THE INTRINSIC EDDINGTON RATIO DISTRIBUTION OF ACTIVE GALACTIC NUCLEI IN STAR-FORMING GALAXIES FROM THE SLOAN DIGITAL SKY SURVEY

Mackenzie L. Jones1, Ryan C. Hickox1, Christine S. Black1, Kevin N. Hainline1,2, Michael A. DiPompeo1, and Andy D. Goulding3

1 Department of Physics and Astronomy, Dartmouth College, Hanover, NH 03755, USA
2 Steward Observatory, University of Arizona, Tucson, AZ 85721, USA
3 Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

Received 2016 February 16; revised 2016 May 4; accepted 2016 May 4; published 2016 July 14

ABSTRACT

An important question in extragalactic astronomy concerns the distribution of black hole accretion rates of active galactic nuclei (AGNs). Based on observations at X-ray wavelengths, the observed Eddington ratio distribution appears as a power law, while optical studies have often yielded a lognormal distribution. There is increasing evidence that these observed discrepancies may be due to contamination by star formation and other selection effects. Using a sample of galaxies from the Sloan Digital Sky Survey Data Release 7, we test whether or not an intrinsic Eddington ratio distribution that takes the form of a Schechter function is consistent with previous work suggesting that young galaxies in optical surveys have an observed lognormal Eddington ratio distribution. We simulate the optical emission line properties of a population of galaxies and AGNs using a broad, instantaneous luminosity distribution described by a Schechter function near the Eddington limit. This simulated AGN population is then compared to observed galaxies via their positions on an emission line excitation diagram and Eddington ratio distributions. We present an improved method for extracting the AGN distribution using BPT diagnostics that allows us to probe over one order of magnitude lower in Eddington ratio, counteracting the effects of dilution by star formation. We conclude that for optically selected AGNs in young galaxies, the intrinsic Eddington ratio distribution is consistent with a possibly universal, broad power law with an exponential cutoff, as this distribution is observed in old, optically selected galaxies and X-rays.

Key words: galaxies: active

1. INTRODUCTION

In the past decade, astronomers have made significant progress in developing a generalized model of the formation and evolution of galaxies, their host bulges, and black holes across cosmic time (for reviews, see Alexander & Hickox 2012; Silk & Mamon 2012). Despite this progress, the evolution of supermassive black holes (SMBH) and their impact on their host galaxies and large-scale structures remains relatively poorly understood. As these central black holes grow via mass accretion, they emit copious amounts of light as active galactic nuclei (AGNs; Salpeter 1964; Lynden-Bell 1969; Shakura & Sunyaev 1973; Soltan 1982; Rees 1984). The masses of these SMBHs are correlated with the properties of their host stellar bulges (Kormendy & Ho 2013) and the volume-averaged galaxy–black hole growth rate is consistent with the black hole–spheroid mass relationship for a wide range of black hole masses (e.g., Heckman et al. 2004). However, other than a potential link through a common supply of cold gas, the physical processes connecting black holes and galaxies are still uncertain (Alexander & Hickox 2012).

AGNs are observed and characterized differently in various wavelength regimes (e.g., Hickox et al. 2009; Mendez et al. 2013; Goulding et al. 2014). Decades of optical observations of AGNs have led to the parameterization of AGNs based on spectroscopic characteristics (e.g., Peterson 1997). Particularly useful for high-z studies due to their brightness in the optical, type 1 AGNs exhibit optical spectra containing blue continua and both narrow- and broad-line emission, while the spectra of type 2 AGNs contain only narrow-line emission (For reviews see Antonucci 1993; Urry & Padovani 1995; Netzer 2015). The most statistically powerful studies of AGNs to date come from optical spectroscopic surveys such as the Sloan Digital Sky Survey (SDSS; York et al. 2000). In surveys of type 2 AGNs, it is possible to connect accretion to the galaxy properties via high-excitation narrow lines without contamination from the optical continuum that is characteristic of type 1 AGNs. Narrow-line AGNs may be selected based on specific emission line ratio strengths (e.g., Veilleux & Osterbrock 1987; Kewley et al. 2001; Kauffmann et al. 2003; Kewley et al. 2006) and the corresponding positions on the Baldwin et al. (1981) BPT diagram.

Likewise, in the infrared, specific emission line ratio strengths can be used to select AGNs (e.g., Pope et al. 2008; Goulding & Alexander 2009; Petric et al. 2011). More commonly, IR-selected AGNs are characterized photometrically using color selection (e.g., Lacy et al. 2007; Jarrett et al. 2011; Donley et al. 2012; Mateos et al. 2012; Stern et al. 2012). In the X-rays, AGNs appear as the most luminous X-ray sources and can be selected based on an X-ray luminosity threshold and a characteristic X-ray power-law spectrum (e.g., Chandra Deep Fields; Brandt et al. 2001; Giacconi et al. 2002; Alexander et al. 2003; Luo et al. 2008, 2010).

