its randomization. The group multiplicity function is a rapidly-declining function of the number of group members, $g(N) \propto N^{-2.5}$, so systems with fewer members dominate the $\Phi(m_{12})$ measurement. Accordingly, the measurement on the higher richness threshold sample is subject to larger errors, though it is possible that this result could indicate that the luminosities of satellite and central galaxies are more strongly correlated in poor groups than they are in rich groups. [**ARZ: Although, if you look at the figure, probably not. It's just that for the poorer groups you have smaller error bars. So I added the last sentence above.**]

The fact that $\psi(m_{12}| > 2) \neq 0$ indicates that the distribution of satellite luminosities, $\Phi_{sat}(L)$ cannot be determined solely by the luminosity of the central galaxy. Additional information about the group is necessary in order to fix the distribution of satellite luminosities. This result contradicts the premise of the CLF (and notice we did not assume a functional form in order to arrive at this contradiction). Standard CLF models, which do not contain any such additional information, have been shown to reproduce a wide range of astronomical data. The gap abundance provides a *new* statistic with which to test galaxy-halo assignment models such as the CLF, and so it would be very interesting to determine what modifications to the standard CLF formalism are required in order to predict $\Phi(m_{12})$ accurately. Yet, before such a result could be established unequivocally, it would first need to be confirmed that $\Psi(m_{12}| > N = 2) \neq 0$ generally, *irrespective* of the details of the group-finding algorithm. We leave the exhaustive study of this issue as a task for future work.

7 CONCLUDING REMARKS

We have used the SDSS Mr19 catalog of galaxy groups to provide a series of new tests of models for the connection between galaxies and dark matter halos. We conclude this paper with a brief summary of our primary results.

(i) We have developed a novel implementation of SHAM that allows for the rapid construction of a mock galaxy catalog with a brightness distribution that *exactly matches* any desired luminosity function, *even after scatter has been included.*

(ii) Our fiducial SHAM model, based on abundance matching on $V_{\max}^{\rm acc}$ with 0.2dex of scatter at the faint end and 0.15dex at the bright end, accurately predicts group

[3] We note that when we run our analysis pipeline on their data set and adopt their fiber collision convention, we recover their results in full quantitative detail.

multiplicity, the abundance of groups as a function of richness, $g(N)$, a new success for the abundance matching prescription.

(iii) The $g(N)$ predictions based on SHAM models using $V_{\max}^{\rm peak}$ and $V_{\max}^{\rm z=0}$ do not match the observed group multiplicity function. In fact, these predictions straddle the $V_{\max}^{\rm acc}$ prediction, so measurements of group multiplicity may provide a promising avenue for constraining models of satellite mass stripping.

(iv) The group galaxy luminosity function $\Phi_{group}(L)$ and field galaxy luminosity function $\Phi_{field}(L)$ are predicted rather poorly by our mock catalogs, with SHAM group galaxies being systematically too bright and SHAM field galaxies systematically too dim. Since our all-galaxy luminosity function exactly matches that of the observed catalog by construction, this shortcoming must be due to an erroneous allocation of galaxies into group and field environments. We find this to be true in all of the variations of SHAM catalogs that we explored, suggesting that this is a generic weakness of the SHAM prescription.

(v) Our fiducial SHAM model, as well as models using $V_{\max}^{\rm peak}$ and $V_{\max}^{\rm z=0}$ with the same amount of scatter, accurately predicts the observed abundance of groups as a function of magnitude gap, $\Phi(m_{12})$, suggesting that the prediction for the relative brightnesses of galaxies in groups is a new success of the SHAM paradigm.

(vi) The gap abundance prediction is quite sensitive to the amount of scatter between luminosity and $V_{\max}$, suggesting that $\Phi(m_{12})$ measurements may be a new way to constrain the scatter in abundance matching.

(vii) The observed gap abundance is inconsistent with the hypothesis that the gap is determined by a set of random draws from a universal luminosity function, contradicting recent results of **XXXXXXXXXXXXX**.

(viii) The hypothesis that satellite galaxy brightnesses are drawn at random from $\Phi_{sat}(L|L_{cen})$ is inconsistent with the SDSS DR7 groups, at least for relatively low-richness groups. This contradicts one of the premises of the CLF formalism and indicates that draws from $\Phi_{sat}(L)$ and $\Phi_{cen}(L)$ are correlated in small groups, a possibility that we will pursue in future work.

