The Diversity of Environments around Luminous Quasars at Redshift $z \sim 6$

Keven Ren$^{1,2}$, Michele Trenti$^{1,2}$, Madeline A. Marshall$^{1,2,3}$, Tiziana Di Matteo$^{4}$, and Yueying Ni$^{4}$

$^1$School of Physics, The University of Melbourne, Parkville, VIC, Australia; kevenr@student.unimelb.edu.au
$^2$ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia
$^3$National Research Council of Canada, Herzberg Astronomy & Astrophysics, 5071 West Saanich Road, Victoria, BC V9E 2E7, Canada
$^4$McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 2020 October 12; revised 2021 June 8; accepted 2021 June 10; published 2021 August 23

Abstract

Significant clustering around the rarest luminous quasars is a feature predicted by dark matter theory combined with number density matching arguments. However, this expectation is not reflected by observations of quasars residing in a diverse range of environments. Here we assess the tension in the diverse clustering of visible $i$-band dropout galaxies around luminous $z \sim 6$ quasars. Our approach uses a simple empirical method to derive the median luminosity–to–halo mass relation, $L_{\text{med}}(M_{\text{h}})$, for both quasars and galaxies under the assumption of lognormal luminosity scatter, $\Sigma_{Q}$ and $\Sigma_{G}$. We show that higher $\Sigma_{Q}$ reduces the average halo mass hosting a quasar of a given luminosity, thus introducing at least a partial reversion to the mean in the number count distribution of nearby Lyman-break galaxies. We generate a large sample of mock Hubble Space Telescope fields of view centered across rare $z \sim 6$ quasars by resampling pencil beams traced through the dark matter component of the BlueTides cosmological simulation. We find that diverse quasar environments are expected for $\Sigma_{Q} > 0.4$, consistent with numerous observations and theoretical studies. However, we note that the average number of galaxies around the central quasar is primarily driven by galaxy evolutionary processes in neighboring halos, as embodied by our parameter $\Sigma_{G}$, instead of a difference in the large-scale structure around the central quasar host, embodied by $\Sigma_{Q}$.

We conclude that models with $\Sigma_{Q} > 0.3$ are consistent with current observational constraints on high-$z$ quasars, and that such a value is comparable to the scatter estimated from hydrodynamical simulations of galaxy formation.

Unified Astronomy Thesaurus concepts: Quasars (1319); Supermassive black holes (1663); Clustering (1908)

1. Introduction

The past two decades have yielded significant advances in the search for high-redshift quasars, or quasi-stellar objects (QSOs). The collective effort from a number of imaging surveys has classified $\gtrsim 10^2$ quasars beyond redshift $z = 6$ (from surveys such as the Sloan Digital Sky Survey (SDSS), Richards et al. 2002; Canada–France High-z Quasar Survey, Willott et al. 2010; Subaru-HSC programs, e.g., SHELLQs, Matsuoka et al. 2018), including two quasars at $z > 7.5$ (Bañados et al. 2017; Yang et al. 2020), when the universe was a mere fraction of its current age, around $\sim 700$ Myr old. The brightest quasars are some of the most luminous objects observed, with bolometric luminosities up to $L_{\text{bol}} \sim 10^{48}$ erg s$^{-1}$ or, equivalently, UV magnitudes at $M_{\text{UV}} \lesssim -29$. The masses of their central supermassive black holes (SMBHs) of order $10^9 M_{\odot}$ and their relative rarities at $\sim 1$ Gpc$^{-3}$ hint at a picture where these quasars reside within the most massive dark matter halos and are therefore tracers of extreme overdense regions of our universe (Springel et al. 2005). As such, high-redshift ($z > 6$) quasars are key objects to an important era of our cosmological timeline in the understanding of early structure formation.

The properties of these quasars are still quite uncertain, with numerous ongoing research questions such as their accretion history, seed formation, impact on galaxy formation, and SMBH growth mechanisms. The potential to shed light on these processes will therefore have a significant impact in constraining models of SMBH growth and galaxy evolution. Fundamentally, these processes are all connected to the halos they are hosted within, yet even in this regard, the properties of quasar halo hosts are still ambiguous. On one hand, theory/simulations have typically determined that the hosts have masses of the order of $\gtrsim 10^{12} M_{\odot}$ at $z \sim 6$ and hence are indicative of massive halos forming in rare ($>4\sigma$) peaks of the density field based on linear theory (Li et al. 2007; Costa et al. 2014; Feng et al. 2015; Marshall et al. 2020). These halos are generally expected to be biased, with an increased abundance of neighboring halos in their vicinity; thus, one should observe significant cross-correlation between the population of luminous quasars and galaxies in the field of view (Overzier et al. 2009; Romano-Diaz et al. 2011; Costa et al. 2014).

Observational studies of the cross-correlation of Lyman-break galaxies and Ly$\alpha$ emitters around luminous quasars at high redshifts have so far yielded no firm conclusion on the local clustering of quasars, with both overdensities (Kim et al. 2009; Utsunomiya et al. 2010; Hsu et al. 2013; Morselli et al. 2014; García-Vergara et al. 2017; Ota et al. 2018) and average underdensities (Stiavelli et al. 2005; Willott et al. 2005; Kim et al. 2009; Bañados et al. 2013; He et al. 2017; Kikuta et al. 2017; Mazzucchelli et al. 2017; Uchiyama et al. 2017; Champagne et al. 2018; Ota et al. 2018) being recorded. One possible factor responsible for the large variance in the environments is the challenge in making direct comparisons of neighbor counts across studies due to the difference in selection criteria applied on the observations of a limited number of quasar fields. Willott et al. (2005), using the Gemini telescope, found no evidence of an overdensity of $i$-band dropouts in the quasar field around $J1030+0524$ relative to the field. However, Stiavelli et al. (2005), observing the same object using the Advanced Camera for Surveys (ACS) from the Hubble Space Telescope (HST), found a number of missed galaxies fainter than the limiting magnitude used in Willott et al. (2005) and suggested that the quasar field is overdense, containing twice the excess of galaxy dropouts compared to the mean GOODS field. Furthermore, it can appear that the
overdensities are apparent on larger scales, beyond the typical field of view of an HST/ACS window. The reobservation of two average/underdense Kim et al. (2009) HST/ACS quasar fields in Morselli et al. (2014) with a larger field of view using the Large Binocular Camera (with a viewing area of ~600 arcmin² compared to HST/ACS’s ~11 arcmin²) finds that the observed fields are overdense at these larger scales. Additionally, a number of wider area surveys have shown a marked deficit of galaxies around the inner ~2.5 pMpc region of the quasar, hinting toward the possibility of significant radiative feedback from the active galactic nucleus (AGN) that quenches star formation in nearby galaxies (Utsumi et al. 2010; Morselli et al. 2014).

In a recent investigation of quasar environments, Habouzit et al. (2019) used the hydrodynamical simulation Horizon-AGN to infer the distribution of neighboring galaxies around massive SMBHs, albeit at lower redshifts of \( z \lesssim 5 \). In their work, they found a high degree of variance for counts for a single projected field of view, consistent with observations. Furthermore, an enhancement of number counts relative to the average field should be observable in the \( 2 \) cMpc radius around massive SMBHs, although it is contingent on probing to a sufficient sensitivity/depth of field. However, we should also note that the case is more extreme for \( z \sim 6 \) quasars, as the clustering bias is strongly halo mass–dependent. We can expect the bias to increase by a factor of \( \sim 1.5 \) (assuming a Sheth & Tormen 1999 bias function) between the rarest \( 10^{-9} \) Mpc\(^{-3} \) halos and the \( 10^{-6} \) Mpc\(^{-3} \) ones available in the Horizon-AGN simulation volume.