Once AGNs are selected through observations, measuring the distribution of the Eddington ratio (a ratio of the observed accretion rate to the maximum possible accretion given by the Eddington limit) yields important information about black hole growth, such as the typical accretion rate for black holes of a given mass, and provides a clue as to the physical properties driving this growth (e.g., Heckman et al. 2004; Aird et al. 2012). Individual galaxies can vary dramatically in AGN luminosity over short timescales (e.g., Schawinski et al. 2010; Hickox et al. 2014; Schawinski et al. 2015). So, measuring the
To further study the intrinsic AGN Eddington ratio distribution, it is useful to examine the connection between the SMBH and host galaxy. There is increasing observational evidence that the average black hole accretion rate is related to the host galaxy star formation rate in AGNs (e.g., Chen et al. 2013), whereas for individual galaxies this dependence is observed to be much weaker (e.g., Shao et al. 2010; Alexander & Hickox 2012; Rosario et al. 2012). This relationship can be reproduced with a tight correlation between star formation and time-averaged AGN activity, along with short-term variability of the AGNs over a wide dynamic range (Thacker et al. 2014; Volonteri et al. 2015). Hickox et al. (2014) find that a Schechter function AGN luminosity distribution can reproduce observed trends.

Direct measurements can also probe the Eddington ratio distribution of galaxies. In the X-rays, which are less susceptible than the optical to extinction and contamination biases (as discussed below), Aird et al. (2012) and Bongiorno et al. (2012) find an observed Eddington ratio distribution that is similar to the Schechter function. The Eddington ratio distribution of quasars can also be observed directly using optical spectroscopy; for example, for luminous AGNs, Kollmeier et al. (2006) and Shen et al. (2008) find that the Eddington ratio distribution is lognormal and peaks at high Eddington, while low-luminosity AGNs can have distributions that extend to lower Eddington ratios (Gavignaud et al. 2008). Schulze & Wisotzki (2010) observe a turnover at low Eddington ratios for low-redshift AGNs in the Hamburg/ESO survey, but attribute this to selection effects and determine that an intrinsic Schechter function better describes the Eddington ratio distribution. There is further evidence that the Eddington ratio distribution evolves with redshift, which may be indicative of a shift in the global black hole accretion density to lower black hole masses at lower redshift (e.g., Nobuta et al. 2012; Shen & Kelly 2012).

Kauffmann & Heckman (2009, hereafter KH09) present a method for determining the AGN Eddington distribution from optical spectroscopic galaxy samples, selecting AGNs based on the position of galaxies with narrow emission lines on the BPT diagram. They determine the distribution of AGN Eddington ratios for different galaxy populations and test how these distributions vary as a function of black hole mass. An outline of this method is provided in Section 3. Two distinct AGN populations in different host galaxies are separated based on a mean stellar age indicator, the relative strength change of the continuum flux across the 4000 Angstrom break (Balogh et al. 1999). A young stellar population is defined as having a break index of D4000 < 1.5 and an old stellar population has a break index of D4000 > 1.5 (Kauffmann et al. 2003). The Eddington ratio distribution for the young galaxy population is found to be lognormal in shape, while the old galaxy population has a shape consistent with a power law.

Both the lognormal and power-law models have successfully been used to fit the Eddington ratio distribution in various studies. In order to fully compare these Eddington ratio distributions, it is necessary to understand any observational biases that could be intrinsic to the methods.

While X-ray observations provide one of the most valuable methods for detecting and characterizing AGNs, studies of the cosmic X-ray background (CXB) and direct X-ray measurements (e.g., Lansbury et al. 2014; Stern et al. 2014) suggest that X-ray observations are biased against the most heavily

The statistical distribution of Eddington ratios can provide a general framework for the long-term process of black hole growth. Many theoretical models and simulations predict an intrinsic Eddington ratio distribution (Novak et al. 2011; Gabor & Bournaud 2013; Thacker et al. 2014) or assume a functional form for the Eddington ratio distribution in modeling the AGN population (Figure 1). More complicated hydrodynamical models can directly yield the intrinsic Eddington ratio distribution, while many analytic or semi-analytic models must assume a distribution because they do not have the physical capability to predict one. The distributions used in these models vary; some theoretical studies utilize a lognormal shape. One particular subset of the lognormal distribution is often referred to as the “scattered lightbulb,” in which the AGN accretion is either “on” or “off” (Conroy & White 2013). Other simulations choose a Schechter function, a power law with an exponential cutoff at high Eddington ratios which is often generalized at low Eddington as a power law (e.g., Hopkins & Hernquist 2009, in which they argue against the “lightbulb” model). Hopkins et al. (2009) also use a Schechter function for the intrinsic AGN distribution:

$$\frac{dt}{d\log L} = t_0 \left( \frac{L}{L_{\text{cut}}} \right)^{-\alpha} \exp(-L/L_{\text{cut}}). \quad (1)$$

In addition to modeling, the quasar luminosity function (QLF) is another probe into the intrinsic Eddington ratio distribution. Building on the work of Conroy & White (2013) in which the QLF is fit with a “lightbulb” model, Veale et al. (2014) compare the lognormal shape to a truncated power law in a simulated fit to the QLF. They find that at high luminosities, both distributions fit the QLF, but at low luminosities the fits are poorly constrained, which they interpret as evidence that quasars spend a majority of their time at low luminosities.

