# Does Matter Matter? Using the Mass Distribution to Distinguish Neutron Stars and Black Holes

Maya Fishbach[1] , Reed Essick[2] , and Daniel E. Holz[1,2,3,4]
[1] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
[2] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA
[3] Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA
[4] Department of Physics, University of Chicago, Chicago, IL 60637, USA

## Abstract

Gravitational-wave detectors have opened a new window through which we can observe black holes (BHs) and neutron stars (NSs). Analyzing the 11 detections from LIGO/Virgo's first gravitational-wave catalog, GWTC-1, we investigate whether the power-law fit to the BH mass spectrum can also accommodate the binary neutron star (BNS) event GW170817, or whether we require an additional feature, such as a mass gap in between the NS and BH populations. We find that with respect to the power-law fit to binary black hole (BBH) masses, GW170817 is an outlier at the 0.13% level, suggesting a distinction between NS and BH masses. A single power-law fit across the entire mass range is in mild tension with (a) the detection of one source in the BNS mass range ($\sim$1–2.5 $M_\odot$), (b) the absence of detections in the "mass-gap" range ($\sim$2.5–5 $M_\odot$), and (c) the detection of 10 sources in the BBH mass range ($\gtrsim$5 $M_\odot$). Instead, the data favor models with a feature between NS and BH masses, including a mass gap (Bayes factor of 4.6) and a break in the power law, with a steeper slope at NS masses compared to BH masses (91% credibility). We estimate the merger rates of compact binaries based on our fit to the global mass distribution, finding $R_{\rm BNS} = 871^{+3015}_{-805}$ and $\mathcal{R}_{\rm BBH} = 47.5^{+57.9}_{-28.8}$ Gpc$^{-3}$ yr$^{-1}$. We conclude that, even in the absence of any prior knowledge of the difference between NSs and BHs, the gravitational-wave data alone already suggest two distinct populations of compact objects.

*Unified Astronomy Thesaurus concepts:* Astrophysical black holes (98); Compact objects (288); Stellar mass black holes (1611); Neutron stars (1108); Gravitational wave astronomy (675); Gravitational waves (678); Astrostatistics (1882)

## 1. Introduction

The mass distribution of neutron stars (NSs) and stellar-mass black holes (BHs) is fundamental to our understanding of stellar evolution, binary formation channels, supernova physics, and the nuclear equation of state (EoS). There has been considerable effort to measure the mass distribution for NSs and BHs based on radio, X-ray, and optical observations of these systems (Valentim et al. 2011; Özel et al. 2012; Kiziltan et al. 2013; Antoniadis et al. 2016; Alsing et al. 2018; Farrow et al. 2019; Farr & Chatziioannou 2020). Indeed, there are several features in the mass distribution that are particularly relevant for understanding the physics of these systems, including the maximum NS mass, the minimum BH mass, and the purported mass gap between the most massive NS and the least massive BH. The maximum possible NS mass is governed by the nuclear EoS, and there has been significant work to extract this value by measuring the masses of electromagnetically identified NSs (see Lattimer 2012 for a review). The maximum mass of the astrophysical population of NSs is currently estimated to be $\sim$2–2.6 $M_\odot$ (Antoniadis et al. 2016; Alsing et al. 2018; Farr & Chatziioannou 2020). Although the maximum mass among astrophysically occurring NSs in binary systems may, in general, differ from the maximum gravitational mass supported by the nuclear EoS (see, e.g., discussions in Miller et al. 2019; Landry et al. 2020), it provides a useful lower bound on this uncertain quantity. Meanwhile, analyses of the BH mass distribution based on the sample of $\sim$20 BHs in X-ray binary systems suggest that the minimum BH mass does not coincide with the maximum NS

mass, implying that there is a mass gap between the two populations (Özel et al. 2010; Farr et al. 2011). However, it has been proposed that this observed mass gap may not be physical, but rather an artifact of X-ray selection effects (Kreidberg et al. 2012). Recently, a low-mass BH, possibly occupying the mass gap, was discovered in radial velocity searches (Thompson et al. 2019), and a candidate mass gap BH was discovered in the compact binary system GW190814 (Abbott et al. 2020d).[5] Understanding whether or not there is a mass gap between NSs and BHs in binary systems has implications for supernova theory and binary physics (Fryer & Kalogera 2001; Belczynski et al. 2012; Breivik et al. 2019).

Gravitational-wave (GW) detections by Advanced LIGO (Aasi et al. 2015) and Virgo (Acernese et al. 2015) provide a rapidly growing sample of binary black hole (BBH) and binary neutron star (BNS) systems. Analyzing the masses of these detections can provide a measurement of the maximum NS mass (Chatziioannou & Farr 2020) and identify the presence of a mass gap between NSs and BHs (Littenberg et al. 2015; Mandel et al. 2015, 2017; Kovetz et al. 2017). This measurement is challenging because large observational uncertainties for the component masses often make it difficult to determine whether individual systems are in the NS mass range, the mass gap, or the BH mass range (Hannam et al. 2013; Littenberg et al. 2015; Mandel et al. 2015). Littenberg et al. (2015) and Mandel et al. (2015, 2017) found that $\sim$100 low-mass detections are required to confidently detect the

---

[5] The secondary component of GW190814, with mass $m_2 = 2.59^{+0.08}_{-0.09}\,M_\odot$, may alternatively be the most massive NS ever observed.

presence of a mass gap and measure the maximum NS mass and minimum BH mass if these features are sharp. Alternatively, it has been proposed that tidal information encoded in the GW signal can be used to distinguish populations of BBH, BNS, and neutron star–black hole (NSBH) systems (Flanagan & Hinderer 2008; Read et al. 2013; Chen & Chatziioannou 2020; Fasano et al. 2020), and Wysocki et al. (2020) recently proposed an analysis to jointly measure the tidal deformability and derived quantities like the EoS together with the mass and spin distribution of the BNS population. However, the imprint of tides is much harder to extract from the GW signal than the masses (Lackey & Wade 2015).

In this paper, we focus on the mass distribution alone and characterize a possible mass feature, such as a gap, between the BNS and BBH populations. To do this, we jointly analyze the masses of the 10 BBH systems and one BNS system detected by the LIGO/Virgo Collaboration (LVC) in their first two observing runs (O1 and O2) and published in the catalog GWTC-1 (Abbott et al. 2019a). We thereby explore whether GW170817, the one BNS system of GWTC-1, is distinguishable from the BBH population based only on its mass.

In addition to the events published by the LVC in GWTC-1, new candidate BBHs have been identified in the public O1 and O2 data (Nitz et al. 2019, 2020; Venumadhav et al. 2020). In order to ensure that we understand the selection function for the catalog (see Section 2.2), we do not analyze these additional systems here, but given that they are relatively high-mass BBHs, we would not expect their inclusion to change our main conclusions. Furthermore, three events from the third observing run (O3) have been published by the LVC to date: GW190425, GW190412, and GW190814 (Abbott et al. 2020b, 2020c, 2020d). Both GW190425, a system with a total mass of $\sim 3\,M_\odot$, and GW190814, a system with a secondary mass of $\sim 2.6\,M_\odot$, are directly pertinent to the subject of this work, as they feature systems that may fall within the mass gap. Without the context of the full set of O3 events, we cannot yet include these additional systems in our population analysis. However, it is clear that the methods described here will be relevant when analyzing events from O3 and beyond.

The remainder of the paper is organized as follows. Section 2 describes the technical details of the analysis, including the parameterization employed for the mass distribution (Section 2.1) and the statistical framework of the population analysis (Section 2.2). Section 3 explores the extension of a BBH power-law fit down to the BNS mass range. We find that a single power law struggles to simultaneously fit the relatively high rate of detections in the BNS mass range (one) compared to BBH detections (10) and the lack of detections in between. In Section 4 we fit for possible features between the NS and BH mass range, including a dip and/or break in the power law, and quantify the preference for these features. In Section 5 we discuss how our results can be used to classify detections into NS and BH categories (Section 5.1) and to infer the merger rate of BNS, NSBH, and BBH systems (Section 5.2), as well as future prospects (Section 5.3). We conclude in Section 6.

## 2. Methods

We describe the parameterization of the mass distribution in Section 2.1 and then discuss the statistical framework upon which we base our inference in Section 2.2.

### 2.1. Mass Model

For our simplest model of the component mass distribution, we consider a power law with a variable minimum mass $m_{min}$, slope $\alpha$, and maximum mass $m_{max}$ (Fishbach & Holz 2017; Kovetz et al. 2017; Abbott et al. 2019b; Wysocki et al. 2019):

$$p_{PL}(m) \propto \begin{cases} m^\alpha & \text{if } m_{min} < m < m_{max}. \\ 0 & \text{else.} \end{cases} \quad (1)$$

This POWER LAW adequately fits the BH mass distribution as inferred from the GWTC-1 BBH detections (Abbott et al. 2019b; Fishbach et al. 2020). When adding the BNS detection to the fit, we gradually build on top of this simple mass distribution, introducing phenomenological features to capture possible deviations from a pure power law.

To allow for the possibility of a dip or gap in the mass spectrum, we multiply the original POWER-LAW mass spectrum by a notch filter,

$$n(m) = 1 - \frac{A}{\left(1 + \left(\frac{\gamma_{low}}{m}\right)^{\eta_{low}}\right)\left(1 + \left(\frac{m}{\gamma_{high}}\right)^{\eta_{high}}\right)}, \quad (2)$$

which suppresses the distribution when $\gamma_{low} < m < \gamma_{high}$. We refer to this model as POWER LAW + DIP. The parameters $\eta_{low}$ and $\eta_{high}$ set the sharpness of the dip's edges, while the amplitude of the dip is set by the parameter $A$. In principle, we can allow the data to inform our knowledge of the sharpness of the gap edges, in addition to their placement and the depth of the gap. However, since we cannot meaningfully constrain all of these features with only 11 events, we fix the edges to be near-infinitely sharp: $\eta_{low} = \eta_{high} = 50$. With sharp edges, $A = 1$ corresponds to an empty gap, while $A = 0$ corresponds to no dip; $A < 0$ corresponds to a bump rather than a dip. We verify that the mass distribution we infer by simultaneously fitting the sharpness parameters $\eta_{low}$ and $\eta_{high}$, in addition to $\gamma_{low}$, $\gamma_{high}$, and $A$, is nearly identical to the mass distribution inferred under the reduced model with $\eta_{low} = \eta_{high} = 50$. This is due to the degeneracies between the parameters of Equation (2) within the large statistical uncertainties on the shape of the gap.