Additionally, Overzier et al. (2009) pointed out that simply placing the hosts in smaller halos can provide one such interpretation that is consistent with the observed clustering data. In a purely heuristic sense, the median host halo mass for a given quasar luminosity is driven by both the amount of scatter in quasar luminosities and the duty cycle (Veale et al. 2014; Ren et al. 2020). Conceptually, having scatter facilitates the odds that the brightest quasar/most massive SMBHs are luminous outliers hosted within relatively common dark matter halos with lower bias (Lauer et al. 2007). Similarly, knowledge of the quasar duty cycle, defined here as the proportion of SMBH accreting in the quasar mode at some cosmic time, is also crucial, as any single pointing at a luminous quasar must be during an instance where the quasar is active. Thus, a lower duty cycle also infers that quasars are generally contained in more common halos. Both parameters are poorly constrained, as their determination through measurements of clustering and number densities of quasars is challenging due to degeneracies (Wyithe & Loeb 2009; Conroy & White 2012; Ren et al. 2020).

In this paper, we examine the environments around SDSS-like \( z \sim 6 \) luminous quasars, taking advantage of the modeling framework we developed previously in Ren et al. (2018) in the context of studying the environment around luminous \( z > 8 \) galaxies. We simulate mock HST fields of view to determine the number of visible Lyman-break dropout galaxies around the most luminous quasars at \( z \sim 6 \). To achieve this, we utilize the large, high-resolution dark matter catalog underpinning the cosmological hydrodynamical simulation BlueTides and populate these halos with quasar and galaxy luminosities obtained from our semiempirical method. Additionally, object-to-object stochasticity in luminosities is included in the modeling through use of a conditional luminosity function (CLF) prescription. We investigate the available parameter space in both galaxy and quasar scatter and the quasar duty cycle that reproduces the available measurements of galaxy counts around quasar fields. This paper is outlined as follows. We describe the framework to construct our mock fields of view and the model used to populate halos with galaxies and quasars in Section 2. In Section 3, we analyze the diversity of environments, computing the average galaxy number counts visible in a field of view for various parameters of scatter. In particular, we evaluate the probability of reproducing the set of HST/ACS observations of Kim et al. (2009) in the same parameter space. In Section 4, we conclude with our key findings. For this work, we adopt the standard WMAP9 cosmological parameters (Hinshaw et al. 2013):
\[
h = 0.697, \quad \Omega_m = 0.2814, \quad \Omega_{\Lambda} = 0.7186, \quad \Omega_b = 0.0464, \quad \sigma_8 = 0.82, \quad \text{and} \quad n_s = 0.971.
\]
All magnitudes are given in the AB magnitude system (Oke & Gunn 1983).

2. Semianalytical Setup

The densities of luminous (\( M_{UV} \lesssim -26 \)) high-\( z \) quasars of order \( 1 \) Gpc\(^{-3} \) play a key role in the difficulty of making any robust statistical inference of their properties. Current state-of-the-art cosmological simulations with volumes comparable to \( 1 \) Gpc\(^3 \), such as the hydrodynamical simulation BlueTides (Feng et al. 2015) or the dark matter \( N \)-body simulation Millennium (Springel et al. 2005), are capable of capturing only a few of these rare quasars. These simulations are therefore susceptible to small number statistics in describing the population of these objects. For our model, we utilize an alternative approach and leverage the natural stochasticity present in quasar luminosities residing in halos to simulate a variety of unique environments without the computational cost of running additional \( N \)-body simulations. Stochasticity increases the variance for the range of halo masses luminous quasars typically reside in, thus effectively boosting the number of possible environments we can investigate. The small quasar clustering measurements at \( z \sim 4 \) (He et al. 2017) have already hinted at the possibility of significant scatter for luminous quasars; however, this possibility is also degenerate with quasars having a low duty cycle, which simply decreases the median halo mass hosting a quasar at fixed luminosity. For our analysis, both the quasar luminosity scatter \( \Sigma_G \) and quasar duty cycle \( \epsilon_{DC} \) are kept as free parameters.

Similarly, the luminous galaxies hosted by neighboring halos are also subject to stochasticity. At minimum, we can consider the variance in the assembly histories for halos to play a nonnegligible role in the variation in the star formation rate and, consequently, also the galaxy luminosity. Following this, we note that the collection of the most luminous galaxies, while generally expected to reside inside moderately massive halos, can be hosted within more common halos, facilitated by the inherent larger number density. The value of \( \Sigma_G \) at high redshifts (\( z > 6 \)) is poorly constrained observationally. An earlier study using clustering measurements of \( z \gtrsim 8 \) galaxies is unable to place limits of \( \Sigma_G \) owing to insufficient depth in the field of view (Ren et al. 2018). However, at lower redshifts, \( \Sigma_G \sim 0.17–0.23 \) at \( z \sim 0.1 \) and is constrained by the bright end shape of the galaxy luminosity function (LF; Yang et al. 2003; Cooray & Milosavljevic 2005). Simple theoretical modeling finds an approximate constraint for the lower limit, \( \Sigma_G \sim 0.2 \) for \( z > 2 \), which is inferred from the distribution of halo assembly times (Ren et al. 2018). Existing modeling of the \( z > 6 \) LF (Ren et al. 2019) and the inferred neutral hydrogen
fraction in the intergalactic medium at $z \sim 7$ (Whitler et al. 2020) assuming this lower limit of scatter have all yielded consistent results within observation limits. In contrast, simulations can have large inferred ranges in $0.2 < \Sigma_G < 0.6$ (additional details provided in Section 3.3). For brevity, we keep the galaxy luminosity scatter, $\Sigma_G$, as a free parameter in this work.

Our approach largely follows that of similar work done in Ren et al. (2018), investigating the environments of bright $z \sim 8$ galaxies. We run a Monte Carlo simulation generating mock fields of view centered around the brightest quasars. We use the dark matter halo catalog of the large cosmological hydrodynamical simulation BlueTides (Feng et al. 2015) and simultaneously populate resolved halos with both a quasar and a galaxy using a CLF prescription. The key input parameters for the CLF are the object’s median luminosity as a function of halo mass, $L_\text{e}(M_h)$; lognormal dispersion in the object’s luminosity given by a scatter parameter, $\Sigma$; and a duty cycle, $\varepsilon_{\text{DC}}$, for the quasar CLF.

For each Monte Carlo iteration, we trace multiple pencil beams across the brightest quasars inside the simulation volume with appropriate pencil beam dimensions to emulate the set of observations from Kim et al. (2009). To reduce the impact of selection bias, we sample for the top six brightest quasars at every iteration. Galaxies close in projection will be defined as neighbors of the bright object. Here we assume that only one object is visible per halo. Additionally, the galaxy host of the central quasar is not counted toward the overall number counts. The set of Monte Carlo iterations spans a log and high quasar duty cycle case, $\varepsilon_{\text{PC}} = 0.01$ and 1. The scatter parameter for galaxies and quasars will be systematically probed to both assess the impact on galaxy neighbor counts and facilitate a comparison with the Kim et al. (2009) observation set containing five quasar fields. The quasars selected in Kim et al. (2009) have magnitudes in the range $m_I = 19.83 - 20.63$ corresponding to number densities of $\sim 1 \text{ Gpc}^{-3}$ and are located at redshifts in the range $z = 5.99 - 6.40$ (see Section 3.3 for details). We provide additional information on the methods and tools used in the following subsections.