![Figure 1. Variety of intrinsic Eddington ratio distributions are used in theoretical models and simulations such as varieties of Schechter and lognormal functions, including the roughly lognormal distribution derived using the method of KH09 (green; refer to Section 3). Our chosen best-fit model of the intrinsic Eddington ratio distribution (red; refer to Section 5) is shown compared to previously published distributions.](image-url)
observed AGNs. The CXB is known to be produced primarily by accretion, the majority of which is AGN accretion, while a smaller contribution is due to X-ray binaries. This background radiation peaks in the hard X-rays (∼30 keV). However, observations have shown that the majority of AGNs are identified in the soft X-rays (0.1–2 keV; e.g., Marshall et al. 1980) and do not produce enough hard X-rays to account for the full hard CXB, implying that we are missing a significant fraction of AGNs due to obscuration (e.g., Clements et al. 2002; Spoon et al. 2004; Iwasawa et al. 2005; Hopkins et al. 2006; Downes & Eckart 2007).

Like the X-ray, optical methods suffer from selection effects. Dilution from host galaxy light and obscuration of the AGNs can affect optical emission lines, either contributing to their flux or causing them to disappear altogether (Goulding & Alexander 2009; Hopkins et al. 2009). Similarly, AGN emission in the IR is also subject to dilution from star formation (Polletta et al. 2007; Mullaney et al. 2011, 2012). Recent work by Trump et al. (2015) finds a strong bias for optical line-ratio-selected AGNs at a given accretion rate, such that for a fixed Eddington ratio, fewer AGNs are identified in low-mass, star-forming host galaxies than in massive and low star-forming hosts. It has also been shown that dilution in spectroscopic samples depends significantly on the size of the aperture used to extract the flux (Maragkoudakis et al. 2014). Additionally, aperture effects can cause extended emission from stellar processes to mimic the emission of a LINER (e.g., Yan & Blanton 2012).

In light of these selection effects, this paper aims to determine the intrinsic Eddington ratio distribution for type 2 AGNs in SDSS galaxies. Specifically, we will test whether an intrinsic Schechter function (e.g., Hopkins et al. 2009; Chen et al. 2013; Hickox et al. 2014) is consistent with the observational results obtained by KH09 for a young galaxy population. We simulate a galaxy population with a known intrinsic AGN Eddington ratio distribution where the distribution acts as a proxy for the amount of time an AGN spends at a particular Eddington ratio, and we apply the method of KH09 to determine the “observed” AGN Eddington ratio distribution. We present the observations used in Section 2 and the application of the KH09 method to observed data in Section 3. Our fiducial model is defined in Section 4, and the description and results of the application of the KH09 method to our simulated data are presented in Section 5. Furthermore, in Section 6, we present a new method for extracting the AGN contribution to the flux using optical emission line ratios. A discussion and summary are given in Section 7.

2. DATA

Our galaxy sample originates from the SDSS Data Release 7 (York et al. 2000; Strauss et al. 2002; Abazajian et al. 2009). We obtained spectroscopic information from the MPA-JHU value-added catalog.1 All emission line fluxes were previously corrected for Galactic reddening within the catalog following O’Donnell (1994). The fluxes were extinction corrected via the Charlot & Fall (2000) prescription following KH09. The extinction-corrected [O iii] fluxes then underwent a mean bolometric correction of 600, as selected by KH09. Star formation rates were available in the catalog and were determined following the method of Brinchmann et al. (2004). We calculated black hole masses via stellar velocity dispersions and using the formula given by Tremaine et al. (2002). The available stellar mass estimates in the value-added catalog were determined based on photometric fits with stellar population synthesis models (Kauffmann et al. 2003; Salim et al. 2007). The chosen sample has a spectroscopic redshift range of 0.0 < z < 0.33. After the extinction and bolometric flux corrections, we select our “real” sample to be all of the observed SDSS objects with a stellar mass measurement.

3. DETERMINING THE EDDINGTON RATIO DISTRIBUTION: OBSERVED SDSS GALAXIES

To determine the observed Eddington ratio distribution, we first separate our “real” sample of SDSS observed galaxies into two populations by mean stellar age, as defined by the D4000 index. The observed young and old galaxy samples are defined as having a D4000 break below and above 1.5, respectively. A break index of 1.5 is chosen based on the results of KH09, in which they find that the observed Eddington ratio distribution showed little dependence on D4000 for all galaxies with D4000 < 1.5. Furthermore, we use this simple definition for our “young” sample to include as large as possible a sample in our analysis, rather than the KH09 representative ranges that separate the populations into young and old by D4000 breaks below 1.4 and above 1.7, respectively. We can then further split the galaxies into groups based on the strength of the emission signal-to-noise ratio (S/N). We define “[O iii]-undetected” galaxies as having an S/N < 3 in [O iii], while galaxies that are “[O iii] detected” are defined as having S/N > 3 in [O iii]. Of the ∼770,000 SDSS galaxies retrieved from the MPA-JHU value-added catalog with mass estimates, there are ∼520,000 that are classified as [O iii] detected while ∼250,000 are [O iii] undetected. Of the [O iii]-detected galaxies, ∼306,000 of these fit our young classification.

We focus on the young sample in our analysis, in part because these galaxies are actively star-forming and suffer greatly from contamination and other selection effects. In addition, previous optical studies have yielded a lognormal Eddington ratio distribution for young galaxies, while the old galaxy population exhibits an Eddington ratio distribution that is consistent with the power-law distribution found with X-ray observations.

With the young galaxy population taken from SDSS DR7, we next reproduce the KH09 prescription for determining the Eddington ratio distribution. Of our real sample of [O iii]-detected young galaxies, we further select galaxies with significant line detection (S/N > 3) in all four emission lines used for the BPT diagram (Hα, Hβ, [N ii], and [O iii]) to directly compare our results with those in KH09. Selecting galaxies with a detection in all four emission lines rather than our original cut in [O iii] results in a decrease of roughly 4% of our initial sample. We thus select our “real” observed sample to consist of 294,934 young galaxies with significant line detections.