Similar to the notch filter that models the mass gap, we can use a low-pass filter that "turns off" the mass distribution when $m_{max} < m$ to model the edge of the upper/pair-instability mass gap, rather than the sharp cutoff at $m_{max}$. In other words, we can replace the condition that $p(m > m_{max}) = 0$ by multiplying $p(m)$ by

$$l(m) = \left(1 + \left(\frac{m}{m_{max}}\right)^n\right)^{-1}, \quad (3)$$

where large $n$ corresponds to a sharp cutoff and small $n$ corresponds to a gradual turnoff. Similar to the low-mass-gap's edges, we do not have enough detections to meaningfully constrain the steepness of the upper mass gap, and we fix the cutoff to be sharp: $n = 50$.

As a final complication, we include the possibility of a break in the power law at $\gamma_{high}$, so that objects below the gap may follow a different power-law slope $\alpha_1$ from objects above the
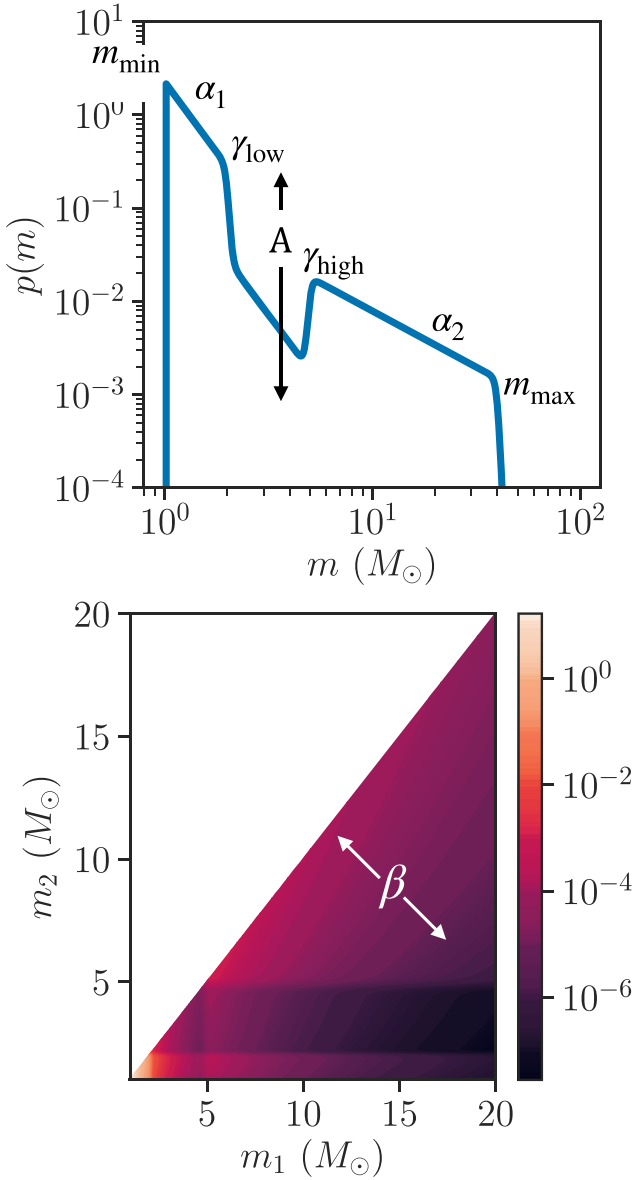
**Figure 1.** Example phenomenological distribution described in Section 2.1. Top: one-dimensional mass distribution parameterized according to Equation (5): a broken power law with slopes $\alpha_1$ and $\alpha_2$ and break at $\gamma_{high}$, with a notch filter between $\gamma_{low}$ and $\gamma_{high}$ with amplitude $A$. Bottom: corresponding two-dimensional distribution, constructed from the one-dimensional distribution with a mass ratio–dependent pairing function following Equation (6). The color bar denotes the probability density $p(m_1, m_2)$.

gap with slope $\alpha_2$:

$$p_{BPL}(m) \propto \begin{cases} m^{\alpha_1} & \text{if } m_{min} < m \leqslant \gamma_{high} \\ m^{\alpha_2} & \text{if } \gamma_{high} < m < m_{max}. \\ 0 & \text{else.} \end{cases} \quad (4)$$

The most general one-dimensional mass distribution we consider is therefore

$$p(m|\lambda) \propto p_{BPL}(m) \times n(m) \times l(m), \quad (5)$$

with free parameters $\{m_{min}, m_{max}, \alpha_1, \alpha_2, A, \gamma_{low}, \gamma_{high}\}$ denoted by $\lambda$. We refer to this as POWER LAW + DIP + BREAK.

Figure 1 demonstrates some of the features of our parameterization. Physically, $\gamma_{low}$ may correspond to the maximum NS mass and $\gamma_{high}$ to the minimum BH mass. However, the mapping between the physical properties, such as the maximum NS mass,

and features in the overall mass distribution, such as the onset of a mass gap, may be more complicated due to, for example, the supernova mechanism or accretion from a binary partner. This idealized model allows us to explore whether a single power law ($A = 0$; $\alpha_1 = \alpha_2$) can fit the entire compact object mass spectrum or whether there is a distinguishing feature between the NS and BH mass spectrum in the form of a dip ($0 < A < 1$), gap ($A = 1$), and/or break in the power law ($\alpha_1 \neq \alpha_2$). If such a feature is found, it can be used to identify subpopulations of the compact object mass distribution. In this case, the total mass distribution can alternatively be modeled as a mixture model of subpopulations (e.g., Kapadia et al. 2020).

As in Doctor et al. (2020) and Fishbach & Holz (2020), we assume a simple pairing function to generate a joint distribution over both component masses that make up a binary system, given a particular component mass distribution:

$$p(m_1, m_2|\Lambda) \propto p(m = m_1|\lambda)p(m = m_2|\lambda)$$

$$\times \left(\frac{m_2}{m_1}\right)^{\beta} \Theta(m_2 \leqslant m_1), \quad (6)$$

where $\Lambda$ represents the total set of free parameters $\{m_{min}, m_{max}, \alpha_1, \alpha_2, A, \gamma_{low}, \gamma_{high}, \beta\}$, or the union of $\lambda$ and $\{\beta\}$, and $\Theta$ is the Heaviside step function that enforces our labeling convention that $m_2 \leqslant m_1$. Here we take the pairing function to be a power law in the mass ratio. More complicated pairing probabilities are possible, and indeed, any function $p(m_1, m_2)$ can be factored into a product of the one-dimensional mass distribution and a pairing function. We stick with this simple model because it adequately reproduces the observed distribution of GWTC-1 (Fishbach & Holz 2020).

### 2.2. Statistical Framework

The analysis presented in this work consists of two main steps: (a) model fitting, i.e., given the GW data, measuring the population parameters of the mass distribution model; and (b) model checking, i.e., simulating sets of observable data from the fit to the model and evaluating how closely they resemble the actual set of observed data. This subsection provides an overview of these calculations; the details are provided in the Appendix.

Using the parameterized mass distributions from Section 2.1, $p(m_1, m_2| \Lambda)$, we construct a hierarchical Bayesian inference to determine the appropriate population-level parameters, $\Lambda$, given the observed data $\{D_i\}$ (Loredo 2004; Mandel 2010; Mandel et al. 2019; Thrane & Talbot 2019). We focus on the mass distribution alone, fixing the spin distribution (uniform in spin magnitude and isotropic in orientation) and redshift distribution (flat in comoving volume and source-frame time). The posterior on the population hyperparameters, $p(\Lambda \mid \{D_i\})$, is evaluated according to the methods in the Appendix. Each draw from this hyperposterior $p(\Lambda \mid \{D_i\})$ corresponds to a mass distribution $p(m_1, m_2| \Lambda)$. Averaging the mass distribution over the hyperposterior yields the posterior population distribution:

$$p(m_1, m_2|\{D_i\}) = \int p(m_1, m_2|\Lambda)p(\Lambda|\{D_i\})d\Lambda. \quad (7)$$