2.1. Simulation Parameters

For our halo catalog, we extract the dark matter component from the BlueTides simulation, a state-of-the-art cosmological hydrodynamical simulation for the first galaxies (Feng et al. 2015). BlueTides uses the smoothed particle hydrodynamics (SPH) code MP-GADGET with $2 \times 7040^{3}$ particles and tracks their evolution in a cosmological volume of $(400\ h^{-1}\ \text{Mpc})^{3}$ from $z = 99$ to 6.56. The corresponding mass resolutions for the dark matter and gas particles (in the initial condition) in BlueTides are $M_{\text{DM}} = 1.2 \times 10^5\ h^{-1}\ M_\odot$ and $M_{\text{gas}} = 2.4 \times 10^8\ h^{-1}\ M_\odot$, respectively. The star particles have a mass of $1/4M_{\text{gas}} = 6 \times 10^5\ h^{-1}\ M_\odot$. The gravitational softening length is 1.8 $\text{ckpc}$, which is indicative of its spatial resolution. Halos are identified in BlueTides with a friends-of-friends algorithm, using a linking length of $b = 0.2$ (Davis et al. 1985). This configuration results in BlueTides being able to resolve halos down to a mass of $\sim 1 \times 10^6\ M_\odot$. Finally, AGN feedback in BlueTides is modeled by injecting a fraction of the black hole (BH) radiation energy as thermal energy to gas particles in a region twice the radius of the SPH smoothing kernel of the BH particle. The large volume and high mass resolution have enabled BlueTides to conduct detailed studies of the first quasars/most massive BHs (Di Matteo et al. 2017; Tenneti et al. 2017; Ni et al. 2018; Tenneti et al. 2018) and the first galaxies (Waters et al. 2016a, 2016b; Wilkins et al. 2017). Additionally, BlueTides has been shown to be in full agreement with a number of observables, such as the UV galaxy LFs (Feng et al. 2015; Wilkins et al. 2017), galaxy clustering (Bhowmick et al. 2017), and the quasar luminosity function (QLF; Marshall et al. 2020; Ni et al. 2020).

For our analysis, we take the final snapshot available at $z = 6.56$ to compare with the $z \sim 6$ set of quasar observations from Kim et al. (2009) that vary between $z = 5.99$ and 6.40. In addition, since added stochasticity in quasars reduces the median halo mass hosting luminous quasars, this suggests that simulations with volumes smaller than 1 Gpc$^3$ can be used to sufficiently model statistics around these luminous quasars when given sufficient scatter. We describe in detail the impact of the smaller volume of BlueTides for the environments of quasars in Section 3.2.

In each Monte Carlo realization, pencil beams are traced across the six brightest quasars. Our pencil beams are parameterized with a cross-sectional area of $5.92 \times 5.92\ h^{-2}\ \text{Mpc}^2$ ($\sim 11.3\ \text{arcmin}^2$) and a depth of $244.5\ h^{-1}\ \text{Mpc}$, corresponding to a photometric uncertainty in redshift measurements, $\Delta z = 0.9$, centered at $z = 6.56$. We select these physical dimensions to be consistent with the observations of Kim et al. (2009), searching for $z \sim 6$, $i$-band dropouts with the HST’s ACS and a flux limit in the $z$ band of $z_{\text{AB}} < 26.5$. Each pencil beam is assigned with a random orientation defined by rotations around two distinct axes. As the size of the simulation volume is larger than the depth of the pencil beam, we have no risk of the pencil beam overlapping with itself.

2.2. Conditional Luminosity Function

We utilize an empirical CLF approach to assign halos with quasar and galaxy luminosities, $L_Q$ and $L_G$. The CLF, $\Phi(L|M_h)$, is a probabilistic description of measuring an object’s luminosity, $L$, given some halo mass, $M_h$. The CLF for quasars is defined by

$$\Phi(L|\mathcal{Q}(M_h)) = \left(1 - \varepsilon_{\text{DC}}\right)\delta(L = 0)$$

$$+ \frac{\varepsilon_{\text{DC}}}{\sqrt{2\pi}\Sigma_Q} \exp\left[-\frac{\left(L - L_{Q,\mathcal{Q}(M_h, \Sigma_Q)}\right)^2}{2\Sigma_Q^2}\right],$$

where $L_{Q,\mathcal{Q}(M_h)}$ is the median quasar luminosity as a function of halo mass $M_h$, $\Sigma_Q$ is the width of the dispersion in dex, $\varepsilon_{\text{DC}}$ is a constant quasar duty cycle defined as the fraction of SMBHs undergoing quasar-mode accretion, and $\delta(x)$ is the Dirac delta function. Similarly, we define the CLF for galaxies as

$$\Phi(L|\mathcal{G}(M_h)) = \frac{1}{\sqrt{2\pi}\Sigma_G} \exp\left[-\frac{\left(L - L_{G,\mathcal{G}(M_h, \Sigma_G)}\right)^2}{2\Sigma_G^2}\right],$$

with $L_{G,\mathcal{G}(M_h)}$, $\Sigma_G$ as the galaxy-equivalent variables for $L_{Q,\mathcal{Q}(M_h)}$, and $\Sigma_Q$. We assume a lognormal form for our CLF, as is standard in the literature for defining galaxy observables as a function of halo mass (luminosities: Cooray & Milosavljevic 2005; Yang et al. 2009; stellar masses: Behroozi et al. 2010; Moster et al. 2010; Reddick et al. 2013). The assumption of a lognormal functional form (to first order at least) is easily justified by...
invoking the multiplicative central limit theorem for the collective processes that link halo mass to luminosities. Additionally, the CLF is related to the LF by

\[
\phi(\log L) = \int_0^\infty \frac{dn}{dM_h} \Phi(\log L|M_h) dM_h, \tag{3}
\]

where \(\frac{dn}{dM_h}\) is the BlueTides halo mass function. We follow the simple procedure outlined in Ren et al. (2019) to derive the median luminosity versus halo mass relation, \(L_c(M_h, \Sigma)\), with enfolded stochasticity. The Ren et al. (2019) procedure, in short, is based on modifying \(L_c(M_h, \Sigma = 0)\), i.e., standard abundance matching without scatter, by a constant scaling factor and substituting a constant luminosity value past a characteristic threshold mass, \(L_c(M_h > M_h') = L_c(M_h')\), to achieve a well-fitting modeled LF using Equation (3). The physical interpretation of \(M_h'\) can be linked to the accretion of hot quasi-static gas around the host galaxy onto its SMBH, i.e., radio-mode AGN feedback, which is expected to be substantial for massive galaxies (Croton et al. 2006). For the abundance matching, we use the Bouwens et al. (2015) \(z \approx 7\) galaxy LF\(^5\) and the Matsuoka et al. (2018) \(z \approx 6\) quasar LF, which are subsequently redshift-evolved to \(z = 6.56\) to match the redshift of the BlueTides snapshot. Heuristically, simply using \(\Sigma > 0\) with the zero-scatter relation \(L_c(M_h, \Sigma = 0)\) will overestimate the bright end of the LF (Cooray & Milosavljević 2005). Thus, in Ren et al. (2020), we demonstrate that using a simple mass cutoff, \(M_h'\), offers a good approximation over standard deconvolution techniques for modeling the bright end of the LF under the influence of nonzero scatter. More importantly, the cutoff, \(M_h'\), has an intuitive interpretation for feedback in \(L_c(M_h)\) that evolves with \(\Sigma\), where a higher \(\Sigma\) requires self-regulation to occur at lower halo masses in order to preserve the LF. This is a feature that may otherwise be hidden due to “overfitting” from using a deconvolution method. The scaling factor and characteristic threshold parameters are calculated by \(\chi^2\) minimization to measurements of the LF, and the resulting \(L_c(M_h)\) is then normalized to the average number of galaxies detected in a random pointing at \(\Sigma = 0\). As our limiting constraint is in the \(z\) band, we convert our 1450 Å magnitudes to \(z\)-band magnitudes by calculating the UV slopes, \(\beta\), and assuming the galaxy UV continuum, \(f_\lambda \propto \lambda^\beta\), where \(\beta = -2 - 0.2(19.5 - M_{UV})\) (Bouwens et al. 2014).

