What You Don't Know Can Hurt You: Use and Abuse of Astrophysical Models in Gravitational-wave Population Analyses

April Qiu Cheng[1,2] , Michael Zevin[3,4,5] , and Salvatore Vitale[1,2]

[1] LIGO Laboratory, Massachusetts Institute of Technology, 185 Albany St, Cambridge, MA 02139, USA; aqc@mit.edu
[2] Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA
[3] Kavli Institute for Cosmological Physics, The University of Chicago, 5640 South Ellis Ave, Chicago, IL 60637, USA
[4] Enrico Fermi Institute, The University of Chicago, 933 East 56th St, Chicago, IL 60637, USA

Abstract

One of the goals of gravitational-wave astrophysics is to infer the number and properties of the formation channels of binary black holes (BBHs); to do so, one must be able to connect various models with the data. We explore benefits and potential issues with analyses using models informed by population synthesis. We consider five possible formation channels of BBHs, as in Zevin et al. (2021b). First, we confirm with the GWTC-3 catalog what Zevin et al. (2021b) found in the GWTC-2 catalog, i.e., that the data are not consistent with the totality of observed BBHs forming in any single channel. Next, using simulated detections, we show that the uncertainties in the estimation of the branching ratios can shrink by up to a factor of ~ 1.7 as the catalog size increases from 50 to 250, within the expected number of BBH detections in LIGO–Virgo–KAGRA's fourth observing run. Finally, we show that this type of analysis is prone to significant biases. By simulating universes where all sources originate from a single channel, we show that the influence of the Bayesian prior can make it challenging to conclude that one channel produces all signals. Furthermore, by simulating universes where all five channels contribute but only a subset of channels are used in the analysis, we show that biases in the branching ratios can be as large as $\sim 50\%$ with 250 detections. This suggests that caution should be used when interpreting the results of analyses based on strongly modeled astrophysical subpopulations.

1. Introduction

Gravitational waves (GWs) emitted by the mergers of compact objects, neutron stars and black holes, encode the properties of their sources including masses, spins, and distance. When one has a large enough data set, information can be combined from the detected sources to infer properties of the underlying astrophysical process—or processes—that created them. Nearly 100 compact binary coalescences[6] (the large majority of which are BBHs) have been revealed in the data of ground-based GW detectors, LIGO (LIGO Scientific Collaboration et al. 2015) and Virgo (Acernese et al. 2015), up to their third observing run (O3) (The LIGO Scientific Collaboration et al. 2021b; Olsen et al. 2022; Nitz et al. 2023), allowing this type of analysis to be performed. The formation scenarios for compact binaries can be broadly separated in two categories: isolated evolution in the galactic field and dynamical assembly in dense environments such as clusters and AGN disks (see, e.g., Mapelli 2021; Mandel & Farmer 2022, for reviews). The overall data set might contain BBHs formed from a combination of these and other evolutionary pathways.

Ideally, one would like to fully characterize any astrophysical formation channels that contribute sources, as well as their relative abundances (branching ratios). In practice, several approaches have been proposed and followed in the literature. They all have merits and shortcomings, and we quickly review them here, focusing on BBHs, which are the topic of our work.

1. *Heuristic models*—The most straightforward analyses rely on heuristic parametric distributions to describe the astrophysical distributions of black hole parameters. For example, the primary (i.e., most massive) black hole mass distribution can be modeled as a mixture model of a power law and a Gaussian (the `power law + peak` model of Abbott et al. 2023, originally introduced by Talbot & Thrane 2018); the spin orientation distribution as a mixture model of an isotropic component and a component nearly aligned with the orbital angular momentum (Talbot & Thrane 2017; Vitale et al. 2017b); etc. The functional forms might be chosen based on computational expediency, or they could be inspired by reasonable astrophysical expectations (e.g., a power-law component in the black hole mass function because the masses of progenitor stars are distributed that way; Kroupa 2001). Meanwhile, mixture models might allow for different subpopulations to be accounted for. The main potential shortfall of this approach is that, if the models are very strong, the resulting posteriors might actually be model-driven, especially for hard-to-measure parameters. This has been shown, for example, in the context of the spin magnitude measurement (Galaudage et al. 2021; Roulet et al. 2021; Callister et al. 2022) and

---

[5] NASA Hubble Fellow.