Following the KH09 analysis, we determine the AGN contribution to the total [O iii] flux using the BPT diagram (Figure 2). An assumption is made that galaxies defined at a particular locus on the star-forming sequence have no AGN contribution, while galaxies at a locus on the far end of the AGN sequence have line emission entirely dominated by an AGN.
4. SIMULATING THE GALAXY AND AGN POPULATION

Once the Eddington ratio distribution is found for our real sample of observed galaxies, we can proceed in simulating a population with a known intrinsic Eddington ratio distribution. By comparing the output from the KH09 method of our simulated galaxy population with that obtained from our real sample, we may better understand the underlying Eddington ratio distribution for SDSS galaxies.

The simulation is built on the assumption that every galaxy has a SMBH with some accretion based on our intrinsic Eddington ratio distribution and thus, by definition, is an “AGN.” However, not all of these galaxies will have an accretion rate that is significant enough to be identified using BPT diagnostics and may not be considered an “observed” AGN. The simulated sample begins with the SDSS galaxies. A schematic of the selection process in making our simulated sample is outlined in Figure 3. We again separate the SDSS galaxies based on [O III] detection and age. Objects in the observed sample that have S/N > 3 in [O III] are considered to be a part of our [O III]-detected galaxy class, while those with S/N < 3 in [O III] are [O III]-undetected and further separated by their position on the BPT diagram. Using the star-forming sequence boundary (Kauffmann et al. 2003), the star-forming galaxies are separated from the composite and AGN regime. Using the BPT to separate our [O III] detected is not strictly accurate for those objects without detections in all four lines, however this accounts for only 4% of our total [O III]-detected sample and furthermore allows for the possibility of those emission lines to become significant with the addition of a simulated AGN component. Our [O III]-undetected galaxy group and star-forming group are then used as a base for our simulated sample.

In order to better compare with the real observed sample population numbers, we also use objects in the Composite and AGN regime that are [O III] detected and located on the BPT diagram above the Kauffmann et al. (2003) boundary. However, these AGNs and Composite galaxies cannot be directly included in the parent sample because we have no way of knowing a priori how much the AGN contributes to the total emission line flux values. Instead, they are matched to our earlier pool of [O III]-undetected and star-forming galaxies. This match is run in three dimensions: black hole mass, D4000 line break, and redshift for ~220,000 AGNs and Composite galaxies. The distributions of stellar mass and star formation...
rate for our matched AGN and Composite galaxies are not significantly different than the distributions of the parent sample with which they are matched.

The star-forming and [O\text{iii}]-undetected galaxies are sorted into a three-dimensional cube of black hole mass (bin size of 0.1), D4000 (bin size of 0.02), and redshift (bin size 0.01). The composite and AGN objects are then sorted in the same way. The bins in each of these two 3D cubes are then matched; every composite and AGN galaxy is paired with a randomly selected star-forming or [O\text{iii}]-undetected galaxy from its matching bin. Those that do not match are then excluded, but this does not amount to a significant decrease in objects (~7000, roughly 3%). Based on the match, these AGNs and Composite galaxies are assigned emission line fluxes and star formation timescales of the matched star-forming or [O\text{iii}]-undetected galaxy before they are included in the parent sample. The black hole mass remains the same in order to match the distribution of the real observed sample, as defined in Section 3. We match to both the [O\text{iii}]-detected, star-forming galaxies and the [O\text{iii}]-undetected galaxies, rather than just the [O\text{iii}]-detected, star-forming galaxies to allow for the possibility that the [O\text{iii}]-undetected galaxies may become significant with the addition of our simulated AGN component.

Once the simulated AGN component is added to our parent sample (as discussed below) we have a comparable number of simulated galaxies to what is observed. The simulated AGN component is drawn randomly from an Eddington ratio distribution characterized by a Schechter function with an exponential cutoff near the Eddington limit (e.g., Hopkins et al. 2009; Hickox et al. 2014), as given by Equation (1). This is then added to the parent sample. We adopt an upper cutoff at the Eddington luminosity ($L_{\text{cut}} = L_{\text{edd}} = 1.38 \times 10^{38} M_{\odot} \text{ erg s}^{-1}$, assuming a radiative efficiency of 0.1), while the lower cutoff, which determines the normalization of the curve (i.e., the average luminosity), can be varied to better match the observed distribution on the BPT diagram. To further motivate our specific choice of a Schechter function, we also tested a simple power-law distribution and found the exponential cutoff is needed to match the observed suppression of high-Eddington AGNs.

We first vary the slope of the Schechter function, $\alpha$, between 0.0 and 0.8 in steps of 0.1 to determine the best fit to the observed Composite and AGN occupation fractions. From this broad search, we further refine our best fit by varying $\alpha$ between 0.28 and 0.48 in steps of 0.002. Changing the slope of the power law makes the largest impact in the distribution of the most luminous objects, as demonstrated in Figure 4 by the variation of the Eddington ratio distribution. In addition to varying $\alpha$ to improve our fit, we also allow the lower cutoff ($log(L_{\text{cut}})$) to vary between $-4.8$ and $-2.8$ in steps of 0.02 which changes the number of observed high Eddington ratio galaxies. We use the ratio of $L([\text{O\text{iii}}])/M_{BH}$ as a proxy for the Eddington ratio using a bolometric correction of 600 (KH09) on the instantaneous bolometric luminosities. This corresponds to an addition of 1.78 to the log of the Eddington ratio distribution. For example, our lower cutoffs would then span a range of $-3.02$ to $-1.02$ in $log(L([\text{O\text{iii}}])/M_{BH})$.