ACKNOWLEDGMENTS

We have benefitted greatly from discussions with Nick Battaglia, Surhud More, Maya Newman, and ... APH and ARZ were supported in part by the Pittsburgh Particle physics, Astrophysics, and Cosmology Center (PITT PACC) at the University of Pittsburgh. APH ... your fermilab source of support goes here. The work of ARZ is also supported in part by the National Science Foundation through grant AST 1108802. AAB ... JAN...

APPENDIX A

In this Appendix we give a detailed account of our implementation of the abundance matching procedure to assign galaxies with r-band luminosities to dark matter halos. As discussed in § 3, a map from the maximum circular velocity of a halo, $V_{\max}$, to an r-band luminosity $M_{\rm r}$ is provided by the implicit relation given by Eq. 1. As we demonstrate in § 4, different choices for the abundance matching parameter (that is, $V_{\max}^{\rm z=0}$, $V_{\max}^{\rm acc}$, $V_{\max}^{\rm peak}$) result in mock galaxy catalogs with different properties, and so this choice has important consequences in the modeling of the galaxy-halo connection. We remind the reader that we denote mock catalogs constructed by abundance matching with $V_{\max}^{\rm z=0}$ as "SHAM0", with $V_{\max}^{\rm acc}$ as "SHAMacc", and with $V_{\max}^{\rm peak}$ as "SHAMpeak", and that the mock catalog referred to in the text as our fiducial model is a SHAMacc catalog. Since the novel features of our SHAM implementation method are the same regardless of this choice, throughout this Appendix we simply refer to the abundance matching parameter as $V_{\rm L}$.

Our SHAM procedure begins by using Eq. (1) to match the distribution of luminosities assigned to dark matter halos and subhalos to the double-Schechter function fit in **?**. We refer to these luminosities as $M_{\rm r}^{\rm init}$. SHAM models with scatter between $V_{\rm L}$ and $M_{\rm r}$ more successfully describe a variety of astronomical data (see **???**, and references therein) than models with no scatter. Accordingly, we introduce scatter as follows. For the $i^{\rm th}$ halo in the catalog, we assign an independently chosen random variable $\delta M_{\rm r}^i$ drawn from a Gaussian distribution of width σ^i. We use these random variables to assign new luminosities to the galaxies in the catalog via $M_{\rm r}^i \rightarrow M_{\rm r}^i + \delta M_{\rm r}^i$. In our fiducial catalog, we choose $\sigma^i = 0.5$ for all halos, which introduces roughly 0.2dex of scatter in the galaxy luminosities at the faint end of the luminosity function and 0.15dex at the bright end, which is very similar to the level of scatter used in **?**. We refer to these brightnesses as $M_{\rm r}^{\rm scatter}$.

Our goal is to construct a mock catalog with a luminosity function that exactly matches that of the Mr19 catalog, rather than the **?** luminosity function. To accomplish this, we rank-order all the halos and subhalos in the simulation by their luminosities $M_{\rm r}^{\rm scatter}$. Because of the scatter we have introduced, this ordering of the halos is non-monotonic in $V_{\rm L}$.

Rank-ordering the observed Mr19 galaxies by their luminosity naturally provides a map from cumulative number density $n_g(< M_{\rm r})$ to $M_{\rm r}$. We use this map to associate r-band magnitudes to halos in Bolshoi. The $i^{\rm th}$ halo in the list, ordered as described above, is assigned a rank-ordered cumulative number density $n_{\rm rank} \equiv i/V_{\rm Bolshoi}$, where $V_{\rm Bolshoi} = 250(h^{-1}{\rm Mpc})^3$ and i is the rank-order of the $i^{\rm th}$ halo. We use $n_{\rm rank}$ to assign luminosities to the halos by linear interpolation of the map from $n_g(< M_{\rm r})$ to $M_{\rm r}$. Halos with rank-ordered cumulative number densities larger than $n_g(< M_{\rm r} = -19)$ are discarded.[4] This procedure gives a luminosity function of the mock galaxies that *exactly matches* the Mr19 luminosity function, and which includes scatter in the mapping

[4] There are only two Bolshoi halos with rank-ordered cumulative number densities less than the value of n_g of the brightest Mr19 galaxy. These halos are not reassigned a new luminosity, but keep the $M_{\rm r}^{\rm scatter}$ value assigned to them by the initial (post-scatter) abundance match to the **?** luminosity function.

between $V_{\rm L}$ and $M_{\rm r}$. The reason for the initial abundance match to the **?** analytical fit is simply that the exact luminosity function of galaxies dimmer than $M_{\rm r} = 19$ that are located in the spatial region occupied by the Mr19 galaxies is not known.