[6] The exact number depends on how conservative of a detection threshold one uses.

the spin orientation (Vitale et al. 2022a). On the positive side, if a parameter can be reliably measured, it often has a clear connection with a meaningful astrophysical quantity (e.g., the slope of a mass power law). Heuristic models also allow correlations between parameters to be probed in a straightforward manner (Callister et al. 2021; Adamcewicz & Thrane 2022; Biscoveanu et al. 2022; Vitale et al. 2022a; Baibhav et al. 2023).

2. *Flexible models*—Flexible model approaches have been proposed as a way of avoiding the risk of forcing features into the data. In such models, 1D posterior distributions (usually fully marginalized, e.g., $p(m_1|d)$), are modeled as splines, Gaussian processes, or using autoregression (Mandel et al. 2017; Vitale et al. 2019; Tiwari 2021, 2022; Veske et al. 2021; Edelman et al. 2022; Edelman et al. 2023; Golomb & Talbot 2022; Rinaldi & Del Pozzo 2022; Callister & Farr 2023). While these approaches are less likely to impose stringent features into the posteriors (though see, e.g., Farah et al. 2023), the number of unknown parameters is typically larger than for heuristic models, and the model parameters are not usually associated to any specific astrophysical quantity. Furthermore, these methods are not well suited to disentangle subpopulations, and are instead better suited to measuring the overall distribution of parameters.

3. *Astrophysically informed models*—Finally, one may use models obtained directly from the output of population synthesis. Given a set of initial conditions and choices for uncertain stellar, binary, and environmental physical parameters, such models make predictions for the anticipated underlying and detectable distribution of compact binaries. These models can be parameterized directly in terms of physically meaningful quantities (e.g., the onset and evolution of binary mass transfer phases, the strength of supernova natal kicks, the efficiency of angular momentum transport), and correlations between parameters are automatically built in the models, both of which are very strong positive factors. In practice, the range of variations in physical uncertainties, as well as the number and complexity of the differing channels one considers, are often limited by the availability of numerical simulations that thoroughly explore the variation of the output (e.g., distribution of spin magnitude in the population) when one of the physical input parameters (e.g., efficiency of common envelope evolution) is varied. This is due to the fact that most population synthesis algorithms require significant computational resources to run, which implies that they cannot be evaluated "on the fly" for any value of their input parameters, but must instead, for example, be evaluated on a sparse grid. Recent efforts have begun to more thoroughly explore compact binary population predictions for individual formation channels over an expansive array of physical and environmental uncertainties (e.g., Broekgaarden et al. 2022). Consideration of multiple formation channels simultaneously and self-consistently proves more difficult, given the diversity of codebases needed to model different channels and the unique physics that affects compact binary populations from each channel. The most expansive multichannel analysis to date was performed in Zevin et al. (2021a), who considered five possible formation channels (see

Section 2.1 below), parameterized by the spin of quasi-isolated black holes at birth (a proxy for the efficiency of angular momentum transport in massive stars) and the efficiency of the common envelope ejection. By analyzing the 45 confident BBH sources of the penultimate (GWTC-2) LIGO–Virgo–KAGRA collaboration (LVK) catalog, Zevin et al. (2021a) found that the data required more than a single formation channel in order to explain the diversity of GW events and the distribution of parameters of the detected binaries. They also found that the data preferred small natal black hole spins, consistent with the fact that most of the LVK BBHs have small spin magnitudes.

In this paper, we analyze both merits and shortcoming of approaches based on models informed by the output of population synthesis codes. First, we repeat the analysis of Zevin et al. (2021a) on the latest LVK catalog (GWTC-3), which comprises 69 BBH with a false-alarm rate of less than $1\,\mathrm{yr}^{-1}$. We find that the model for common envelope evolution can explain up to 95% of the BBHs in the underlying population (while contributing up to 37% of the *detectable* BBHs).

Then, we create catalogs of simulated BBH signals with parameters drawn from our models, for some assumed values of the common envelope efficiency, quasi-isolated natal black hole spins, and branching fractions across channels. We analyze the performance of the analysis as the number of BBH sources in the catalog increases from 50 to 250. We find that, for channels that produce higher number of detectable sources (and hence are proportionally more represented in the catalog, even if the true underlying branching fractions are not higher), the uncertainty on the underlying branching fraction can improve by a factor of up to ∼1.7 as the number of sources increases to 250.