Once we have our simulated $L([\text{O\text{iii}}])$, we can find the AGN contribution to the other BPT emission lines. Each parent galaxy is assigned an AGN $[\text{O\text{iii}}]$ luminosity randomly selected from the given Eddington ratio distribution. We then select two regions in BPT space where we could expect an addition of AGN flux to move our parent sample; a high- or low-ionization circular region with a distribution given as a function of radius (radius $= 0.075$) and centered near the KH09 AGN track points. Based on the real observed distribution of galaxies in these two regions, we construct a probability for an object in our parent sample to move to either region once an AGN has been added, such that an object moves toward a random point within the high-ionization circular region roughly 48% of the time.

We now have an AGN contribution in [O\text{iii}], the total emission line ratios $[\text{O\text{iii}}]/\text{H}_\alpha$ and $[\text{N\text{ii}}]/\text{H}_\beta$ from the randomly selected endpoint, and the observed flux in our four emission lines from the parent sample. By adding the AGN contribution in [O\text{iii}] to the observed [O\text{iii}] flux and using the total emission ratios, we can determine the total (AGN+SF) emission in each four lines. Since we know the SF component, we can extract the simulated AGN contribution to the remaining emission lines.

By adding these simulated AGN components to the emission from the parent sample in [O\text{iii}], [N\text{ii}], $\text{H}_\alpha$, and $\text{H}_\beta$, we obtain a simulated sample of young galaxies that contain line emission from stellar processes, as well as an AGN component with $L_{\text{bol}}$ drawn from our given Schechter function distribution. Thus, we have a sample with a known intrinsic Eddington ratio distribution that we can then use to model the unknown intrinsic Eddington ratio distribution of the real SDSS observations.

We compare our simulated set of galaxies to observations in three ways. We first examine our simulated sample and the observed galaxies in BPT space to compare the general shape of our sample distribution. As shown in Figure 5, our sample is able to mimic the observed distribution of young galaxies with significant detection in all four BPT emission lines.
We then check the output of our simulation by determining the fraction of galaxies that lie above the Kauffmann et al. (2003, hereafter Ka03) and Kewley et al. (2001, hereafter Ke01) demarcation lines. The observed occupation fractions above the demarcation lines are 14.7% and 3.1%, respectively, for young galaxies with $S/N > 3$ in all four BPT flux lines. We match the observed and simulated values by adjusting the simulated intrinsic Eddington ratio distribution, specifically the lower limit and $\alpha$. The occupation fraction residuals of these fits are shown in Figure 6 as orange to white lines. As a further comparison, we calculate the Eddington ratio distribution of our simulated galaxies following the procedure discussed in Section 5.

5. DETERMINING THE EDDINGTON RATIO DISTRIBUTION: SIMULATED SDSS GALAXIES

We apply the KH09 method to galaxies in our simulated sample that have emission line fluxes in all four BPT emission lines with an $S/N > 3$, following the exact procedure used for the observed objects in Section 3. The fraction of $[\text{O} \text{\textsc{iii}}]$ due to the AGN component is similarly found for the entire set of simulated objects. The same two tracks from Section 3 have been selected to represent the AGN and LINER populations and the sample is split between these two ionization regimes using the same separation line (see Figure 2).

Thus, the “observed” contributions of AGN and star formation are computed for the simulated galaxies. It is useful to note that our simulated galaxies were built with a known AGN luminosity in $[\text{O} \text{\textsc{iii}}]$ and are being compared to the AGN component determined empirically by this procedure. The differential Eddington parameter, as defined in Section 3, is plotted in Figure 7. This Eddington parameter distribution is approximately lognormal in shape and significantly different than the input intrinsic Eddington ratio distribution.

Figure 5. Our simulated sample of young galaxies in BPT space (red) plotted over the real sample of young galaxies from SDSS (blue). The distribution Schechter function was chosen so that the simulated sample would best fit the real data.

Figure 6. Comparison of the output of our simulated BPT diagram to the observed SDSS galaxies, as a function of the parameters $\alpha$ and $L_{\text{cut}}$. The orange to white contours show residuals of the absolute difference in the Composite/AGN fraction between the simulated and real data. The blue contours in the background show the dependence of the $\chi^2$ between the simulated and observed distributions of the Eddington parameter, calculated using the method of KH09. Indicated as “Best Fit” from the KH09 method is the point of overlap with the minimum $\chi^2$ and residual fit with slope of $\alpha = 0.38$, and lower cutoff of $L_{\text{cut}} = -3.9$ (See Section 5).