We take a conservative lower limit of galaxy luminosity scatter, \(\Sigma_G = 0.1\), based on the analytical estimate of \(\Sigma_G \gtrsim 0.2\) from halo assembly processes (Ren et al. 2018). Similarly, we tentatively select \(\Sigma_G = 0.5\) as our upper limit, guided by the average \((\Sigma_G) \gtrsim 0.4\) from Meraxes (Ren et al. 2018). The lower limit of the quasar luminosity scatter, \(\Sigma_Q = 0.4\), is used to allow for a greater range of halo masses when selecting for luminous quasars, thus reducing the impact of selection bias. We find that \(\Sigma_Q \lesssim 0.3\) does not achieve enough variety in halo masses with our catalog, leading to a skewed distribution of galaxy neighbors for the \(\Sigma_Q = 0.3\) case due to the presence of a lower-density environment around a particularly massive halo.

---

\(^5\) The choice of the Bouwens et al. (2015) galaxy LF results in a higher number of \(m_\text{c} < 26.5\) galaxies compared to other \(z \sim 7\) galaxy LFs from Finkelstein et al. (2015) and Bowler et al. (2016) by \(\gtrsim 30\%\). Thus, using the Bouwens et al. (2015) LF remains a more conservative case, as it increases the likelihood of finding overdense environments.

We note that the effect of selection bias can be reduced using a larger simulation box or taking the aggregate result of multiple viewings. We adopt an upper limit of \(\Sigma_Q = 0.8\), as it becomes challenging to determine an \(L_Q(M_h)\) that adequately fits the knee in the Matsuoka et al. (2018) QLF for values that are significantly greater, \(\Sigma_Q > 0.7\). Thus, we vary our free parameters over the range \(\Sigma_G \in [0.1, 0.5]\) and \(\Sigma_Q \in [0.4, 0.8]\), while assuming values of \(\varepsilon_{\text{DC}} = 0.01\) and 1 for quasars. In Figure 1, we show the derived median luminosity against halo mass relations, \(L_c(M_h)\), with the limiting values of \(\Sigma\) for each object type. The resulting modeled LFs computed from \(L_c(M_h)\) using Equation (3) are consistent with observations within uncertainties.

3. Results

3.1. The Diversity of Environments around Luminous Quasars

Our model samples over the parameter space \(\Sigma_G \in [0.1, 0.5]\), \(\Sigma_Q \in [0.4, 0.8]\) and considers two duty cycles, \(\varepsilon_{\text{DC}} = 0.01\) and 1. For each point in the \((\Sigma_G, \Sigma_Q, \varepsilon_{\text{DC}})\) parameter space, we run a total of \(\sim 10^3\) Monte Carlo realizations. We follow the Kim et al. (2009) observation parameters using a magnitude limit \(m_\text{c} < 26.5\) with a pencil beam cross-sectional area matching the HSC/ACS field of view of 11.3 arcmin\(^2\). We collect the number of visible galaxies centered over the top six most luminous quasars in each Monte Carlo realization to reduce the impact of selection bias. With the targeted fields, we follow the same definition from Kim et al. (2009), where the number of galaxies visible in the quasar field does not include the host galaxy. In addition, for every targeted quasar pointing, we also collect the number of visible galaxies from a random field pointing in the same simulation volume; i.e., we collect six random fields per Monte Carlo realization. The average number of visible galaxies per pointing in the field throughout the simulation volume is \(\sim 3.7 \pm 2.6\) galaxies, calculated by averaging random pointings across all \(\Sigma_G\). This value is consistent with the number of galaxies expected in a GOODS ACS field, between three and eight galaxies (Kim et al. 2009). We find that the average number of visible galaxies per pointing is insensitive to \(\Sigma_G\), which is expected as we calibrate our LF at every \(\Sigma_G\). However, we find a small increase in the dispersion of galaxy counts from the random field pointings with \(\Sigma_G\) from 2.3 galaxies with \(\Sigma_G = 0.1\) to 2.9 galaxies with \(\Sigma_G = 0.5\).

In Figure 2, we show select snapshots demonstrating the diversity in galaxy neighbors around highly luminous quasars using our simple empirical approach (bottom panels). In addition, we also include analogs to our quasar fields as predicted by the full hydrodynamical suite BlueTides for comparison (top panels; see Feng et al. 2015; Di Matteo et al. 2017; Marshall et al. 2020 for details on the subgrid prescription used in BlueTides). For the BlueTides sample, we similarly trace pencil beams across the simulation volume and record halos captured inside the beam. We compute the galaxy luminosity by summing the luminosities from all star particles associated with a halo. The UV luminosity from each star particle was derived using the Binary Population And Spectral Population Synthesis models (BPASS v2.2; Stanway & Eldridge 2018) with the particle mass, age, and metallicity as inputs (Wilkins et al. 2017). We apply the photometric correction in a similar fashion to our empirical modeling, converting the magnitude from UV to the \(z\) band. Our approach
does not distinguish between satellite galaxies, as their luminosity contribution to the overall luminosity is small (one can estimate this contribution using the stellar masses of \( \varepsilon = 7.5 \) satellites and centrals from BlueTides; Bhowmick et al. 2018). Finally, we use the same magnitude limit of \( m_i < 26.5 \) to select for our galaxy neighbors. In BlueTides, we note a diversity of environments selected from the sample of the top six most massive SMBHs, with the neighbor count ranging from eight to 36 galaxies (three of which are shown in Figure 2) compared to the baseline value of 11.4 ± 6.2 galaxies from \( 2 \times 10^2 \) random pointing pencil beams. We note that the results given here are not corrected for dust; thus, both the number counts of neighbors and the magnitudes for the objects can be potentially overestimated (Marshall et al. 2020; Ni et al. 2020). Despite not being able to determine the extent of dust attenuation in our BlueTides quasar fields, the panels in Figure 2 still qualitatively show the diversity of environments with respect to both halo mass and central quasar luminosity. In BlueTides, there is a large variation between halo mass and quasar luminosity that arises as a consequence of SMBH masses being correlated with a low-tidal environment (Di Matteo et al. 2017; Huang et al. 2020; Ni et al. 2020). Phenomenologically, low-tidal fields have the capability to induce rapid growth of the central SMBH through direct radial accretion of cold gas from filaments. This correlation facilitates the large variation in quasar luminosities and subsequently can lead to a variety of densities around luminous quasar fields.

We also check for any correlation between quasar luminosity and the environment in BlueTides by individually comparing the set of nine most luminous quasars to other lower-luminosity quasars with similar halo masses. We define lower-luminosity quasars as being fainter by at least 3 mag than our target quasar and take similar mass halos from the range \( \Delta log(M_h) = 0.1 \) of the target. We account for projection variation by simulating \( 1 \times 10^2 \) randomly oriented pencil beams around each target luminous AGN candidate. We find that this variation can be significant, having a dispersion up to 50% of the average neighbor count around the quasar. Keep in mind that AGN feedback in BlueTides is not modeled by radiative transport, meaning it does not capture the possible effects of quasar radiation across megaparsec scales onto nearby central galaxies. Thus, any correlation would be an indication that the conditions that enable luminous quasars also facilitate star formation in nearby galaxies.

In Figure 3, we show reference single random pointings around the nine most luminous quasars, together with the distribution of the number of galaxy neighbors from the set of fainter quasars. We find that the quasar luminosity is not significantly correlated to the number of visible galaxy neighbors in the field of view. For example, as shown in the upper left panel, the set of fainter quasars inside similar mass halos to the most luminous quasar are considered to be overdense compared to the field average, with a count of \( \sim 20.0 \pm 8.3 \) galaxies to 11.4 galaxies. However, the most luminous quasar field containing \( 24.9 \pm 8.8 \) visible galaxy neighbors is well within 1\( \sigma \) of the sample average of the set of fainter AGN counterparts.