We often present the mass distribution in terms of the astrophysical merger rate density, denoted by $dN/dm_1 dm_2 dV_c dt_s$, where $V_c$ is the comoving volume[6] and $t_s$ is the time as measured
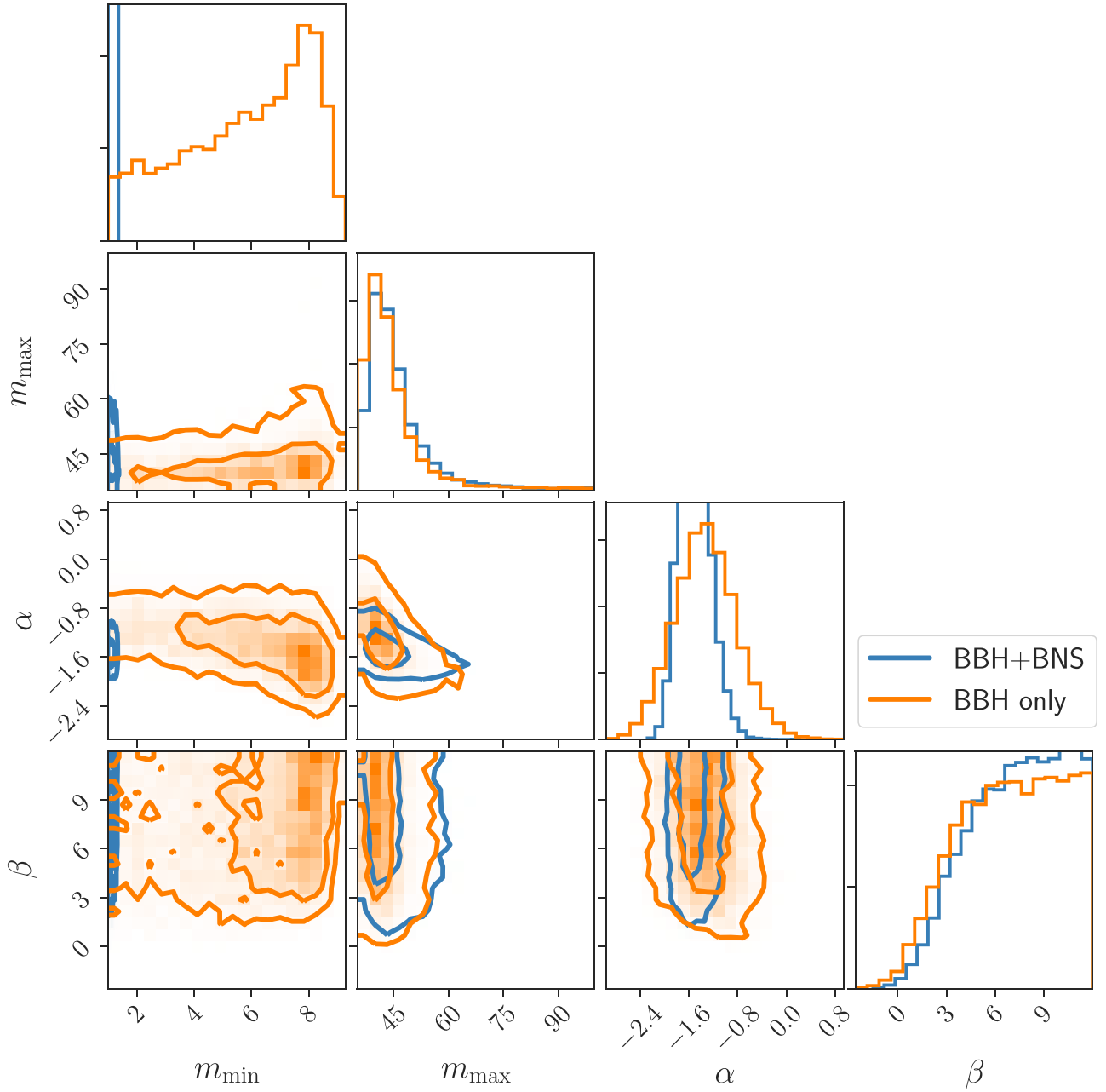
**Figure 2.** Corner plot (Foreman-Mackey 2016) comparing (orange) the power-law fit to the 10 BBHs and (blue) the fit to all 11 events. Contours show 50% and 90% credible regions. The main effect of adding the BNS event GW170817 to the power-law fit is on the $m_{\min}$ constraints (first column), particularly the joint $\alpha$–$m_{\min}$ constraints (first column, third row). The low mass of GW170817 forces $m_{\min} \lesssim 1.3$ but prefers a relatively steeper power-law slope compared to the BBH-only joint fit for $\alpha$–$m_{\min}$. However, the power-law fit to all 11 events remains consistent with the BBH-only fit within the 90% level.

in the source frame, rather than the probability density $p(m_1, m_2)$. While the probability density $p(m_1, m_2)$ integrates to unity, the rate density integrated over the masses $(m_1, m_2)$ yields the overall merger rate, $dN/dV_c dt_s$. The rate density is related to the probability density according to Equations (A1)–(A3).

Once we fit the population model, we perform a posterior predictive check by comparing the distribution of observed masses as implied by the fit to the model, or the posterior predictive distribution (PPD), to the actual set of observed events, or the empirical distribution function (EDF). This comparison provides strong goodness-of-fit tests, along with further insight into why certain features of the overall mass distribution are favored.

## 3. Can a Single Power Law Fit NS and BH Masses?

In this section, we discuss the ability of a simple power law (POWER LAW of Equation (1)) to reproduce the 11 detections of GWTC-1. We ask whether the BNS detection, GW170817, is distinguishable from the BBH population based on its mass alone. If we did not know (based, for example, on its electromagnetic counterpart or prior astrophysical information) that GW170817 belonged to a separate class of compact objects, would we have classified it as a population outlier based on its mass? Do the GW data alone suggest the existence of distinct populations of NSs and BHs?

We begin by exploring whether the same power law that fits the BBH detections can also accommodate the BNS
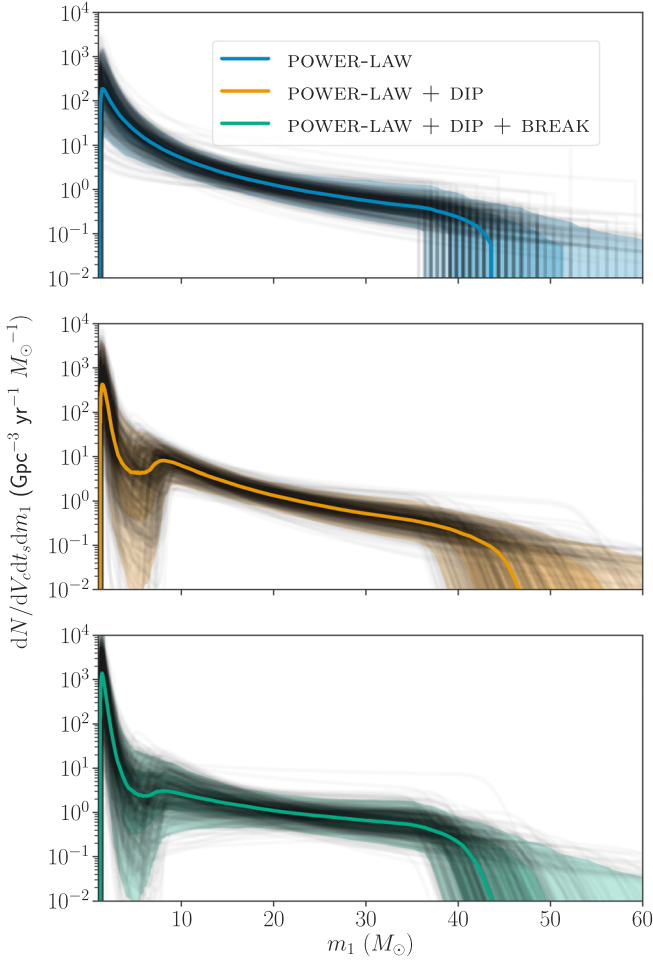
**Figure 3.** Posterior population distributions for three models, in order of increasing complexity: POWER LAW, POWER LAW + DIP, and POWER LAW + DIP + BREAK. Each panel shows the differential merger rate density, $dN/dV_c dt_s dm_1$, as a function of primary mass $m_1$. The colored lines show the median rate density, and the colored shaded bands show $1\sigma$ (68%) and $2\sigma$ (95%) credible intervals. In gray, we plot 500 draws from the population posterior under each model.

detection, GW170817. The POWER-LAW fit to the 10 GWTC-1 BBHs yields $m_{min} = 6.2^{+2.4}_{-4.5} M_\odot$, $\alpha = -1.34^{+0.87}_{-0.80}$, $m_{max} = 42.2^{+20.2}_{-5.5} M_\odot$, and $\beta = 7.2^{+4.4}_{-5.4}$ (median and 90% equal-tailed intervals; see Figure 2). Unlike previous power-law fits to the BBH (Abbott et al. 2019b; Roulet & Zaldarriaga 2019; Fishbach & Holz 2020), we allow the prior on the minimum mass to extend down to $1 M_\odot$, as opposed to 3 or $5 M_\odot$. Specifically, we assume flat priors on all hyperparameters in the range $m_{min} \in [1, 10], m_{max} \in [30, 100], \alpha \in [-4, 2], \beta \in [-4, 12]$. We find that although the posterior on $m_{min}$ peaks at $8.2 M_\odot$, there remains posterior support down to the lower prior boundary of $m_{min} = 1 M_\odot$. The posterior probability at the peak $m_{min} = 8.2 M_\odot$ is ~three times larger than at $m_{min} = 1 M_\odot$.

While the POWER-LAW fit to the BBH does not rule out masses as low as $1 M_\odot$, we would not expect to detect them very often. Based on the BBH-only POWER-LAW fit, we would expect to detect a GW170817-like system, with primary mass $m_1 \lesssim 2 M_\odot$, in a set of 11 total detections only 0.13% of the time, suggesting that GW170817 is a fairly atypical system with respect to the BBH population.
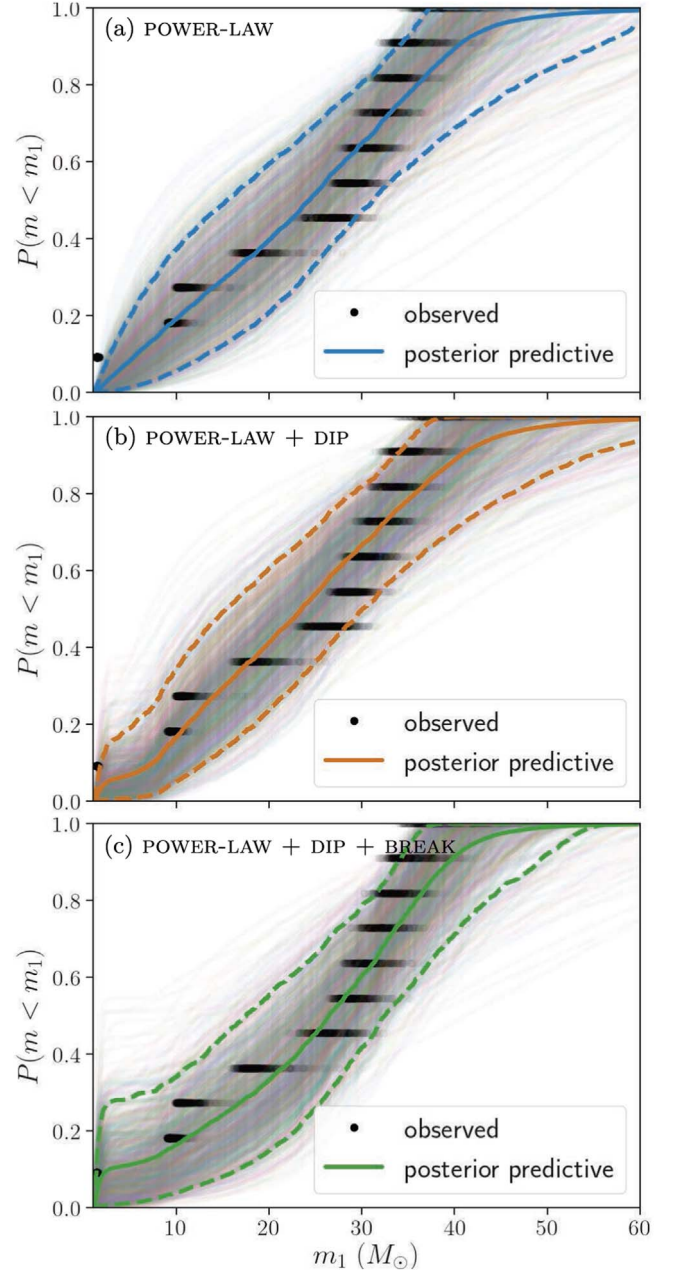


**Figure 4.** Posterior predictive check comparing the observed primary mass distribution as predicted from our models (thin colored curves; each curve corresponds to a draw from the hyperparameter posterior) with the empirical distribution from the 11 GWTC-1 events (black; each point corresponds to a draw from the population-informed single-event posterior). The solid colored line in each panel corresponds to the posterior average (mean) of the predicted distributions, while the two dashed colored lines denote the symmetric 90% interval around the predicted curves. The model is a good fit to the data if the empirical distribution (black) is contained within the range of model predictions (colored). The top panel shows that the POWER-LAW model has trouble accounting for GW170817, while the other models account for all 11 events.