Figure 7. Eddington ratio distribution extracted from the KH09 method for the observed SDSS data (red) appears lognormal, as well as the observed sample with only star-forming galaxies (No AGN; green). Our best-fit fiducial model (dark blue) for the intrinsic Eddington ratio distribution is also shown. Note that our intrinsic distribution is a power law with an exponential cutoff but becomes lognormal and consistent with the observed SDSS data after the KH09 method is applied (light blue).
After applying the KH09 method to our simulated data and the real SDSS data, we test the accuracy of our intrinsic function to reproduce the observed, real Eddington ratio distribution using $\chi^2$. The best fit is selected based on the overlap between the observed-calculated occupation fraction residuals from Section 4 and the $\chi^2$ values of the fit to the real data, as shown in Figure 6. These parameters follow a similar trend, but a fit of $\alpha = 0.38$ and lower cutoff $L_{\text{Cut}} = -3.9$ are the best overlap minimum for which the AGN fractions (defined above and below the Ka03 and Ke01 demarcations) are 14.7% and 4.8%, respectively. This fit gives us 294,994 simulated young galaxies with S/N > 3 in [O III] which is comparable to the number of real young galaxies with the same signal strength, of which there are 294,934.

To test the uncertainty of our black hole mass calculations, we introduce an artificial scatter of 0.3 dex in log($M_{\text{BH}}$) and run our simulation with our best-fit Schechter function. When comparing our scattered Eddington ratio distribution to the best-fit Eddington ratio distribution, we find that the recovered distributions agree within 0.03 dex in Eddington parameter space. With our best fit, we also find good agreement with the observed Eddington ratio distribution within 0.125 dex (see Figure 7). When comparing the intrinsic to the observed Eddington ratio distribution, the most obvious difference is the turnover around $-0.5$ dex in log($L_{\text{II}}/M_{\text{BH}}$). It is necessary to point out that when the KH09 method is applied to [O III]-detected star-forming galaxies, before the simulated AGN is added, we extract an AGN contribution that is lognormal in shape with a similar turnover (Figure 7).

Thus, the observed turnover is likely due to the intrinsic width of the star-forming sequence caused by observational and intrinsic scatter within the star-forming sequence and believed to be caused by the fluctuations and efficiency of accretion and star formation (e.g., Kewley et al. 2001; Whitaker et al. 2014). The intrinsic width of the star-forming sequence contributes to hiding low-luminosity AGNs. A star-forming galaxy spectrum with emission line ratios that are offset from the selected “pure” star-forming loci will be found to have a significant AGN contribution, even if it has little or no AGN activity. This “overestimates” the AGN contribution for a substantial number of galaxies, such that this method cannot recover significant populations with small AGN contributions. Due to this scatter, any method to extract an AGN contribution from galaxy integrated fluxes in star-forming galaxies cannot yield a low AGN contribution.

6. A NEW METHOD FOR DETERMINING THE AGN CONTRIBUTION BY POSITION ON THE BPT DIAGRAM

In this section, we develop a method for extracting the AGN contribution that accounts for the scatter of the star-forming sequence and preserves the probability distribution of finding AGN flux in each region on the BPT diagram with the aim of better reproducing the Eddington ratio distribution down to low values. Our procedure uses a simulated set of galaxies created following the process outlined in Section 5; a schematic of this method is given in Figure 8.

To recap, our simulation consists of adding an AGN luminosity component randomly drawn from a Schechter function distribution to a galaxy component selected as [O III] undetected or star-forming based on the detection of [O III] and/or position on the BPT diagram. We select the Schechter function parameters that produce a BPT diagram distribution consistent with the observations. Thus, we have created a set of simulated galaxies with known AGN contributions across the BPT diagram.

Using the positions of our simulated galaxy sample on the BPT diagram, we create a grid in BPT space in which each bin contains a sample of our simulated galaxies with the particular distribution of the AGN contribution fractions. This tells us the probability of a galaxy in each square having a specific AGN component. For a grid built from our Schechter function, the average AGN fraction in each of the selected bins is shown on the right side of the top panel of Figure 9. While the average is depicted for simplicity, a distribution is computed for each bin so that the probability is conserved (Figure 8; right panels). As expected, the upper right portion of the BPT has the highest average AGN contributions to the total flux. We have thus built a tool from which we can extract the AGN contribution fraction based on the position of observed galaxies on the BPT diagram.

We examine the AGN contribution fraction distributions across select regions of the BPT diagram corresponding to high and low ionization regions, as described in Section 4 (Figure 10). This allows us to verify that the grid method is able to extract low AGN contributions even in regions where AGN emission may be diluted by star formation. In comparison, we show that by assigning a single AGN fraction for every position on the BPT diagram based on the distance from a “pure star-forming” locus, as in KH09, we are unable to identify very small AGN contributions. We find that the grid method can more robustly extract a broad range of the AGN contribution to the flux while the KH09 method overestimates...
the AGN fraction for galaxies in the star-forming and composite regions in BPT space. This overestimation of the AGN fraction can account for the lognormal shape of the output Eddington ratio distribution and its large value at $L/L_{\text{Edd}} \sim 0.01$ that differ markedly from the input Schechter function. These differences can explain the discrepancy between the Eddington ratio distribution obtained by KH09 and those derived from X-ray surveys (see Figure 1).

We divide the real SDSS data using this grid, separating our observed galaxies into bins based on the same grid for which our model gives us known AGN contribution fraction distributions. We assign each observed galaxy an AGN contribution fraction drawn randomly from the full distribution corresponding to that particular grid position on the BPT. Thus we extract AGN contributions to the flux of the real galaxies that preserve the distribution in AGN fraction within that region in BPT space. We follow the methods outlined in Section 5 to determine the “observed” Eddington ratio distribution.