For the demonstration of our empirical model, we select the values \( \Sigma_G = 0.3, \Sigma_Q = 0.5, \) and \( \varepsilon_{DC} = 1 \). The different densities shown in Figure 2 are relative to the bright quasar pointings as opposed to the field, with the lower middle panel in Figure 2 representing the average environment of a bright quasar field. Both BlueTides and our empirical model are qualitatively consistent in that the majority of luminous quasars are typically hosted by massive halos and intrinsically lie in highly biased regions but also follow the relationship of having fewer luminous quasars in less dense environments. For the selected set of parameters, the average number of visible

---

**Figure 1.** Modeled median luminosity versus halo mass relation, \( L_i(M_h) \), for galaxies (top left) and quasars (bottom left), assuming minimum (blue curves) and maximum (red curves) scatter and \( \Sigma \) cases. For quasars, we also vary the duty cycle, \( \varepsilon_{DC} = 0.01 \) and 1 (dashed and solid). Here \( L_i(M_h) \) is calibrated to the LF, \( \phi(M_{UV}) \) (right), by iteratively solving for \( \phi(M_{UV}) \) in Equation (3). The range of calibrated LFs across \( \Sigma \) is given by the shaded region: galaxies (top right) and quasars (bottom right). The solid black points and dashed black curves are the observational data points plus best-fit lines, respectively, for Bouwens et al. (2015; galaxies, top right) and Matsuoka et al. (2018; quasars, bottom right). Note that some error bars in the galaxy LF are smaller than their data points.
Note that the BlueTides results are not dust-corrected; hence, the number counts and magnitudes computed may be overestimated. Both magnitude and halo mass of the central quasar are annotated at the top left.

Figure 2. Selected mock fields of view centered over a bright quasar showing the diverse quasar environments ordered by density from left to right. The top panels are outputs from the full hydrodynamical suite, BlueTides. The bottom panels are views from our empirical model with the parameters \( \Sigma_G = 0.3, \Sigma_Q = 0.5 \), and \( \varepsilon_{DC} = 1 \). Note that the BlueTides results are not dust-corrected; hence, the number counts and magnitudes computed may be overestimated. Both magnitude and halo mass of the central quasar are annotated at the top left (bottom right) of the top (bottom) panels. The colors show the luminosities of galaxies (solid lines) with the flux limit set at \( m_z < 26.5 \). The central quasar (red dashed line) has its z-band magnitude explicitly stated. The size of the objects is exaggerated to represent changes in log halo mass and is not an indication of the actual halo extent.

Additionally, we highlight in Figure 1 that \( \Sigma_G \) not only increases the variation of galaxy luminosities at a given halo mass but also sets the scale for radio-mode AGN feedback, as indicated by the mass where \( L_c/M_h \) flattens. For example, \( \Sigma_G = 0.3 \) points to a feedback scale corresponding to \( M_h \sim 10^{11.5} M_\odot \), in general agreement with the peak star formation efficiency, which occurs at \( M_h \sim 10^{12} M_\odot \), at high redshift (Tacchella et al. 2018; Behroozi et al. 2019). Seen this way, \( \Sigma_G \) offers two distinct mechanisms to explain an underdense observation: (1) a large variation in galaxy luminosities increases the variance of visible galaxy number counts in the field, boosting the odds of a serendipitous underdensity observation, and (2) neighboring galaxies are increasingly likely to be self-regulated, especially at high \( M_h \gtrsim 10^{11.5} M_\odot \).

This result eases the tension in models finding strong clustering conditions around quasars with assumed \( \varepsilon_{DC} \sim 1 \). These are typical outputs from using simpler models for populating halos with galaxies, such as assigning Lyman-break galaxies through a constant halo mass cut (Romano-Diaz et al. 2011; Buchner et al. 2019). In these instances, the use of a halo mass cut does not consider the event where the massive surrounding halos may be self-regulated, thus leading to an overestimation of visible galaxy neighbors.

3.2. Average Number of Galaxy Neighbors in the QSO Field

In Figure 4, we show the average number of galaxy neighbors visible in a single bright quasar field as a function of our free parameters \( \Sigma_G, \Sigma_Q, \) and \( \varepsilon_{DC} \). Perhaps unremarkably,
the average bright quasar field tends to be in excess of neighbors compared to a random field pointing (∼3.7 galaxies) ranging from an average of (∼4–20) neighbors for $\varepsilon_{DC} = 1$ and (∼3–12) neighbors for $\varepsilon_{DC} = 0.01$. Here the highest neighbor counts occur at the lowest $\Sigma_G$ values. However, the variance in each of these fields can also be significant, as mentioned in Section 3.1.

The two most prominent parameters that determine the extent of clustering are $\Sigma_G$ and $\varepsilon_{DC}$. It is relatively intuitive to understand how either of these parameters should impact clustering around quasars. As $L_{c,G}(M_h, \Sigma_G)$ is calibrated to the LF, having $\Sigma_G > 0$ effectively redistributes the population of galaxies at a fixed luminosity across an increasing range of halo masses. This decreases the probability for massive halos with similarly massive neighbors to contain luminous galaxies and reduces the overall number of galaxies in the average quasar field of view. On the contrary, the duty cycle operates independently to $\Sigma_G$ by assuming a constant probability, $\varepsilon_{DC}$, for a halo to contain an active SMBH. This reduces the nominal halo mass for the most luminous quasars, implying generally less clustered environments. Finally, the effect of the quasar scatter, $\Sigma_Q$, should be degenerate with $\varepsilon_{DC}$, where both of these parameters would alter the distribution of host halo masses for quasars of fixed luminosity.

Curiously, the average galaxy number counts centered on luminous quasars appear to only show weak to no dependence...
in \( \Sigma_Q \) across the entire parameter space. We explore this discrepancy by deriving the distribution of the linear bias factor for the halos containing our brightest sources in the simulation,

\[
p(b_{\text{brightest}}) \propto \int \frac{dn}{dM_h} (M_h(b)) \Phi (\log L|M_h) p(L_{\text{brightest}}) d \log L,
\]

where \( M_h(b) \) is the inversion of the Sheth & Tormen (1999) halo mass bias relation, \( \Phi (\log L|M_h) \) is our usual CLF, and \( p(L_{\text{brightest}}) \) is the distribution of luminosities for the brightest object, given by the following relation:

\[
p(L_{\text{brightest}}) \propto \frac{d}{dL_{\text{brightest}}} \left( \int_{L_{\text{min}}}^{L_{\text{max}}} \phi(L) dL \right)^n(V).
\]

Here \( \phi(L) \) is the QLF, and \( n(V) \) is the expected number of objects in a volume \( V \) between the range of magnitudes \( (L_{\text{min}}, L_{\text{max}}) \),

\[
n(V) \approx \int_{L_{\text{min}}}^{L_{\text{max}}} V \phi(L) dL.
\]

We take our limits \( (L_{\text{min}}, L_{\text{max}}) \) to be \( M_{UV} = (-20, -30) \), respectively.

In Figure 5, we show the probability distributions for the bias of halos hosting our brightest quasars at our limiting \( \Sigma_Q \) values and the magnitude of the brightest object in a cosmic volume, \( V \). The duty cycle is taken to be \( \varepsilon_{DC} = 1 \), as changes in duty cycle do not have any significant impact on the spread of bias for a quasar at a given luminosity (Ren et al. 2020). Note that the sharp peak is an artifact from our modeling method by assuming a sudden cutoff in the median quasar luminosity versus halo mass relation (Figure 1). We find that for a simulation volume similar to that of BlueTides \( (400 h^{-1} \text{Mpc})^3 \), our analysis should measure only a small dependence on \( \Sigma_Q \) for clustering around the brightest quasar. In the parameter space where \( \Sigma_Q \) is low, the brightest quasar should still have a significant probability of being hosted inside halos with \( M_h < M_c \), where \( M_c \) is the characteristic cutoff in \( L_c(M_h, \Sigma_Q) \). Thus, the instance where we do not see tangible differences between \( \Sigma_Q \) values is just a limitation of having an insufficient simulation volume to fully capture \( M_h > M_c \) halos. Additionally, we would also expect any measured trends to be further diluted, as our analysis uses the average of the visible galaxy counts around the top six most luminous quasars, rather than the single brightest quasar. We also show the expected distribution of bias around the brightest quasar assuming sufficient volume in the bottom panels of Figure 5.