If we now include GW170817 and fit the POWER-LAW model to all 11 events in GWTC-1, we find constraints on the hyperparameters that are broadly compatible with the BBH-only fit; see the comparison in Figure 2. The largest shift between the BBH-only fit and the all-event fit is in the joint $(m_{min}, \alpha)$ posterior, as seen in the third row of the first column of Figure 2. Due to the correlation between $\alpha$ and $m_{min}$, if we
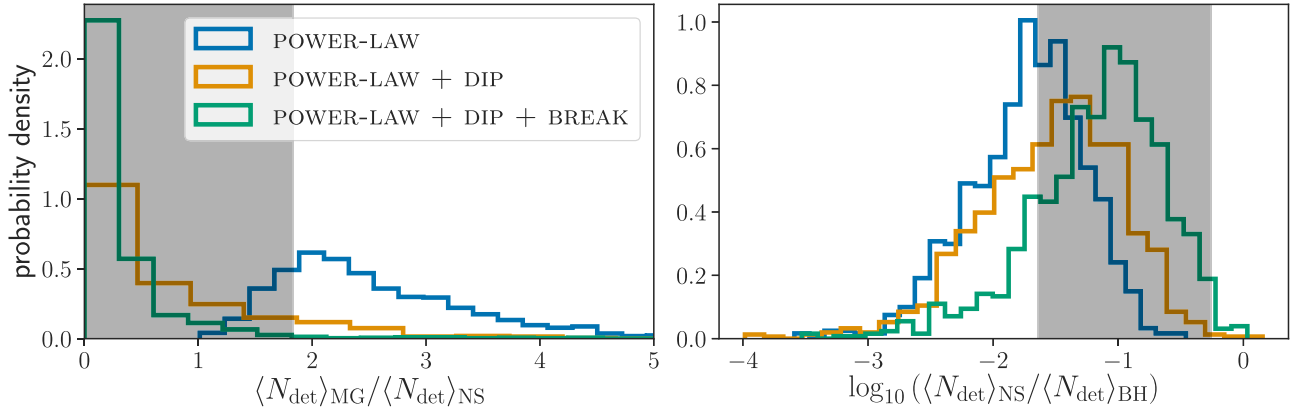
**Figure 5.** Left: ratio of the expected number of detections with primary mass in the MG (defined here as 2.5–5 $M_\odot$) compared to the expected number of NS detections (defined here as 1–2.5 $M_\odot$). Right: ratio of the expected number of detections with an NS primary mass (in the mass range 1–2.5 $M_\odot$) compared to a BH primary mass (in the mass range 5 $M_\odot$–100 $M_\odot$). Here GWTC-1 contains zero events with $m_1$ in the MG, one BNS, and 10 BBHs, leading us to measure an mass-gap–NS ratio of ∼zero and an NS–BH ratio of ∼0.1; the 90% highest posterior density credible intervals on these values are shown by the gray shaded regions (see the text for more detail). The simple POWER-LAW model predicts at least as many mass gap detections as NS detections. Meanwhile, when we allow for both a dip and a break in the power law, we lower the expected fraction of mass-gap detections and raise the expected fraction of NS detections relative to BH detections, allowing us to better fit the observed number of detections in each bin.

constrain $m_{\min} < 2\,M_\odot$ in the BBH-only fit, we find a shallower slope, $\alpha = -1.09^{+0.79}_{-0.56}$, compared to the slope in the all-events fit, $\alpha = -1.52^{+0.45}_{-0.35}$. Nevertheless, the hyperparameter posteriors agree within the 90% levels between the two fits.

The inferred primary mass distribution for the POWER-LAW fit to the 11 GWTC-1 events can be seen in the top panel of Figure 3, while the goodness of fit of the POWER-LAW model can be visualized in the posterior predictive check of Figure 4. The thin colored curves of Figure 4 show 1000 draws from the PPD of observed primary masses. Each curve corresponds to a different draw from the hyperparameter posterior. The solid line corresponds to the mean of these 1000 realizations, and the dashed lines show 90% credible bounds. In black, we show 1000 draws from the EDF. Each draw from the EDF is found by reweighting the single-event posteriors to the population prior. From the updated posterior for each of the 11 events, we draw an $m_1$ sample and order these 11 points from smallest to largest. The EDF passes through (∼1.6 $M_\odot$, 1/11), driven by the primary mass of GW170817, which is above the 90% predictive band. While this is suggestive, it is not terribly unexpected from noise fluctuations affecting the most extreme members of a set (Weibull 1951; Fishbach et al. 2020). Based on the POWER-LAW fit to all events, we expect to detect an event with primary mass $m_1 < 2\,M_\odot$ in a set of 11 events 17% of the time, a significant shift from the expected 0.13% for the BBH-only fit.

Another way of evaluating the goodness of fit of the models is shown in Figure 5, which compares the expected fraction of detections in different primary mass bins—the NS mass range $m_1 \in [1, 2.5]\,M_\odot$, the mass gap range $m_1 \in [2.5, 5]\,M_\odot$, and the BH mass range $m_1 \in [5, 100]\,M_\odot$—as predicted from the POWER-LAW fit (in blue). These values are found by integrating the PPD cumulative distribution functions of Figure 4 between the specified mass boundaries. The boundaries used here are chosen only for illustrative purposes, although the maximum mass achievable by NSs is an area of active study (e.g., Margalit & Metzger 2017; Abbott et al. 2018a, 2020a; Rezzolla et al. 2018; Shibata et al. 2019; Essick et al. 2020; Landry et al. 2020; Essick & Landry 2020). Regardless of the precise boundaries, the true mass distribution should be able to

accurately predict the fraction of events detected in each category. For the remainder of this work, we will use the NS, mass gap, and BH labels to refer to these bins in primary mass, unless stated otherwise.

Despite the large measurement uncertainties on the primary mass, each of the GWTC-1 events falls clearly into one of these mass-based categories, so that we can trivially count one NS, zero mass gaps, and 10 BHs. Letting $f_{\rm NS}$, $f_{\rm mass\ gap}$, and $f_{\rm BH}$ denote the expected fraction of detections in the NS, mass gap, and BH mass range according to the true underlying mass distribution, the observed number of detections in each category, out of $N$ total detections, follows a trinomial distribution:

$$p(N_{\rm NS}, N_{\rm MG}, N_{\rm BH} | f_{\rm NS}, f_{\rm MG}, f_{\rm BH})$$
$$= \frac{N!}{N_{\rm NS}! N_{\rm MG}! N_{\rm BH}!} f_{\rm NS}^{N_{\rm NS}} f_{\rm MG}^{N_{\rm MG}} f_{\rm BH}^{N_{\rm BH}}. \quad (8)$$

Given $N_{\rm NS} = 1$, $N_{\rm mass\ gap} = 0$, and $N_{\rm BH} = 10$, we can calculate the posterior on $f_{\rm NS}$, $f_{\rm mass\ gap}$, and $f_{\rm BH}$ according to

$$p(f_{\rm NS}, f_{\rm mass\ gap}, f_{\rm BH} | N_{\rm NS}, N_{\rm mass\ gap}, N_{\rm BH})$$
$$\propto p(N_{\rm NS}, N_{\rm mass\ gap}, N_{\rm BH} | f_{\rm NS}, f_{\rm mass\ gap}, f_{\rm BH}) \quad (9)$$
$$\times p_0(f_{\rm NS}, f_{\rm mass\ gap}, f_{\rm BH}).$$

We take the Jeffreys prior for $p_0(f_{\rm NS}, f_{\rm mass\ gap}, f_{\rm BH})$, which is a symmetric Dirichlet distribution with a concentration parameter $\alpha = 0.5$. The posterior is then given by a Dirichlet distribution with parameters $\boldsymbol{\alpha} = (N_{\rm NS} + 0.5, N_{\rm mass\ gap} + 0.5, N_{\rm BH} + 0.5)$. The posterior on $f_{\rm NS}$, $f_{\rm mass\ gap}$, and $f_{\rm BH}$, produced by drawing from a Dirichlet distribution in this way, is used to produce the gray shaded regions in Figure 5. Here the expected number of events in category $\mathcal{C}$, $\langle N_{\rm det} \rangle_{\mathcal{C}}$, is related to the expected fraction $f_{\mathcal{C}}$ by $\langle N_{\rm det} \rangle_{\mathcal{C}} = N f_{\mathcal{C}}$, so $\langle N_{\rm det} \rangle_{\rm mass\ gap} / \langle N_{\rm det} \rangle_{\rm NS} = f_{\rm mass\ gap} / f_{\rm NS}$, and so on.