To test this method and the completeness of the results, we chose a grid built from our “best-fit” Schechter function and refer to it as our “calibration” function. We then use this grid to extract AGN contribution fractions from simulated data, assuming a known input models for the AGN Eddington ratio distribution: an intrinsic Schechter function distribution and an intrinsic lognormal distribution. Furthermore, we can fit an intrinsic distribution function to the extracted Eddington ratio distribution of the real data by applying our “calibration” grid to simulations built with our Schechter function with a large range in $\alpha$ and $L_{\text{cut}}$ as defined in Section 4. Using the calculated $\chi^2$ values, we develop a finer range of $\alpha$ and $L_{\text{cut}}$ to determine the best fit. This “best grid fit” is given by $\alpha = 0.40$ and $L_{\text{cut}} = -3.75$.

As an additional test, we can vary our calibration grid to be built from a lognormal function rather than our “best grid fit” Schechter function. The lognormal function we have elected to present is similar to what is found for young galaxies in KH09. Additional lognormal functions were tested with varying slopes and centered at different Eddington ratios. These were only able to reproduce the observed low Eddington regime at the expense of under predicting the number of AGNs at high Eddington ratios, and vice versa. We conclude that a single lognormal is not able to reproduce the full Eddington ratio distribution. Using two calibration grids allows us to verify how robust our technique is in recovering the input distribution and assess any potential biases due to the choice of “calibration” model.

The results of the following analysis are outlined below and referenced in Figure 11. The top of this figure refers to the results of our first calibration grid selected, the “best grid fit” Schechter function with slope $\alpha = 0.40$ and $L_{\text{cut}} = -3.75$ where the grid and input function are shown in dark blue.

1. We apply this grid to the real SDSS sample and extract an Eddington ratio distribution (red) that is similar to the calibration Schechter function (dark blue) in shape and exponential cutoff, if not amplitude, down to $-0.5$ dex in Eddington parameter.

2. This grid application is repeated with a sample of known intrinsic Eddington ratio distribution, one that is built from the same Schechter function used for the calibration (dark blue). We recover an Eddington ratio distribution (light blue) that matches the extracted distribution of the real SDSS sample (red) to within 0.07 dex in each bin of Eddington ratio.

3. As a further check, we change the intrinsic Eddington ratio distribution such that the simulated input is a lognormal function with width $\sigma = 0.5$ (dark green). Our calibration grid extracts an Eddington ratio distribution (light green) which does not match the shape of the real Eddington ratio distribution until around $-2$ dex in Eddington ratio space where the turnover to lower Eddington ratios exhibits a similar slope albeit different normalization.

We then consider a different calibration function to test for potential bias in the grid creation process. We plot the results of our alternate calibration grid based on a lognormal Eddington ratio distribution...
ratio distribution with width $\sigma = 0.5$ (the input function is shown in dark green) in the bottom of Figure 11.

1. With this new grid, we extract the Eddington ratio distribution of the real SDSS sample (red) and find a wide peaked shape that does not match the input lognormal in slope, amplitude, or maxima.

2. We then apply the lognormal calibration grid to our “best grid fit” Schechter function (dark blue) and extract an Eddington ratio distribution (light blue) that again closely resembles the extracted Eddington ratio distribution of the real SDSS sample. They are consistent within 0.06 dex in each bin of Eddington ratio, to an Eddington ratio around $-2$ dex where they diverge slightly.

3. Once again, we switch the test function to a lognormal intrinsic distribution, using the same lognormal function as the calibration grid (dark green), and the grid is applied to this new distribution. The extracted lognormal Eddington ratio distribution (light green) is narrower and exhibits a slightly different maximum compared to the real data.

For both of the calibration grids, the extracted Eddington ratio distribution of the intrinsic Schechter function is consistent with the Eddington ratio distribution of the real data. These are then closely correlated in shape, if not amplitude, with the intrinsic Eddington ratio distribution to $-0.5$ dex in Eddington parameter. This technique is able to recover the shape of the Eddington ratio distribution to an order of magnitude lower in Eddington ratio than previously published methods.

For the KH09 method of extracting AGN contributions from the BPT diagram on our sample of young observed galaxies to reproduce the lognormal Eddington parameter distribution found by KH09. To test the consistency with our model, we create a simulated set of galaxies with an observed star-forming and simulated AGN component built from our fiducial model. To test our simulation, we compare it to the observational data by reproducing the general shape of the observed BPT diagram and population fractions. We apply the KH09 method to our young simulated sample, which returns a lognormal function consistent with the real observed young data. This suggests that the recovered lognormal distribution could be an artefact of the method due to a bias against low-luminosity AGNs in star-forming galaxies.

7. DISCUSSION AND SUMMARY

We use the galaxy spectroscopic catalog SDSS DR7 to investigate the intrinsic Eddington ratio distribution of galaxies obtained from optical spectroscopy.