Next, we study the biases that can be introduced in this type of inference if the analysis is not using a suite of models fully representative of what is actually realized in nature. This problem was first indirectly shown by Franciolini et al. (2022), who added a channel for primordial black hole formation to those used by Zevin et al. (2021a), and obtained that the inference on the fraction of primordial back holes was significantly affected by which of the *other* channels were included in the analysis. To do so, we generate mock universes where each of the five channels contributes some known fraction of the underlying population, and run the inference excluding in turn one of the five models. We show how this introduces biases in the inference of the remaining four channels' branching fractions. The channels that are most heavily biased are the ones that can most easily produce sources similar to the one channel excluded from the analysis, as well as those with the lowest detection efficiencies. Finally, we generate universes where the totality of the BBH sources are produced by one of the five channels, and run the analysis with the same five models. We show that, while the natal spin can be inferred correctly, it is usually not possible to exclude that more than one channel contributes to the population after 100 events, a caveat to our result from our inference on GWTC-3 data that multiple channels contribute to both the underlying and detected BBH population.

The rest of the paper is organized as follows: in Section 2, we review the basics of hierarchical inference and the models

used in this work, and apply these tools to GWTC-3's BBHs. In Section 3, we apply the method to different simulated catalogs in the ideal scenario where models and true populations match. In Section 4, we focus on biases in the inference. We conclude in Section 5.

2. Hierarchical Inference on GWTC-3 Data

2.1. Hyperinference Method

We use hierarchical Bayesian inference on the branching fractions between different astrophysical formation channels of BBHs. Our methods mostly follow the analysis developed by Zevin et al. (2021a), adapting their codebase, Astrophysical Model Analysis and Evidence Evaluation (AMA $\mathcal{Z}$ E), for our work. Here, we outline the essentials of this method, as well as the key differences.

We consider five different formation channels: three isolated evolution (field) channels and two dynamical formation channels. The common envelope (CE) (e.g., Paczynski 1976; van den Heuvel 1976; Tutukov & Yungelson 1993; Bethe & Brown 1998; Belczynski et al. 2002, 2016; Dominik et al. 2012; Eldridge & Stanway 2016; Stevenson et al. 2017; Giacobbo & Mapelli 2018) and stable mass transfer (SMT) (van den Heuvel et al. 2017; Neijssel et al. 2019; Gallegos-Garcia et al. 2021; van Son et al. 2022) scenarios are field channels that involve unstable and stable mass transfer, respectively, following the formation of the first black hole. In the chemically homogeneous evolution (CHE) channel, stars in a close, tidally locked binary rotate rapidly, causing temperature gradients that lead to efficient mixing of the stars' interiors. The stars do not undergo significant expansion of the envelope, preventing significant post-main-sequence wind mass loss and premature merging, resulting in higher-mass BBHs (de Mink & Mandel 2016; Mandel & de Mink 2016; Marchant et al. 2016). Finally, the two dynamical formation channels lead to merging black holes via strong gravitational encounters that harden the binary in cluster cores (e.g., McMillan et al. 1991; Hut et al. 1992; Sigurdsson & Phinney 1993; Portegies Zwart & McMillan 2000; Miller & Hamilton 2002; Gültekin et al. 2006; O'Leary et al. 2006; Fregeau & Rasio 2007; Downing et al. 2010; Samsing et al. 2014; Ziosi et al. 2014; Rodriguez et al. 2015, 2016; Antonini & Rasio 2016; Askar et al. 2017; Samsing & Ramirez-Ruiz 2017; Kremer et al. 2020), where heavy black holes migrate toward due to dynamical friction (Lightman & Shapiro 1978; Sigurdsson & Hernquist 1993); we consider dynamical formation of BBHs in globular clusters (GCs) and nuclear star clusters (NSCs). See Zevin et al. (2021a) for a more detailed description of the astrophysical models considered in this work.

Each of these channels is modeled to predict the four-dimensional distribution of the BBHs it forms with parameters $\theta = [\mathcal{M}_c, q, \chi_{eff}, z]$, where $\mathcal{M}_c$ is the source-frame chirp mass, q is the mass ratio (defined to be $0 < q \leqslant 1$), χ_{eff} is the effective dimensionless spin parameter, and z is the redshift; these are constructed into a probability distribution using a four-dimensional kernel density estimate (KDE) bounded by the physical constraints of each parameter. The models also depend on two additional parameters encoding uncertainties in the physical prescription: χ_b, the dimensionless spin of a black hole formed in quasi-isolation directly following core collapse, and α_{CE}, which parameterizes the efficiency of common