According to the POWER-LAW fit to all 11 GWTC-1 detections, we should detect one BNS system per $48^{+370}_{-38}$ systems containing a BH. This is in mild tension with our detection of one BNS system per 10 BBH systems, which implies that the detection ratio is $\langle N_{\rm det} \rangle_{\rm BH} / \langle N_{\rm det} \rangle_{\rm NS} = 8.8^{+51.0}_{-6.5}$ (median and 90% symmetric

interval). Meanwhile, according to the POWER-LAW fit, we expect $2.37^{+2.06}_{-0.98}$ systems with a mass gap primary mass per BNS event. Again, this is in mild tension with our observation of no systems in the mass gap and one BNS system, which suggests the detection ratio $\langle N_{\rm det}\rangle_{\rm MG}/\langle N_{\rm det}\rangle_{\rm NS} < 1.9$ at 90% credibility.

In summary, we find that within the statistical uncertainties, the POWER-LAW model provides a marginally adequate fit to the 11 GWTC-1 detections. However, tensions emerge, hinting at possible features in the mass spectrum between the NS and BH mass range.

(a) *The event GW170817 is a low-mass outlier with respect to the BBH population.* Based on the BBH-only POWER-LAW fit, we would expect to detect a system with $m_1 < 2\,M_\odot$, given 11 total detections, only 0.13% of the time. When we update the POWER-LAW fit with all 11 detections, the hyperparameters shift to accommodate GW170817 (see Figure 2), and this increases to 17% of the time.

(b) *The mass gap is too empty.* The POWER-LAW fit to the 11 detections overpredicts the number of mass gap detections compared to NS detections. We would expect to detect a greater number of BNS systems than mass-gap systems, given 11 detections, only 12% of the time.

(c) *The event GW170817 is a surprise.* The POWER-LAW fit underpredicts the number of NS detections compared to BH detections; out of 11 total detections, we would expect to detect one NS primary mass and 10 BH primary masses only 9% of the time based on this fit.

In the following section, we characterize possible features between the NS and BH mass spectrum, including a mass gap and a power-law break, and explore how their presence alleviates these tensions.

### 4. Characterizing a Feature between NSs and BHs

The previous section examined the ability of a single power law to fit the BNS and BBHs of GWTC-1. Below, we fit the full mass distribution of Equation (5) to the GWTC-1 events. We investigate the presence of a feature between the NS and BH mass spectrum, quantifying the evidence in favor of a mass dip, gap, or break between NS and BH masses.

The first extension we consider to a power-law mass spectrum is POWER LAW + DIP, parameterized by the notch filter of Equation (2), which suppresses the merger rate for masses $\gamma_{\rm low} < m < \gamma_{\rm high}$. The free parameters of this model are the minimum NS mass $m_{\rm min}$, maximum BH mass $m_{\rm max}$, power-law slope $\alpha$, amplitude of the dip $A$, dip boundaries $\gamma_{\rm low}$ and $\gamma_{\rm high}$, and mass-ratio power-law slope $\beta$. While it is important to remember that the lower edge of the dip may or may not correspond to the maximum NS mass, depending on whether BHs exist below the gap, this subtlety does not affect our analysis. For convenience, we introduce a parameter describing the gap width $w = \gamma_{\rm high} - \gamma_{\rm low}$ and set flat priors on $m_{\rm min} \in [1\,M_\odot,\ 1.4\,M_\odot]$, $m_{\rm max} \in [35\,M_\odot,\ 100\,M_\odot]$, $\alpha \in [-5, 2]$, $A \in [0,\ 1]$, $\gamma_{\rm low} \in [1.4\,M_\odot,\ 3\,M_\odot]$, $w \in [2\,M_\odot,\ 6\,M_\odot]$, and $\beta \in [-4, 12]$.

With this choice of priors, we allow for a mass gap starting at $1.4\,M_\odot < \gamma_{\rm low} < 3\,M_\odot$ with a width of $2\,M_\odot < w < 6\,M_\odot$. Our priors on the dip location are externally motivated by observational and theoretical expectations for NS masses (see, e.g.. Özel & Freire 2016; Essick et al. 2020; Farr & Chatziioannou 2020; Landry et al. 2020). We verify that our results are not driven by the prior choice with a "look-elsewhere"

test, fitting the full POWER LAW + DIP + BREAK model to only the 10 GWTC-1 BBH detections and finding that, although we cannot rule out the presence of a second dip in the mass spectrum, there is no compelling evidence for a dip in the BBH mass range; we simply recover the prior on $A$.

The fit to the primary mass distribution under POWER LAW + DIP is shown in the middle panel of Figure 3. Comparing to the top panel, which shows the POWER-LAW fit, we can see that the data prefer some decrease in the merger rate between $\gamma_{\rm low} = 2.2^{+0.6}_{-0.5}$ and $\gamma_{\rm high} = 6.7^{+1.0}_{-1.5}\,M_\odot$.[7] The posterior on the amplitude of the dip, $A$, peaks at $A = 1$, corresponding to a perfect mass gap, with a tail down to $A = 0$, corresponding to an uninterrupted power law. We find that a perfect mass gap is preferred over a POWER LAW by a factor of 4.6. As an additional test, we fix $A = 1$ within our parameterization of Equation (2) and allow the sharpness of the gap edges, $\eta_{\rm low}$ and $\eta_{\rm high}$, to vary between $\eta_{\rm low} = \eta_{\rm high} = 0$ (no gap) and $\eta_{\rm low} = \eta_{\rm high} = 50$ (perfect mass gap).[8] We find that within this model, a perfect gap ($\eta_{\rm low} = \eta_{\rm high} = 50$) is preferred over an uninterrupted power law by a Bayes factor of 6.0, similar to the preference we recover in the default model. While this is suggestive, noise fluctuations are expected to occasionally produce Bayes factors at least this large (see, e.g., Agathos et al. 2014). Indeed, Jeffreys (1961) suggested that the ratio would need to be $\gtrsim 100$ to be decisive.

The second feature we allow in the mass distribution is a break in the power law in addition to a dip. The fit to the primary mass distribution under the POWER LAW + DIP + BREAK model is shown in the bottom panel of Figure 3. In this model, we adopt the same priors on the free parameters $m_{\rm min}$, $m_{\rm max}$, $A$, $\gamma_{\rm low}$, $w = \gamma_{\rm high} - \gamma_{\rm low}$, and $\beta$. Within the full model, we recover similar constraints on the parameters describing the dip, still favoring a mass gap $A = 1$ over a simple power law $A = 0$ by a factor of $\sim 4$. Again, this is suggestive but not definitive.

The POWER LAW + DIP + BREAK model additionally allows for a break in the power law at $\gamma_{\rm high}$, with a power-law slope $\alpha_1$ at $m < \gamma_{\rm high}$ and slope $\alpha_2$ at $m > \gamma_{\rm high}$. This allows for the possibility that the NS and BH mass spectra are described by different power laws, with a possible gap between them. We set flat, uninformative priors on $\alpha_1$ and $\alpha_2$: $\alpha_1, \alpha_2 \in [-8, 2]$. The joint posterior on $\alpha_1$ and $\alpha_2$ can be seen in Figure 6. We find that $\alpha_1$, the NS power-law slope, is likely steeper than $\alpha_2$, the BH power-law slope. The data prefer $\alpha_1 < \alpha_2$ at 91% credibility, with $\alpha_1 = -2.58^{+0.72}_{-0.87}$ and $\alpha_2 = -1.16^{+0.50}_{-0.45}$. The inferred value of the BH power-law slope, $\alpha_2$, is very similar to the power-law slope inferred with the BBH-only fit, $\alpha = -1.34^{+0.87}_{-0.80}$, and prefers to be slightly shallower than the power-law slope inferred under the POWER-LAW model fit to all events, $\alpha = -1.52^{+0.45}_{-0.35}$; see the comparison in Figure 6.

We can understand the preference for both a dip and a break between NS and BH masses by returning to the posterior predictive checks of Figures 4 and 5. The simple POWER-LAW fit underpredicts the fraction of detections in the NS range while overpredicting the fraction of detections in the mass gap range, compared to the current observations of one BNS, zero mass gaps, and 10 BBHs. In the POWER-LAW model, the rate of

---

[7] The posterior on $\gamma_{\rm low}$ is not well constrained and follows the prior, which is to be expected from only one detection between $m_{\rm min}$ and $\gamma_{\rm low}$ and the low sensitivity at low masses.

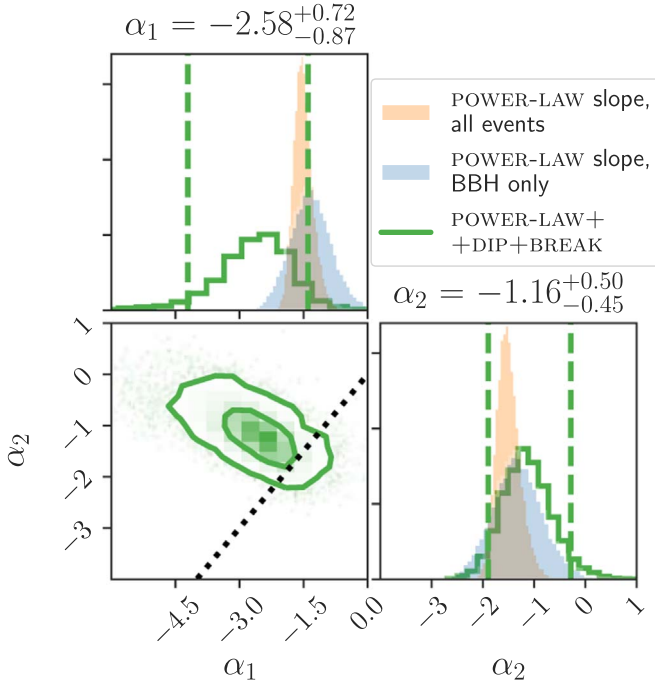[8] Recall that our default choice throughout this work is to fix $\eta_{\rm low} = \eta_{\rm high} = 50$.

**Figure 6.** Joint posterior on the power-law slopes $\alpha_1$ (NS mass range) and $\alpha_2$ (BH mass range) for the full POWER LAW + DIP + BREAK model (green). The two-dimensional contours show 50% and 90% credible regions, while vertical dashed lines show one-dimensional 90% credible intervals. For comparison, we show the power-law slope inferred under the POWER-LAW fit to (blue) only the BBH and (orange) all GWTC-1 events. We recover $\alpha_1 < \alpha_2$ with 91% credibility; $\alpha_1 = \alpha_2$ in this model reduces to the POWER LAW + DIP model.

mass gap detections must be at least 1.5 times as large as the NS detection rate (at 90% credibility); increasing the fraction of NSs within the POWER-LAW model necessitates an increase in the fraction of mass gap events.