---

**Figure 4.** Average number of visible galaxies \((m_z < 26.5)\) in a quasar field as a function of scatter parameters \( \Sigma_G \) and \( \Sigma_Q \) assuming duty cycle values \( \varepsilon_{DC} = 1 \) (left) and 0.01 (right). As with Kim et al. (2009), the galaxy neighbor count excludes the central galaxy that contains our bright quasar.

**Figure 5.** Modeling the probability distribution of the bias for the brightest object, \( p(b_{\text{brightest}}) \), inside a simulation volume, \( V_{\text{tot}} \), of \( (400 h^{-1} \text{Mpc})^3 \), corresponding to the catalog used in this work, BlueTides (upper left), and a hypothetical N-body simulation with a larger volume of \( (1500 h^{-1} \text{Mpc})^3 \) (lower left). For each of these volumes, we look at the two limiting cases of \( \Sigma_Q = 0.4 \) (solid blue) and 0.8 (solid red). The panels on the right show the distribution of magnitudes for the brightest quasar inside their respective volumes (BlueTides, upper right; larger simulation, lower right) by solving Equation (5).
We do not expect moving to a larger simulation to have any noticeable qualitative impact given the overlap in the bias distribution even with a volume such as (1.5 Gpc)$^3$. However, if we were to still repeat our Monte Carlo analysis using the larger volume (1.5 Gpc)$^3$, then it would be easier to infer a small dependency in $\Sigma_Q$ for clustering around the brightest quasars compared to this analysis using a volume of $(400 \ h^{-1} \text{ Mpc})^3$.

Another interesting aspect from Figure 5 is that variance in the bias, or the diversity of environments around the brightest quasar, is larger for smaller $\Sigma_Q$. Intuitively, we would expect the opposite effect, as a larger $\Sigma_Q$ implies a wider range of luminosities at a given $M_h$. The rationale is that this effect is offset by the combination of a decrease in flattening threshold mass $M_h$ as $\Sigma_Q$ increases in combination with the strong dependence on $M_h$ for the bias. Additionally, the figure also shows how $\Sigma_Q$ impacts the luminosity-dependent nature of clustering. We see that the distribution of bias for the brightest object does not notably change, even as the volume is increased for the highest values of $\Sigma_Q$. This implies that the clustering around quasars brighter than the quasar luminosity $L_q$, where $\phi_q(L) = 1/V$, is insensitive to luminosity. Here $V$ is defined as the smallest volume such that $p(\text{bias}_{\text{brightest}})$ does not change with increasing $V$. Conversely, we see an evolution in the bias around the brightest quasars for low values of $\Sigma_Q$ when increasing volume. This indicates that there is some degree of luminosity-based clustering until we reach this limiting luminosity. We can understand this in the context of our introduced flattening from the radio-mode AGN feedback in our modeling of the median luminosity versus halo mass relation, where the median luminosity is no longer dependent on halo mass past $M_h > M_h^c$.

### 3.3. Comparison with the Observations of Kim et al.

Kim et al. (2009) presented a sample of five quasar fields at $z \sim 6$, including a single field from Stiavelli et al. (2005), analyzing the number of i-band dropout galaxies (limiting magnitude of $z_{\text{AB}} < 26.5$) in the vicinity of the central quasar. The redshift range of the quasars spans $z = 5.99-6.40$, slightly lower than the redshift $z = 6.56$ used in this study. The magnitude of the central quasar spans $z_{\text{AB}} = 19.98-20.63$, corresponding to a number density of $\sim 10^{-9} \text{ Mpc}^{-3}$. The quasar fields examined by Kim et al. (2009) find a large variance in galaxy number counts, yielding a total of two underdense fields (relative to the normalized average number count in the GOODS survey), two overdense fields, and one field with an average density. A number of simulations have attempted to derive a theoretical understanding behind the variation in these number counts around massive SMBHs at high-$z$ (Romano-Diaz et al. 2011; Costa et al. 2014; Habouzit et al. 2019). A common thread between these simulations is that the environments around the most massive SMBHs are overdense relative to the average counts in the field. Critically, this finding persists even with a smaller-scale simulation (100 Mpc)$^3$ at lower redshift, $z \sim 5$, implying that clustering should be further enhanced with larger masses and at higher $z$. Additionally, both Costa et al. (2014) and Habouzit et al. (2019) also reported significant variations in the fields centered around massive SMBHs. In Figures 4 and 5, we show that the modeling described here is qualitatively consistent with the results of these simulations in both density and variance of fields.

While these results have generally found that the local neighborhoods of massive SMBHs tend to be clustered, we arrive at a source of tension for the multiple observational accounts of underdense fields around these luminous quasars. Various attributed physical explanations can include having a variable galaxy duty cycle (Romano-Diaz et al. 2011), strong galactic winds from supernovae (Costa et al. 2014), or radiative, quasar-mode feedback from surrounding AGNs across megaparsec scales (Habouzit et al. 2019). Another plausible explanation is that the Kim et al. (2009) fields are too shallow, and the Poisson noise becomes comparable to the signal; hence, deeper imaging may be required in order to robustly claim a lack of clustering signal for these underdense fields. However, there is still scope in using these shallow number counts to inform the range of our scatter parameters $\Sigma_Q$, $\Sigma_G$ and see if scatter presents a viable explanation in relieving the tension in these findings.

We assess this tension by thoroughly investigating the possibility of claiming multiple detections of $>3$ overdense fields (equivalent to the complement of finding $>2$ underdense fields) over a quintuple set of images. For brevity, we remove the average density classification and label the fields as either overdense or underdense. Here overdense is defined as having a galaxy number count in excess compared to the average random pointng of $\sim 3.7$ galaxies (i.e., four or more galaxies), while underdense is simply defined as the inverse. We group the individual quasar pointings into sets of five and derive the probability of replicating the Kim et al. (2009) sample after assuming values of $\Sigma_G$ and $\Sigma_Q$. On a fundamental level, this probability depends on both the average number and the variance in galaxy number counts in the quasar pointings. This Monte Carlo approach complements previous studies of investigating the environments by having the capacity to generate a sample size large enough for statistical inferences. In Figure 6, we show the confidence level of obtaining $>3$ overdense fields out of a set of five. As expected, the tension in having multiple underdense fields in a set of five fields can be high. With our nominal assumption of maximal duty cycle, $\varepsilon_{\text{DC}} = 1$, the $2\sigma$ contour (corresponding to an $\sim 5\%$ probability of matching the Kim et al. 2009 results) is restricted to the curve $\Sigma_G \sim 0.3$ and lies relatively independent of $\Sigma_Q$. On the contrary, a lower duty cycle, $\varepsilon_{\text{DC}} = 0.01$, opens up the parameter space and implies that it is possible that such a scenario is plausible for all values of $\Sigma_G$.

Similar to the results of Figure 4, the most prominent parameters that determine the likelihood of replicating the observations of Kim et al. (2009) are $\Sigma_G$ and $\varepsilon_{\text{DC}}$. While our model shows that setting a low duty cycle presents a suitable explanation, the determination of $\varepsilon_{\text{DC}}$ at high redshifts is poorly constrained in practice and remains an open question of research. In a heuristic sense, a low duty cycle is theoretically constrained by observations of $10^9 \ M_{\odot}$ SMBHs by $z \sim 7$, which effectively places further constraints on the formation modes of seeds and early accretion processes. For example, hydrodynamical simulations (BlueTides, Appendix A; Illustris, DeGraf & Sijacki 2016) and theoretical modeling (Aversa et al. 2015) have inferred duty cycles of order unity for $z > 6$ SMBHs. However, observational determinations of the duty cycle at high redshifts $z > 4$, typically using quasar–galaxy cross-correlation measurements, have alluded to a wide range of duty cycles in the range $1 \times 10^{-3} \ < \ varepsilon_{\text{DC}} < 1$ (Shen et al. 2007; Shankar et al. 2010a, 2010b; He et al. 2017). Addressing these
tensions on \( \epsilon_{DC} \) is beyond the scope of this study. However, we will note that derivations of the duty cycle based on luminosity measurements can be highly sensitive to obscuration effects from both the quasar and galaxy (Chen & Gnedin 2018; Trebitsch et al. 2019) or even simply from the definition of the minimum luminosity that constitutes an “on” quasar (DeGraf & Sijacki 2016). In a similar way, the measurements of the duty cycle can also be potentially underestimated if there is a population of feedback-affected galaxies that are unseen but exist at fainter magnitudes, as detailed in Section 3.2. Taking the uncertainty in low duty cycle scenarios, we investigate if a cohesive picture can be built upon the worst-case assumption of maximal clustering with a duty cycle of order unity.