envelope ejection (see, e.g., Ivanova et al. 2013).[7] We note that the choice of natal black hole spin χ_b does not set all black holes in a given population to merge with this exact spin, because tidal spin-up processes (Qin et al. 2018; Zaldarriaga et al. 2018; Bavera et al. 2021) and hierarchical mergers (Pretorius 2005; González et al. 2007; Buonanno et al. 2008) can increase the spin of black holes that participate in BBH mergers. We assume these two parameters take on a grid of discrete values ($\chi_b \in [0.0, 0.1, 0.2, 0.5]$, $\alpha_{CE} \in [0.2, 0.5, 1.0, 2.0, 5.0]$) over which we compute the models, although α_{CE} only affects the CE channel in our models. A plot of the detection-weighted marginalized model KDEs for $\alpha_{CE} = 1.0$, $\chi_b = 0.2$ is shown in Figure 1 as an example. Further details, including formation models, detection weighting, and mathematical framework, can be found in Zevin et al. (2021a).

Overall, we perform hierarchical inference on seven hyperparameters[8] $\Lambda = [\beta, \chi_b, \alpha_{CE}]$, where $\beta = [\beta_{CE}, \beta_{CHE}, \beta_{SMT}, \beta_{GC}, \beta_{NSC}]$ are the five astrophysical formation channel branching fractions. The steps involved in performing hierarchical inference on GW populations given a set of posterior samples ($\theta = [\mathcal{M}_c, q, \chi_{eff}, z]$ in our case) for N_{obs} sources in the presence of selection effects have been thoroughly discussed in the literature (Farr et al. 2015; Mandel et al. 2019; Thrane & Talbot 2019; Vitale et al. 2022b; Abbott et al. 2023), and we therefore do not review them here. As in Zevin et al. (2021a), we use a broad, agnostic prior for the hyperparameters: a flat symmetric Dirichlet prior for β and a uniform prior for χ_b and α_{CE} over the allowed discrete values. The outputs of the inference are samples of the hyperposterior $p(\beta|\chi_b, \alpha_{CE})$ over the grid of hypermodels (i.e., allowed values of χ_b and α_{CE}), from which we can compute the marginalized hyperposterior $p(\beta)$ as well as the Bayes factors BF_b^a of hypermodel a compared to hypermodel b. Bayes factors >1 indicate that model a is more supported than model b, with $BF_b^a > 10$ (100) indicating strong (decisive) support for model a over model b (Kass & Raftery 1995).

We also recover the *detectable* branching fractions β^{det}, which represent the fraction of detectable BBHs originating from each channel. These are computed by rescaling each underlying branching fraction by its detection efficiency, defined as $\xi_j^{\chi_b,\alpha_{CE}} = \int P_{det}(\theta)p(\theta|\mu_j^{\chi_b,\alpha_{CE}})\,d\theta$, where $P_{det}(\theta)$ is the probability of detecting a BBH with parameters θ and $p(\theta|\mu_j^{\chi,\alpha})$ is the probability of formation channel j producing a BBH with parameters θ, dependent on the model μ_j and choice of physical prescription χ_b and α_{CE} (Zevin et al. 2021a). We apply this method to both real (Section 2.2) and simulated (Sections 3 and 4) BBH data.

2.2. Application to GWTC-3

We extend the work of Zevin et al. (2021a) by applying the inference to confident BBH detections up to the second half of the third observing run (O3b). We use the publicly released priors and posterior samples from the GWTC-2.1 (The LIGO Scientific

---

[7] The choice of these hyperparameters come from the availability of self-consistent astrophysical models; in particular, the globular cluster simulations were run on a grid of discrete values in χ_b, and α_{CE} variations only affect the parameter distributions of the CE model, which incurred less computational cost to explore additional variations than other models. See Zevin et al. (2021a) for further discussion on the reasoning and motivations for the physical prescriptions considered.