By introducing a dip, we decrease the expected number of mass gap detections to <1.9 per NS detection (90% credibility), while slightly increasing the expected fraction of NS detections to one NS detection per $28^{+288}_{-22}$ BH detections.

Introducing a break in the power law in addition to a dip allows us to further increase the expected number of NS detections to one NS detection per $13^{+141}_{-10}$ BH detections, bringing it close to the GWTC-1 observation of one BNS per 10 BBH detections while maintaining a low rate of mass gap detections (<0.80 per NS detection).

## 5. Discussion

Given the evidence for a feature between NSs and BHs, we now consider several implications, including delineating our knowledge about specific objects in Section 5.1, updated astrophysical rates in Section 5.2, and prospects for the coming years in Section 5.3.

### 5.1. Updated Single-event Classification

Although not conclusive, the GWTC-1 detections show hints of a feature between NS and BH masses. Future detections will allow us to resolve this feature with increased precision, which may provide a natural boundary between the NS and BH populations. Meanwhile, our inference of the compact object mass distribution allows us to update the mass measurements of individual events, often allowing for much tighter constraints than the posteriors inferred under uninformative
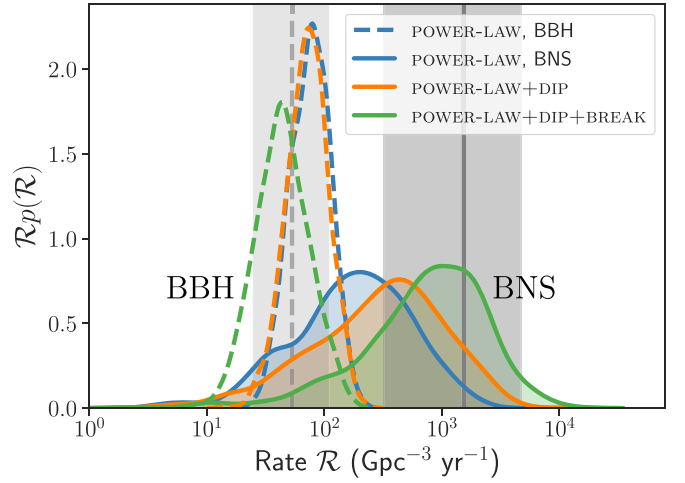


**Figure 7.** Astrophysical merger rate within two different mass bins: for BNS, $1\,M_\odot < m_2 < m_1 < 2.5\,M_\odot$, and for BBH, $5\,M_\odot < m_2 < m_1 < 100\,M_\odot$, as inferred by each of the models. The dashed open probability density curves centered below $\mathcal{R} \sim 10^2\,\mathrm{Gpc}^{-3}\,\mathrm{yr}^{-1}$ show the BBH rate inference and the solid filled curves show the BNS rate inference for the POWER LAW (blue), POWER LAW + DIP (orange), and POWER LAW + DIP + BREAK (green) models. For comparison, the gray shaded regions show the median and 90% credible intervals of the BBH (dashed) and BNS (solid) rates inferred by the LVC in Abbott et al. (2019a, 2019b). Allowing for a dip and break between NS and BH masses tends to decorrelate the merger rates, increasing the inferred BNS merger rate and decreasing the BBH merger rate.

priors (Galaudage et al. 2019; Fishbach et al. 2020; Miller et al. 2020). For example, if the population fit reveals a mass gap between the NS and BH masses, applying the population prior to events for which the likelihood measurement uncertainty is broad and overlaps with the gap will significantly tighten the mass posteriors, forcing the posterior support to lie below or above the gap (Fishbach et al. 2020). Simultaneously fitting the population distribution and the masses of events can self-consistently classify detected sources into NSs and BHs (Farr et al. 2015; Mandel et al. 2015). However, we note that the feature that emerges in the mass distribution may not necessarily correspond to the boundary between NS and BH masses but instead may be complicated by accretion, hierarchical mergers, or primordial BHs (Carr et al. 2016; Yang et al. 2018). External priors on the NS maximum mass may also be applied, together with the population fit, in order to aid in the classification (Abbott et al. 2020d; Essick & Landry 2020).

### 5.2. Compact Object Merger Rates

Regardless of whether there exists a mass feature that naturally distinguishes between subpopulations of compact objects, our fit to the full mass distribution allows us to derive the compact object merger rate in different mass bins without explicitly counting the number of events in each category (Farr et al. 2015; Kapadia et al. 2020). Defining the BNS category as $1\,M_\odot < m_2 < m_1 < 2.5\,M_\odot$ and the BBH category as $5\,M_\odot < m_2 < m_1 < 100\,M_\odot$, we calculate the merger rate for each category by integrating the inferred rate density $dN/dV_c\,dt_s\,dm_1\,dm_2$ under the POWER LAW, POWER LAW + DIP, and POWER LAW + DIP + BREAK models within the specified ($m_1$, $m_2$) region.

The results are shown in Figure 7. In the POWER-LAW model, the BNS and BBH merger rates are closely correlated, but adding the features of the POWER LAW + DIP and POWER LAW + DIP +

BREAK models increases the BNS merger rate estimate while decreasing the BBH rate estimate. The BBH rate under the POWER-LAW model is $\mathcal{R}_{\mathrm{BBH}} = 77.7^{+60.6}_{-38.5}$ Gpc$^{-3}$ yr$^{-1}$, while the full POWER LAW + DIP + BREAK model yields $47.5^{+57.9}_{-28.8}$ Gpc$^{-3}$ yr$^{-1}$. Meanwhile, the inferred BNS merger rate is $\mathcal{R}_{\mathrm{BNS}} = 199^{+817}_{-173}$ Gpc$^{-3}$ yr$^{-1}$ under POWER LAW and $871^{+3015}_{-805}$ Gpc$^{-3}$ yr$^{-1}$ under POWER LAW + DIP + BREAK. The full POWER LAW + DIP + BREAK model better matches the rate estimates of Abbott et al. (2019a, 2019b), which assumed separate BNS and BBH populations, of $R_{\mathrm{BBH}} = 53.2^{+55.8}_{-28.2}$ and $\mathcal{R}_{\mathrm{BNS}} = 1540^{+3200}_{-1220}$ Gpc$^{-3}$ yr$^{-1}$. These trends for the astrophysical rates are consistent with the detection rates explored in Figure 5.

We can also extrapolate our models to calculate the rate in the NSBH category ($5\,M_\odot < m_1 < 100\,M_\odot$, $1\,M_\odot < m_2 < 2.5\,M_\odot$), although we caution that this is a significant extrapolation, since our simple pairing function of Equation (6) may not apply to the NSBH mass region. Because the GWTC-1 detections are all consistent with having equal component masses, our fits strongly disfavor unequal mass pairings and predict a low NSBH rate, with an upper 95% limit of 8.2 Gpc$^{-3}$ yr$^{-1}$.

We reiterate that our choice of mass bins to classify NS, mass gap, and BH sources and calculate the corresponding rates is only illustrative. Future detections will enable us to set the mass bins according to the measured feature in the mass distribution or external measurements of the NS maximum mass while accounting for uncertainty in the bin edges (Essick & Landry 2020). Additional detections will also allow us to meaningfully constrain the rate as a function of redshift (Fishbach et al. 2018; Abbott et al. 2019b).

### 5.3. Looking Ahead

As seen in Section 4, the preference for a dip/break between NS and BH masses can be understood by the dearth of detections between $\sim$2 and $8\,M_\odot$ relative to the number of detections below the purported mass gap (GW170817) and above the gap (10 BBHs) in GWTC-1. Recall that, defining the NS range as $m_1 \in [1, 2.5]\,M_\odot$, the mass-gap range as $m_1 \in [2.5, 5]\,M_\odot$, and the "low-mass" BH range as $m_1 \in [5, 10]\,M_\odot$, the POWER-LAW fit would have us detect $2.40^{+1.95}_{-1.02}$ mass-gap systems for every BNS system and one mass-gap system for every $2.33^{+1.92}_{-0.91}$ low-mass BH system (90% credibility). For a wider MG region, defined between 2.5 and $7\,M_\odot$, we expect $4.63^{+5.55}_{-2.33}$ mass-gap systems per BNS system and $1.39^{+0.79}_{-0.54}$ mass-gap systems per low-mass BH system ($7\,M_\odot < m_1 < 10\,M_\odot$).

While the current preference for a dip/break is suggestive, with only 11 events, the statistical uncertainties remain large. The situation will improve with future detections. For example, with 100 detections, the featureless POWER-LAW model would require at least $N_{\mathrm{BNS}} - 2$ detections with true primary masses in the mass range 2.5–5 $M_\odot$, where $N_{\mathrm{BNS}}$ is the number of detections in the range 1–2.5 $M_\odot$ (95% credibility). For a wider MG region between 2.5 and $7\,M_\odot$, POWER LAW predicts at least $N_{\mathrm{BNS}} + 1$ mass gap detections. Additionally, the POWER-LAW model predicts no more than nine BNS detections given a total of 100 detections. Observing a smaller number of mass-gap detections or a larger number of BNS detections would provide further evidence for a feature between the NS and BH mass range.

Of course, it is important to carry out the full population analysis that takes into account the measurement uncertainties of detected systems. Even in the presence of an absolute gap between NSs and BHs, some observed masses will be scattered into the gap due to noise fluctuations (Mandel et al. 2017; Fishbach et al. 2020). For the power-law type of distributions explored here, we expect $\sim$2% of detected BHs to lie within $1\sigma$ of the MG and $\sim$45% of detected NSs to lie within $1\sigma$ of the gap, assuming $1\sigma$ measurement uncertainties of 20%. For a gap width $w > \sigma \gtrsim 1\,M_\odot$, we expect that $\sim$50% of the masses close to the gap on either side will be erroneously observed within the gap, or $\sim$1% of BHs and $\sim$20% of the detected NS primary masses. If there exists an absolute mass gap, and we observe five detections with NS primary masses along with 50 detections with BH primary masses, we would expect $\sim$1.5 erroneous observations in the gap, while the expectation from a continuous power law would be $\sim$10 $\pm \sqrt{10}$ detections, making it possible to identify the presence of a mass gap at $\sim 3\sigma$ with fewer than 100 detections (see a similar argument in Mandel et al. 2015). These predictions are consistent with the population analysis of Mandel et al. (2017), which employed a clustering analysis on simulated data and found that $\sim$20 detections on either side of the mass gap would enable its confident detection.