We use the KH09 method of extracting AGN contributions from the BPT diagram on our sample of young observed galaxies to reproduce the lognormal Eddington parameter distribution found by KH09. To test the consistency with our model, we create a simulated set of galaxies with an observed star-forming and simulated AGN component built from our fiducial model. To test our simulation, we compare it to the observational data by reproducing the general shape of the observed BPT diagram and population fractions. We apply the KH09 method to our young simulated sample, which returns a lognormal function consistent with the real observed young data. This suggests that the recovered lognormal distribution could be an artefact of the method due to a bias against low-luminosity AGNs in star-forming galaxies.
Furthermore, we present a new method for determining the AGN contribution to the flux using the BPT optical diagnostic diagram. Our grid method confirms our earlier results that a Schechter function is a good description for the intrinsic Eddington ratio distribution of young galaxies. In addition, the method recovers the shape of our Schechter function to more than one order of magnitude in Eddington parameter than is possible using previously published methods.

Our best-fit model to the AGN Eddington ratio distribution using the KH09 method is a Schechter function with exponential cutoff described by an $\alpha$ of 0.38 and a lower cutoff of $-3.9$. Figure 1 depicts a representation of this function compared to other popular models. The best-fit model to the Eddington ratio distribution using the grid method is very similar with $\alpha$ of 0.40 and a lower cutoff of $-3.75$. Our best-fit slopes of the Schechter function ($\approx 0.4$) are modestly flatter than those found by KH09 for passive galaxies ($\approx 0.6$), similar to the findings of recent hydrodynamical simulations that find that galaxies with higher gas fractions have a flatter slope to the Eddington ratio distribution (Gabor & Bournaud 2013).

For both methods, the extracted “observed” Eddington ratio distribution from our simulated galaxies is comparable to the distribution obtained from the real data, suggesting that an...
intrinsic Schechter function distribution is still consistent with the “observed” lognormal distribution found by KH09. If both the young and old galaxies have a Schechter function distribution, as suggested by X-ray observations (e.g., Aird et al. 2013), it is possible that AGN accretion is universal and the fueling mechanisms for all galaxies are much more comparable. Differences between individual AGNs are most likely due to small-scale stochastic variations rather than processes on galactic scales connected with star formation.

This work highlights the difficulties in identifying rapidly growing black holes in star-forming galaxies for any optical measurement with galaxy integrated fluxes due to the systematic bias against low-luminosity AGNs. Furthermore, it illustrates the extent to which emission from star formation in these host galaxies can dilute the observed AGN emission, whether it is a low-luminosity AGN or an AGN in a particularly strong star-forming galaxy. It is especially difficult to separate the concomity if the AGN flux is on scale with, or smaller than, the flux from the star formation. This is in agreement with the results of Hopkins et al. (2009), Goulding & Alexander (2009), and Trump et al. (2015) in which host galaxy star formation and obscuration are found to impact the observed flux. Despite these difficulties, it successfully shows that a Schechter function Eddington ratio distribution is consistent with the SDSS observations.

We summarize in the following points.

1. An intrinsic Eddington ratio distribution for galaxies following a Schechter function is consistent with the observed lognormal distribution found by KH09 after accounting for the method of analysis.
2. Our grid method for extracting the AGN contribution further suggests that a Schechter function is a good description for the intrinsic Eddington ratio distribution of young galaxies.
3. We confirm the bias against observing and extracting a low- to moderate-luminosity AGN component in star-forming galaxies.
4. The Eddington distribution of young galaxies appears to be consistent with a universal power law with exponential cutoff, suggesting that the fueling mechanism between young and old galaxy populations may be more complex and that differences in Eddington ratio are most likely due to small-scale, stochastic variations rather than galaxy-scale processes.

We thank our collaborators as well as the anonymous referee for constructive comments that improved the paper. This work was supported in part by an NSF GK-12 Fellowship, Dartmouth Graduate Fellowship, as well as supported by the National Aeronautics and Space Administration under grant number NNX15AU32H issued through the NASA Education Minority University Research Education Project (MUREP) through the NASA Harriott G. Jenkins Graduate Fellowship activity. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. We acknowledge the use of the Garching Value Added catalog and the Moustakas dust correction package. This research has made use of NASA’s Astrophysics Data System.
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shao, L., Lutz, D., Nordon, R., et al. 2010, A&A, 518, L26
Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, ApJ, 680, 169
Shen, Y., & Kelly, B. C. 2012, ApJ, 746, 169
Silk, J., & Mamon, G. A. 2012, RAA, 12, 917
Soltan, A. 1982, MNRAS, 200, 115
Spoon, H. W. W., Moorwood, A. F. M., Lutz, D., et al. 2004, A&A, 414, 873
Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
Stern, D., Lansbury, G. B., Assef, R. J., et al. 2014, ApJ, 794, 102
Strauss, M. A., Weinberg, D. H., Lupton, R. H., et al. 2002, AJ, 124, 1810
Thacker, R. J., MacMackin, C., Wurster, J., & Hobbs, A. 2014, MNRAS, 443, 1125
Tremaine, S., Gebhardt, K., Bender, R., et al. 2002, ApJ, 574, 740
Trump, J. R., Sun, M., Zeimann, G. R., et al. 2015, arXiv:1501.02801
Urry, C. M., & Padovani, P. 1995, PASP, 107, 803
Veale, M., White, M., & Conroy, C. 2014, MNRAS, 445, 1144
Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295
Volonteri, M., Capelo, P. R., Netzer, H., et al. 2015, MNRAS, 449, 1470
Whitaker, K. E., Rigby, J. R., Brammer, G. B., et al. 2014, ApJ, 790, 143
Yan, R., & Blanton, M. R. 2012, ApJ, 747, 61
York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579