The natural question would be to inquire about the current constraints in \( \Sigma_G \). An estimated theoretical lower limit at \( z > 6 \) is \( \Sigma_G \sim 0.2 \) (Ren et al. 2018), derived from the variance in dark matter halo assembly times. More sophisticated modeling has yielded various estimates of galaxy luminosity scatter at \( z > 6 \), all higher than the derived lower limit: \( \Sigma_G \sim 0.32 \) (BlueTides, hydrodynamical; Appendix B), \( \sim 0.38–0.58 \) (Meraxes, semi-analytical; Mutch et al. 2016), >0.2 (\( z \)-independent galaxy evolution model, empirical; Tacchella et al. 2018), >0.25 (UniverseMachine, empirical; Behroozi et al. 2019), and >0.3 (IllustrisTNG, hydrodynamical; Vogelsberger et al. 2020). The latter three constraints are considered lower limits in \( \Sigma_G \) as they are the scatter derived from the stellar mass–to–halo mass and UV luminosity–to–stellar mass relation, respectively. Here we just assume the existence of some additional processes that link their listed scatter to \( \Sigma_G \) to be summed in quadrature. We can see that on face value, most of the galaxy scatter measured by simulations has largely not excluded the Kim et al. (2009) observations to a 2\( \sigma \) level. However, it is also worthwhile to note that our result shows a surprising sensitivity in \( \Sigma_G \) to the likelihood of finding multiple underdense fields. In the nominal case, \( \epsilon_{DC} = 1 \), we see that only a minor boost of \( \Delta \Sigma_G \sim 0.05 \) from \( \Sigma_G = 0.3 \) to 0.35 is needed for a significant increase in overall probability of \( \sim 5\%–32\% \). While tight, hydrodynamical simulations tend to hover close around the value \( \Sigma_G > 0.3 \); hence, it is possible that any observed underdensities in terms of galaxy neighbors fall completely within the expectations.

4. Conclusion and General Remarks

In this paper, we revisit the existing tension in the diversity of environments around \( z \sim 6 \) quasars between theory/simulations predicting high clustering and observations with measurements of weak/average clustering. We approach this problem with an empirical method, populating halos in a high-resolution \( N \)-body dark matter–only simulation from BlueTides at \( z = 6.56 \) with quasar and galaxy luminosities. The simulation’s box volume of \( (400 \, h^{-1} \, \text{Mpc})^3 \) is sufficiently large to track the rare host halos that could plausibly host SDSS-like quasars. The relations between halo mass and the object’s luminosity are calibrated to their LFs. We then create mock field-of-view images by tracing pencil beams through luminous quasars and recording the number of visible galaxies within a \( 5.92 \times 5.92 \, h^{-2} \, \text{Mpc}^2 \) area plus a depth corresponding to the photometric uncertainty \( \Delta z \sim 0.9 \) at \( 244.5 \, h^{-1} \, \text{Mpc} \). Specifically, we explore the possibility of scatter when populating galaxy or quasar luminosities \( (\Sigma_G, \Sigma_Q) \) as a source to alleviate tensions between modeling and the specific observations of Kim et al. (2009). In addition to scatter, we also investigate the impact of different values of a constant quasar duty cycle, \( \epsilon_{DC} \), defined here as the relative proportion of BHs actively undergoing quasar-mode accretion.

The innovation of our work comes from leveraging stochasticity to generate a large sample of distinct fields. Using a Monte Carlo method, we resample galaxy and quasar luminosities in our simulation volume over \( \gtrsim 10^7 \) times for every point in parameter space. Scatter reduces the average halo mass hosting an object of given luminosity by increasing the probability for common halos to accommodate a quasar with an outlier accretion rate; thus, the number of independent halo conditions accessed increases exponentially for the rarest,
This method allows rapid exploration of the parameter space \( (\Sigma_G, \Sigma_Q) \) that is relatively statistically robust at minimal cost, hence maintaining a competitive advantage over traditional modeling methods.

We summarize our key results as follows.

1. We show that for \( \varepsilon_{DC} = 1 \), our model converges to the interpretation that rare luminous quasars reside in overdense halo environments, with hosts having a halo bias of order \( b \sim 10^3 \) (Figure 5). Furthermore, the high halo bias value persists irrespective of how high we set \( \Sigma_Q \). However, we find that even with similar halo biases, the galaxy number counts from field to field can still vary significantly (Figure 2). Fundamentally, this suggests that the diversity in clustering of luminous galaxies around a quasar host is driven by galaxy evolution processes in neighboring halos instead of the large-scale structure phenomenology of the underlying density field. This is consistent with the findings of Costa et al. (2014), also alluding that galaxy feedback processes play a major role.

2. In our modeling, the extent of visible galaxy clustering is sensitive to \( \Sigma_G \) and \( \varepsilon_{DC} \). The variation in galaxy luminosities set by \( \Sigma_G \) directly increases the variance of visible galaxy counts in a quasar field. Here \( \Sigma_G \) also implicitly sets a scale for radio-mode AGN feedback where massive neighboring halos are increasingly likely to have self-regulated star formation, thus facilitating the odds for a underdense quasar field. Similarly, the impact of changing \( \varepsilon_{DC} \) would alter the distribution of halo masses that hold quasars. Decreasing \( \varepsilon_{DC} \) reduces the probability for an SMBH to be accreting in any instance, and would skew the most luminous quasars toward residing within the less dense common halos.

3. Setting \( \varepsilon_{DC} = 0.01 \) potentially relieves the tension in the weak clustering around quasar fields; however, this can be difficult to justify on a theoretical basis of achieving \( 10^9 \, M_\odot \) SMBHs by \( z \sim 7 \). In addition, \( \varepsilon_{DC} \) can be difficult to constrain, as it is sensitive to obscuration effects potentially underestimating the value (Chen & Gnedin 2018; Trebitsch et al. 2019). Here \( \Sigma_G \) adds to this by suggesting an unaccounted-for population of feedback-affected galaxies beyond the depth of the observation window.

4. We also see that \( \Sigma_G \) has a small/no impact on the clustering of bright objects. We find that the reason is related to the simulation volume used (Figure 5). For low \( \Sigma_G \), the BlueTides N-body simulation did not have the necessary volume to probe enough halos past \( M_h > M_h^\star \). However, we expect the overall impact from this effect to be fairly small, even with sufficient volume, as there is significant overlap in the bias distribution across the range of \( \Sigma_G \).

5. We find that clustering eventually becomes independent of quasar luminosity (Figure 5). This is a natural consequence of assuming some feedback scale in \( L_Q(M_h) \). In our model, the median quasar luminosity does not change past \( M_h > M_h^\star \). As the number density of massive halos drops exponentially, we expect the most luminous quasars to tend toward having a bias value, \( b(M_h^\star) \).