In summary, the ability of future detections to precisely measure features in the mass spectrum depends on the depth and width of the purported mass gap, as well as the sharpness of the features relative to the uncertainty of the observed masses. If the features are sharp, we expect to converge on their location relatively quickly, scaling with the number of detections $N$ as $N^{-1}$, but if they are less abrupt, we expect to converge as $N^{-0.5}$ (Chakrabarty et al. 2003; Mandel et al. 2015).

The discussion throughout this paper has focused mainly on the one-dimensional primary mass distribution, because the GWTC-1 events all consist of nearly equal component masses. In the future, looking for structure in the two-dimensional mass distribution will be important for characterizing a potential population of NSBH systems. The approach described here to simultaneously fit the mass distribution of all compact objects in binary systems will allow us to explore how the component masses of NSBH systems relate to the component NSs and BHs found in BNS and BBH systems.

### 6. Conclusion

This work presents the first analysis to jointly fit the mass distribution of NSs and BHs in merging binary systems using data from the first two observing runs of Advanced LIGO/ Virgo. We assume no external knowledge of NS and BH subpopulations and ask whether GW170817, the least massive event detected, can be identified as an outlier in the BBH-only population based only on its mass, the property that is easiest to measure with GWs. We find that in the context of the BBH population, the masses of GW170817 are exceptional; based on the BBH-only fit, we would expect to detect an event with $m_1 < 2\,M_\odot$ out of a set of 11 events only 0.13% of the time.

We next try to fit a continuous power law across the entire mass range, finding that it is *possible* to extend the POWER-LAW fit to the BBH population of Abbott et al. (2019b) down to the masses of GW170817, but some tensions emerge. Namely, the power-law fit underpredicts the number of

detections in the BNS mass range while overpredicting the number of detections in the MG range.

While more events are required to judge whether the tensions in the POWER-LAW fit are statistically significant, we find that these tensions can be alleviated by allowing for a dip and/or break in the mass distribution between NS and BH masses. When we include the possibility of a dip, we find that a mass gap of width $>2\,M_\odot$ is preferred over an uninterrupted power law by a factor of 4.6 in GWTC-1. When we further allow the power law to take a different slope $\alpha_1$ at low (NS) masses, compared to $\alpha_2$ at high (BH) masses, we find that this is also preferred, with $\alpha_1$ steeper than $\alpha_2$ at 91% credibility.

Considering only the 11 GW events from O1+O2, we find preliminary evidence for two distinct populations of sources, with hints of a gap in between. Jointly fitting the mass distribution of NSs and BHs in binary systems allows us to self-consistently calculate merger rates in different categories and pool our knowledge regarding NS and BH masses across BNS, BBH, and NSBH systems. The methods described here will provide important insights going forward, especially in light of the latest discoveries from LIGO/Virgo. These include a high-mass BNS system with a total mass of $\sim3\,M_\odot$ (Abbott et al. 2020b) and a highly asymmetric BBH or NSBH system with a secondary mass of $\sim2.6\,M_\odot$ (Abbott et al. 2020d). With 100 events, as might be expected by the end of O3 or early in O4 (Abbott et al. 2018a), GW data alone may provide a clear indication of the existence of separate NS and BH populations, as well as important constraints on the existence and associated parameters of a mass gap between NSs and BHs.

## Appendix
## Analysis Details

We write the number density of sources as

$$\frac{dN}{dz\,dm_1\,dm_2} = \mathcal{N}p(z)p(m_1, m_2|\Lambda), \qquad (A1)$$

where the merger rate density is given by

$$\left.\frac{dN}{dm_1\,dm_2\,dV_c\,dt_s}\right|_z = \frac{dN}{dm_1\,dm_2\,dz} \times (T_{\text{obs}}/(1 + z))^{-1} \\ \times (dV_c/dz)^{-1}, \qquad (A2)$$

with $V_c$ the comoving volume, $t_s$ the time as measured in the source frame, and $T_{\text{obs}}$ the total observing time, or 169.7 days for O1+O2 (Abbott et al. 2019a). We assume that the rate density $\mathcal{R}$ is constant in redshift, consistent with the GWTC-1 detections (Fishbach et al. 2018; Abbott et al. 2019b), so that the normalization term $\mathcal{N}$ is related to the astrophysical merger rate $\mathcal{R}$ by

$$\mathcal{N} = T_{\text{obs}}\mathcal{R}\int_0^{z_{\text{max}}} dz(dV_c/dz)(1 + z)^{-1}, \qquad (A3)$$

and the redshift distribution is

$$p(z) = \frac{(dV_c/dz)(1 + z)^{-1}}{\int_0^{z_{\text{max}}} dz(dV_c/dz)(1 + z)^{-1}}. \qquad (A4)$$

For convenience, we denote $A(z_{\text{max}}) = \int_0^{z_{\text{max}}} dz(dV_c/dz)(1 + z)^{-1}$, and we pick $z_{\text{max}} = 1$, as no sources are detectable beyond this redshift for O1/O2 (Abbott et al. 2018a). Marginalizing away the normalization term $\mathcal{N}$ with a prior $p(\mathcal{N}) \propto \mathcal{N}^{-1}$, we obtain a posterior for the population hyperparameters ($\Lambda$) given the observed data $\{\mathcal{D}_i\}$ (Mandel et al. 2019),

$$p(\Lambda|\{\mathcal{D}_i\}) = p(\Lambda)\prod_i^N \frac{Z(\mathcal{D}_i|\Lambda)}{\beta(\Lambda)}, \qquad (A5)$$

where

$$Z(\mathcal{D}_i|\Lambda) \equiv \frac{1}{A(z_{\text{max}})}\int dm_1\,dm_2\,dz \\ \times \left[p(m_1, m_2|\Lambda) \times \left(\frac{1}{1 + z}\right)\left(\frac{dV_c}{dz}\right)p(\mathcal{D}_i|m_1, m_2, z)\right] \qquad (A6)$$

is the marginal likelihood for the $i$th event, and

$$\beta(\Lambda) \equiv \frac{1}{A(z_{\text{max}})}\int dm_1\,dm_2\,dz \\ \times \left[p(m_1, m_2|\Lambda) \times \left(\frac{1}{1 + z}\right)\left(\frac{dV_c}{dz}\right)P(\text{det}|m_1, m_2, z)\right] \qquad (A7)$$

is the fraction of events we expect to detect in a population described by $\Lambda$. The term $P(\text{det}|m_1, m_2, z)$ in Equation (A7) is the probability of detecting a specific system with component masses $m_1$ and $m_2$ at redshift $z$. We calculate this term following the semianalytic calculation described in Abbott et al. (2016, 2019b), which assumes that sources are detected if they have a single-detector signal-to-noise ratio $\rho > 8$, calculated with the Advanced LIGO Early High Sensitivity noise power spectral density (Abbott et al. 2018a).

Note that in addition to the fixed redshift distribution, we assume a fixed spin distribution (uniform spin magnitudes with isotropic orientations) and focus on only the mass distribution. As such, we neglect the possible impact of spins within the

selection function $P(\det|m_1, m_2, z)$, as detectability is predominantly determined by the component masses and the redshift.

Sampling from the posterior distribution of Equation (A5) produces our main results regarding the shape of the mass distribution. However, we can obtain estimates of the overall rate from

$$
\begin{aligned}
p(\mathcal{N}|\{\mathcal{D}\}) &= \int d\Lambda \, p(\mathcal{N}|\Lambda, \{\mathcal{D}_i\}) p(\Lambda|\{\mathcal{D}_i\}) \\
&= \int d\Lambda \, \frac{1}{\mathcal{N}} \mathcal{N}^{N_{\mathrm{obs}}} e^{-\mathcal{N}\beta(\Lambda)} p(\Lambda|\{\mathcal{D}_i\})
\end{aligned}
\tag{A8}
$$

and use Equation (A3) to convert $\mathcal{N}$ to the astrophysical merger rate $\mathcal{R}$. Here $N_{\mathrm{obs}}$ is the number of observations, and we have again assumed a prior $p(\mathcal{N}) \propto \mathcal{N}^{-1}$.

We estimate $Z(\mathcal{D}_i|\Lambda)$ by reweighting publicly available posterior samples (Abbott et al. 2019c) that were originally drawn assuming a prior $p_0(m_1, m_2, z)$,

$$
Z(\mathcal{D}_i|\Lambda) \propto \frac{1}{N_i} \sum_j^{N_i} \frac{p(m_1^{(j)}, m_2^{(j)}|\Lambda)(1 + z^{(j)})^{-1}(dV_c/dz|_{z^{(j)}})}{p_0(m_1^{(j)}, m_2^{(j)}, z^{(j)})},
\tag{A9}
$$

where $m_1^{(j)}, m_2^{(j)}, z^{(j)} \sim p(\mathcal{D}_i|m_1, m_2, z) p_0(m_1, m_2, z)$. For the samples of GWTC-1, the single-event sampling prior is (Abbott et al. 2019b)

$$
p_0(m_1, m_2, z) \propto d_L(z)^2 (1 + z)^2 \left( d_C(z) + \frac{(1 + z) d_H}{E(z)} \right),
\tag{A10}
$$

where $d_C$ is the comoving distance, $d_H = c/H_0$ is the Hubble distance, and $E(z) = H(z)/H_0$ (Hogg 1999).

When carrying out posterior predictive checks, we estimate the EDF by drawing samples from the single-event posterior. For each draw $\Lambda$ of the hyperposterior, we draw a sample from each of the 11 single-event posteriors, with weights $p(m_1, m_2|\Lambda) p(z)/p_0(m_1, m_2, z)$ (Galaudage et al. 2019; Fishbach et al. 2020; Miller et al. 2020).