APPENDIX B

Fiber collisions often occur when two or more galaxies are located within an angular separation of 55 arcseconds from one another. This angular separation is the minimum separation permitted by the width of the optical fibers used in the spectral measurements. When this occurs, the fiber is positioned to measure the spectrum of a randomly chosen galaxy from the two or more "fiber-collided" galaxies. In the data used to construct the DR3-based group catalog, the remaining galaxies of the fiber-collided set are assigned the redshift *and* brightness of the randomly-chosen galaxy. In the DR7-based group catalog that we utilize in this study, only the redshift of the randomly-chosen galaxy is assigned to the remaining fiber-collided galaxies. The absolute r-band magnitudes of the remaining fiber-collided galaxies are inferred from their apparent r-band magnitudes using the redshift of the randomly-chosen, spectroscopically-observed galaxy. This different treatment has important consequences for magnitude gap measurements, and in fact constitutes the primary reason for the difference between our conclusions and those drawn in **?**.

The **?** treatment of fiber collisions differs from ours in two important ways. First, we define m_{12} to be the r-band magnitude difference between the two brightest non-fiber collided members of the group, whereas **?** use fiber-collided galaxies in their m_{12} definition. Second, recall from § 2 that in the DR3 sample studied in **?**, fiber-collided galaxies are assigned the r-band magnitude of their nearest neighbor on the sky. As a consequence of this, if one uses fiber-collided galaxies to define m_{12} then all groups with a member that is fiber-collided with the brightest member are assigned magnitude gaps precisely equal to zero. This is a relatively common scenario since brightest group galaxies are typically found in the densest regions of the sky. As a result, with this treatment of fiber collisions combined with this definition of the magnitude gap, there is an artificial "spike" at $m_{12} = 0$. The authors in **?** attempted to account for this by discarding all groups with $m_{12} = 0$ when performing their KS test. However, in addition to incorrectly enhancing the abundance of $m_{12} = 0$ groups, this treatment of fiber collisions also results in an abundance of large- and moderate-gap groups that is systematically too low, since 100% of such groups with a galaxy that is fiber-collided with the brightest member are incorrectly removed from large- and moderate-gap bins and assigned $m_{12} = 0$. As one can see by the sign of $\Psi(m_{12}| > N)$ in Fig. 6, the sense of this systematic error is such that it improves the agreement between $\Phi(m_{12}| > N)$ and its Monte Carlo relative to our treatment.

The $m_{12} = 0$ spike and concomitant moderate- and lage-gap decrement does not occur in the DR7 treatment. *In our DR7 sample the gap abundance is the same regardless of whether not fiber-collided galaxies are used in the definition of* m_{12}*, whereas in DR3 the gap abundance is very sensitive to this choice.* Moreover, the DR3 gap abundance when fiber-collided galaxies are excluded is in excellent agreement with DR7 gap abundance, and so we conclude that the treatment of fiber collisions in **?** leads to incorrect measurements of $\Phi(m_{12})$ that result in the erroneous conclusion that the observed gap abundance is consistent with the random draw hypothesis.

Figure 6. Randomization tests of magnitude-gap statistics. In blue diamonds we plot $\Psi(m_{12}| > N)$ (see Eq. 3) observed in SDSS Mr19 when the data randomization procedure is to draw brightnesses for all the group members from the global $\Phi(L| > N)$. We show results for a richness threshold of $N > 2$ in the top panel, and $N > 9$ in the bottom panel. In red squares we plot $\Psi(m_{12}| > N)$ when the randomization procedure is to keep fixed the luminosity of the brightest group member and to randomly draw brightnesses for the remaining galaxies from $\Phi_{sat}(L|L_{cen})$. See text for further details about the data randomizations. Each of the $\Psi(m_{12}| > N)$ are distinct from zero, indicating that the distribution of magnitude gaps within groups is inconsistent with either of the simple random-draw hypotheses examined in these panels.