6. In Figures 4 and 5, we show that our results are in agreement with existing simulations finding quasar environments to be generally overdense with a high degree of variance in galaxy counts (Romano-Diaz et al. 2011; Costa et al. 2014; Habouzit et al. 2019). However, there is a marginal difficulty for these simulations to reproduce the underdense fields of Kim et al. (2009). Justifications include invoking a low Lyman-break galaxy duty cycle (Romano-Diaz et al. 2011), strong galactic winds (Costa et al. 2014), or, tentatively, radiative feedback from neighboring quasars (Habouzit et al. 2019).

7. Our model requires \( \Sigma_G \sim 0.3 \) to remain consistent with the observations of Kim et al. (2009; Figure 6). Additionally, current state-of-the-art models of high-redshift galaxy evolution (e.g., Mutch et al. 2016; Tacchella et al. 2018; Behroozi et al. 2019; Vogelsberger et al. 2020 and BlueTides) all tend to be close to this value of \( \Sigma_G \), suggesting that the reported underdense fields are not necessarily unlikely events.

8. One limitation to note is that the independent placement of galaxies and quasars in halos neglects any large-scale (two-halo) environmental effects. Currently, there is limited evidence that suggests that UV quasar radiation can be intense enough to suppress star formation in nearby central galaxies, albeit in relatively small halos, \( M_h < 10^9 \, M_\odot \) (Kashikawa et al. 2007). Habouzit et al. (2019) suggested that the quasar-mode radiation from the AGN is capable of ionizing neutral hydrogen gas up to \( \sim 10 \, c \text{Mpc} \). However, they contended that more detailed radiative transfer simulations are required to measure the significance of this type of feedback on galaxy growth. In the context of our model, any additional feedback would further serve as another source to relieve the tension between the observations and our modeling.

This research was conducted by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D) through project No. CE170100013. K.R. and M.M. are additionally supported through the Research Training Program Scholarship from the Australian Government. K.R. acknowledges support through the Postgraduate Writing-up Award granted by the David Bay Fund. M.M. acknowledges support from the National Research Council of Canada’s Plaskett Fellowship.

### Appendix A

**Quasar Duty Cycle in BlueTides**

The BHs in BlueTides use the subgrid model developed for its predecessors, MassiveBlack I and II (Matteo et al. 2005; Springel et al. 2005), with additional modifications consistent with Illustris (DeGraf & Sijacki 2016). The BHs are initially seeded with a mass of \( M_\bullet = 5 \times 10^6 \, h^{-1} \, M_\odot \) in halos more massive than \( M_h = 5 \times 10^{10} \, h^{-1} \, M_\odot \). In BlueTides, BHs accrete gas via Bondi–Hoyle accretion,

\[
M_\bullet = \frac{4\pi \alpha G^2 M_h^2 \rho}{\rho (c_s^2 + v^2)^{3/2}},
\]

where \( \alpha \) is a dimensionless parameter, \( G \) is the gravitational constant, \( \rho \) is the local density of gas, \( c_s \) is the local sound speed, and \( v \) is the relative velocity between the BH and the nearby gas. BlueTides also allows for super-Eddington accretion, with a maximum permitted rate of two times the...
Eddington limit,

\[ M_{\text{Edd}} = \frac{4\pi G M_p m_p}{\eta \sigma_T c}, \]

where \( m_p \) is the proton mass, \( \sigma_T \) is the Thomson cross-section of an electron, and \( \eta = 0.1 \) is the radiative efficiency. The conversion from accretion rate to bolometric luminosity, \( L_{\text{bol}} \), is given as

\[ L_{\text{bol}} = \eta \dot{M} c^2, \]

where \( c \) is the speed of light. In Figure A1, we show the distribution of SMBH masses and their AGN luminosities in BlueTides at \( z = 6.56 \). We select for established BHs using the criterion \( M_e > 10^{6.5} M_\odot \) to avoid any possible uncertainty originating from the BlueTides seeding prescription. For the quasar duty cycle, we follow the definition provided in DeGraf & Sijacki (2016) and compute the fraction of quasars that exceed some selected luminosity threshold. The calculated quasar duty cycle is \( \varepsilon_{\text{DC}} \sim 1 \) for \( L_{\text{bol}} > 10^{44} \text{erg s}^{-1} \), consistent with the projection for the \( z > 6 \) duty cycle of DeGraf & Sijacki (2016) using Illustris. Additionally, 86% (53%) of BHs in BlueTides have AGN luminosities brighter than \( L_{\text{bol}} > 10^{44.3-44.45} \text{erg s}^{-1} \), corresponding to 50% (75%) of the Eddington limit for an \( M_e = 10^{6.5} M_\odot \) BH.

Appendix B
Galaxy Luminosity Scatter in BlueTides

In Figure B1, we show the scatter in galaxy luminosities, \( \Sigma_G \), in various halo mass bins of size \( \Delta \log(M_h) = 0.12 \) in BlueTides at \( z = 6.56 \). The final mass bin contains all halos \( M_h > 10^{12} M_\odot \), and is selected in a way such that the number of objects is comparable in size to the mass bin before it. The galaxies in BlueTides are not corrected for dust; hence, the scatter values are underestimated, as the inclusion of dust will add an additional degree of variability. In BlueTides, we find that the mean luminosity scatter across all bins is \( \langle \Sigma_G \rangle \sim 0.32 \), which is comparable in magnitude to the hydrodynamical simulation IllustrisTNG at the same redshift, with \( \langle \Sigma_G \rangle > 0.3 \) (estimated from the scatter in the UV luminosity versus stellar mass relation; Vogelsberger et al. 2020). The trend of decreasing \( \Sigma_G \) as \( M_h \) increases is consistent with both IllustrisTNG and the semianalytical model Meraxes for \( z \approx 8 \) galaxies (Ren et al. 2018).

ORCID iDs
Keven Ren @ https://orcid.org/0000-0001-7347-5953
Michele Trenti @ https://orcid.org/0000-0001-9391-305X
Madeline A. Marshall @ https://orcid.org/0000-0001-6434-7845
Tiziana Di Matteo @ https://orcid.org/0000-0002-6462-5734

References
Aversa, R., Lapì, A., de Zotti, G., Shankar, F., & Danese, L. 2015, ApJ, 810, 74
Bahados, E., Venemans, B., Walter, F., et al. 2013, ApJ, 773, 178
Bahados, E., Venemans, B. P., Mazzucchelli, C., et al. 2017, Natur, 553, 473
Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143
Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379
Bhowmick, A. K., Campbell, D., Matteo, T. D., & Feng, Y. 2018, MNRAS, 480, 3177
Bhowmick, A. K., Matteo, T. D., Feng, Y., & Lanusse, F. 2017, MNRAS, 474, 5993
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, ApJ, 793, 115
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34
Bowler, R. A. A., Dunlop, J. S., McLure, R. J., & McLeod, D. J. 2016, MNRAS, 466, 3612
Buchner, J., Treister, E., Bauer, F. E., Sartori, L. F., & Schawinski, K. 2019, ApJ, 874, 117
Champagne, J. B., Decarli, R., Casey, C. M., et al. 2018, ApJ, 867, 153
Chen, H., & Gnedin, N. Y. 2018, ApJ, 868, 126
Conroy, C., & White, M. 2012, ApJ, 762, 70
Cooray, A., & Milosavljević, M. 2005, ApJ, 627, L89
Costa, T., Sijacki, D., Trenti, M., & Haehnelt, M. G. 2014, MNRAS, 439, 2146
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371
DeGraf, C., & Sijacki, D. 2016, MNRAS, 466, 3331
Di Matteo, T., Croft, R. A. C., Feng, Y., Waters, D., & Wilkins, S. 2017, MNRAS, 467, 4243
Feng, Y., Di-Matteo, T., Croft, R. A., et al. 2015, MNRAS, 455, 2778
Feng, Y., Matteo, T. D., Croft, R., & Khandai, N. 2014, MNRAS, 440, 1865
Finkelstein, S. L., Ryan, R. E., Papovich, C., et al. 2015, ApJ, 810, 71

12