Meanwhile, we calculate $\beta(\Lambda)$ with a Monte Carlo integral over a population of $N_{\mathrm{inj}} = 2^{26}$ simulated signals $m_1^{(j)}, m_2^{(j)}, z^{(j)} \sim p_{\mathrm{draw}}(m_1, m_2, z)$ (Tiwari 2018; Farr 2019),

$$
\begin{aligned}
\beta(\Lambda) = \frac{1}{N_{\mathrm{inj}}} \frac{1}{A(z_{\max})} \sum_k^{N_{\mathrm{inj}}} &\left[ P(\det|m_1^{(k)}, m_2^{(k)}, z^{(k)}) \right. \\
&\left. \times \frac{p(m_1^{(k)}, m_2^{(k)}|\Lambda)(1 + z^{(k)})^{-1}(dV_c/dz|_{z^{(k)}})}{p_{\mathrm{draw}}(m_1^{(k)}, m_2^{(k)}, z^{(k)})} \right],
\end{aligned}
\tag{A11}
$$

accounting for the uncertainty in the Monte Carlo integral according to Farr (2019). To sample efficiently, $p_{\mathrm{draw}}$ should resemble the true population distribution. We pick

$$
\begin{aligned}
p_{\mathrm{draw}}(m_1, m_2, z) \\
= \frac{1}{A(z_{\max})} (1 + z)^{-1} (dV_c/dz) p_{\mathrm{draw}}(m_1, m_2),
\end{aligned}
\tag{A12}
$$

with

$$
\begin{aligned}
p_{\mathrm{draw}}(m_1, m_2) \\
= \sum_{i=1}^{2} f_i \, p_{\mathrm{PL}}(m_1|\alpha_i, m_{\min} = 1, m_{\max}^i) p_{\mathrm{PL}} \\
\times (m_2|\alpha = 0, m_{\min} = 1, m_{\max} = m_1),
\end{aligned}
\tag{A13}
$$

with $f_i = [0.5, 0.5]$, $\alpha_i = [-4, -2]$, and $m_{\max}^i = [10, 100]$. We calibrate $\beta(\Lambda)$ to the actual selection function from the O1/O2 search (Usman et al. 2016; Sachdev et al. 2019) by dividing the above calculation of $\beta(\Lambda)$ by a constant factor of 1.7; see Figure 9 in Abbott et al. (2019b). Fishbach & Holz (2020) showed that this constant calibration factor is sufficient in recovering the population results of Abbott et al. (2019b). This same set of injections is used to estimate PPDs, which are found by reweighting the injections according to $p(m_1, m_2, \Lambda)/p_{\mathrm{draw}}(m_1, m_2)$, with $\Lambda$ drawn from the hyperposterior of a given model. We sample from the population hyperposterior using PyMC3 (Salvatier et al. 2016).

## ORCID iDs

Maya Fishbach ⓘ https://orcid.org/0000-0002-1980-5293
Reed Essick ⓘ https://orcid.org/0000-0001-8196-9267
Daniel E. Holz ⓘ https://orcid.org/0000-0002-0175-5064

## References

Aasi, J., Abadie, J., Abbott, B. P., et al. 2015, CQGra, 32, 074001
Abbott, B., Abbott, R., Abbott, T. D., et al. 2020b, ApJL, 892, L3
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvX, 6, 041015
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018a, LRR, 21, 3
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018b, PhRvL, 121, 161101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019a, PhRvX, 9, 031040
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019b, ApJL, 882, L24
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2020a, CQGra, 37, 045006
Abbott, R., Abbott, T. D., Abraham, S., et al. 2019c, arXiv:1912.11716
Abbott, R., Abbott, T. D., Abraham, S., et al. 2020c, arXiv:2004.08342
Abbott, R., Abbott, T. D., Abraham, S., et al. 2020d, ApJL, 896, L44
Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, CQGra, 32, 024001
Ade, P. A. R., Aghanim, N., Arnaud, M., et al. 2016, A&A, 594, A13
Agathos, M., del Pozzo, W., Li, T. G. F., et al. 2014, PhRvD, 89, 082001
Alsing, J., Silva, H. O., & Berti, E. 2018, MNRAS, 478, 1377
Antoniadis, J., Tauris, T. M., Ozel, F., et al. 2016, arXiv:1605.01665
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Belczynski, K., Wiktorowicz, G., Fryer, C. L., Holz, D. E., & Kalogera, V. 2012, ApJ, 757, 91
Breivik, K., Chatterjee, S., & Andrews, J. J. 2019, ApJL, 878, L4
Carr, B., Kühnel, F., & Sandstad, M. 2016, PhRvD, 94, 083504
Chakrabarty, D., Morgan, E. H., Muno, M. P., et al. 2003, Natur, 424, 42
Chatziioannou, K., & Farr, W. M. 2020, arXiv:2005.00482
Chen, H.-Y., & Chatziioannou, K. 2020, ApJL, 893, L41
Doctor, Z., Wysocki, D., O'Shaughnessy, R., Holz, D. E., & Farr, B. 2020, ApJ, 893, 35
Essick, R., & Landry, P. 2020, arXiv:2007.01372
Essick, R., Tews, I., Landry, P., Reddy, S., & Holz, D. E. 2020, arXiv:2004.07744
Farr, W. M. 2019, RNAAS, 3, 66
Farr, W. M., & Chatziioannou, K. 2020, RNAAS, 4, 65
Farr, W. M., Gair, J. R., Mandel, I., & Cutler, C. 2015, PhRvD, 91, 023005
Farr, W. M., Sravan, N., Cantrell, A., et al. 2011, ApJ, 741, 103
Farrow, N., Zhu, X.-J., & Thrane, E. 2019, ApJ, 876, 18
Fasano, M., Wong, K. W. K., Maselli, A., et al. 2020, arXiv:2005.01726
Fishbach, M., Farr, W. M., & Holz, D. E. 2020, ApJL, 891, L31
Fishbach, M., & Holz, D. E. 2017, ApJL, 851, L25
Fishbach, M., & Holz, D. E. 2020, ApJL, 891, L27
Fishbach, M., Holz, D. E., & Farr, W. M. 2018, ApJL, 863, L41

Flanagan, E. E., & Hinderer, T. 2008, PhRvD, 77, 021502
Foreman-Mackey, D. 2016, JOSS, 1, 24
Fryer, C. L., & Kalogera, V. 2001, ApJ, 554, 548
Galaudage, S., Talbot, C., & Thrane, E. 2019, arXiv:1912.09708
Hannam, M., Brown, D. A., Fairhurst, S., Fryer, C. L., & Harry, I. W. 2013, ApJL, 766, L14
Hogg, D. W. 1999, arXiv:astro-ph/9905116
Jeffreys, H. 1961, Theory of Probability (3rd ed.; Oxford: Oxford Univ. Press)
Kapadia, S. J., Caudill, S., Creighton, J. D. E., et al. 2020, CQGra, 37, 045007
Kiziltan, B., Kottas, A., de Yoreo, M., & Thorsett, S. E. 2013, ApJ, 778, 66
Kovetz, E. D., Cholis, I., Breysse, P. C., & Kamionkowski, M. 2017, PhRvD, 95, 103010
Kreidberg, L., Bailyn, C. D., Farr, W. M., & Kalogera, V. 2012, ApJ, 757, 36
Lackey, B. D., & Wade, L. 2015, PhRvD, 91, 043002
Landry, P., Essick, R., & Chatziioannou, K. 2020, PhRvD, 101, 123007
Lattimer, J. M. 2012, ARNPS, 62, 485
Littenberg, T. B., Farr, B., Coughlin, S., Kalogera, V., & Holz, D. E. 2015, ApJL, 807, L24
Loredo, T. J. 2004, in AIP Conf. Ser. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint (Melville, NY: AIP), 195
Mandel, I. 2010, PhRvD, 81, 084029
Mandel, I., Farr, W. M., Colonna, A., et al. 2017, MNRAS, 465, 3254
Mandel, I., Farr, W. M., & Gair, J. R. 2019, MNRAS, 486, 1086
Mandel, I., Haster, C.-J., Dominik, M., & Belczynski, K. 2015, MNRAS, 450, L85
Margalit, B., & Metzger, B. D. 2017, ApJL, 850, L19

Miller, M. C., Chirenti, C., & Lamb, F. K. 2019, ApJ, 888, 12
Miller, S., Callister, T. A., & Farr, W. M. 2020, ApJ, 895, 128
Nitz, A. H., Capano, C., Nielsen, A. B., et al. 2019, ApJ, 872, 195
Nitz, A. H., Dent, T., Davies, G. S., et al. 2020, ApJ, 891, 123
Özel, F., & Freire, P. 2016, ARA&A, 54, 401
Özel, F., Psaltis, D., Narayan, R., & McClintock, J. E. 2010, ApJ, 725, 1918
Özel, F., Psaltis, D., Narayan, R., & Santos Villarreal, A. 2012, ApJ, 757, 55
Read, J. S., Baiotti, L., Creighton, J. D. E., et al. 2013, PhRvD, 88, 044042
Rezzolla, L., Most, E. R., & Weih, L. R. 2018, ApJL, 852, L25
Roulet, J., & Zaldarriaga, M. 2019, MNRAS, 484, 4216
Sachdev, S., Caudill, S., Fong, H., et al. 2019, arXiv:1901.08580
Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, PeerJ Computer Science, 2, e55
Shibata, M., Zhou, E., Kiuchi, K., & Fujibayashi, S. 2019, PhRvD, 100, 023015
Thompson, T. A., Kochanek, C. S., Stanek, K. Z., et al. 2019, Sci, 366, 637
Thrane, E., & Talbot, C. 2019, PASA, 36, e010
Tiwari, V. 2018, CQGra, 35, 145009
Usman, S. A., Nitz, A. H., Harry, I. W., et al. 2016, CQGra, 33, 215004
Valentim, R., Rangel, E., & Horvath, J. E. 2011, MNRAS, 414, 1427
Venumadhav, T., Zackay, B., Roulet, J., Dai, L., & Zaldarriaga, M. 2020, PhRvD, 101, 083030
Weibull, W. 1951, JAM, 18, 293
Wysocki, D., Lange, J., & O'Shaughnessy, R. 2019, PhRvD, 100, 043012
Wysocki, D., O'Shaughnessy, R., Wade, L., & Lange, J. 2020, arXiv:2001.01747
Yang, H., East, W. E., & Lehner, L. 2018, ApJ, 856, 110