[8] Due to the restriction $\sum_i \beta_i = 1$ (the branching fractions must add up to 1), we technically only perform inference on six hyperparameters.
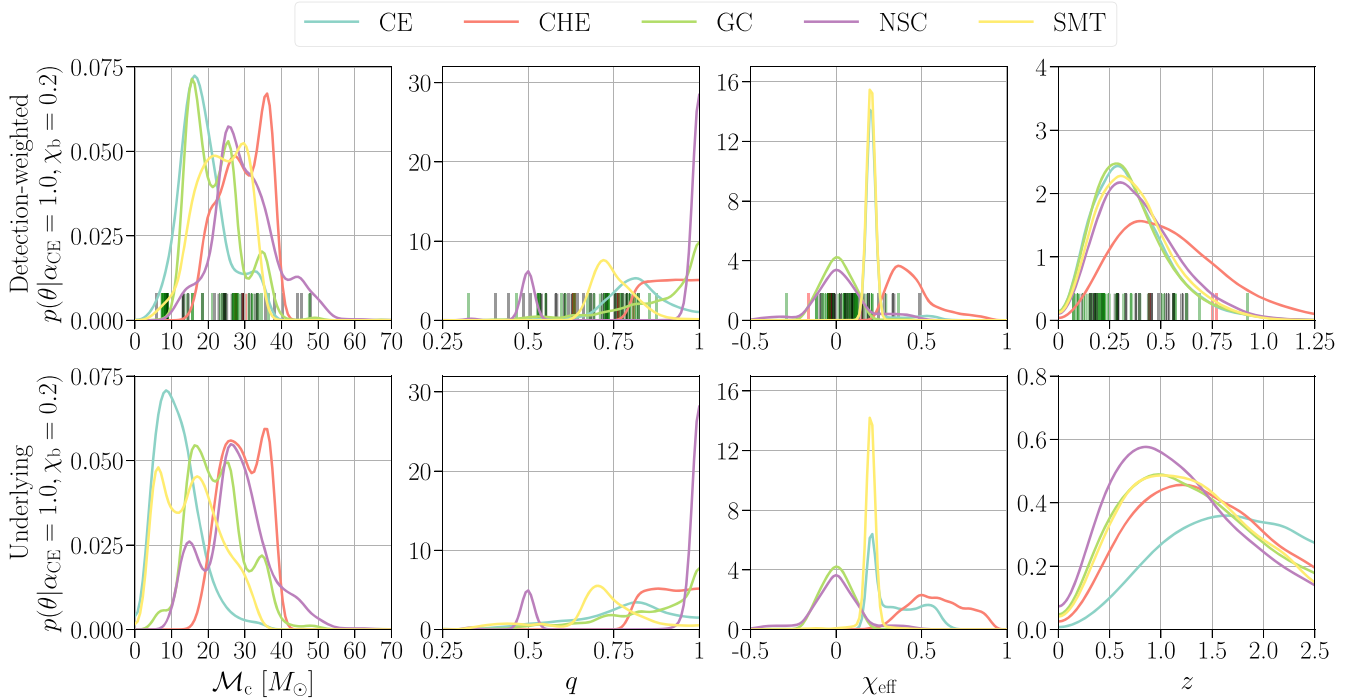
Figure 1. Detection-weighted (top) and underlying (bottom) marginalized KDEs of our four BBH parameters, $\theta = [\mathcal{M}_c, q, \chi_{eff}, z]$, for our five different formation channels with $\alpha_{CE} = 1.0$, $\chi_b = 0.2$. It is worth noting the unique features of each formation channel and their dependence on the chosen model $\chi_b = 0.2$: while the χ_{eff} marginalized KDE for field channels CE and SMT peak at $\chi_{eff} = \chi_b = 0.2$, with the CE channel having support for larger χ_{eff} through tidal spin-up, dynamical channels GC and NSC peak at $\chi_{eff} = 0$, due to the independently isotropically oriented spins of the BBHs with symmetric wings to more extreme χ_{eff} values from hierarchical mergers, and the CHE channel has χ_{eff} exceeding $\chi_b = 0.2$, due to significant tidal spin-up of both black hole progenitors. The first row is the same as the third row of Figure 1 from Zevin et al. (2021a), except here we have calculated the detection weighting with O4 sensitivities, and the black ticks show the median values of the parameter estimation posteriors for each event in GWTC-2, while the green and red ticks mark the events added to and removed from, respectively, this updated analysis that includes GWTC-3.

Collaboration et al. 2021a) and GWTC-3 (The LIGO Scientific Collaboration et al. 2021b) analyses; detection probabilities are calculated for the LIGO-Hanford, LIGO-Livingston, and VIRGO network operating at `midhighlatelow` sensitivities (Abbott et al. 2018). There are two key differences between our analysis and that of Zevin et al. (2021a). First, we apply a more stringent detection threshold of FAR $\leqslant 1\,\mathrm{yr}^{-1}$ (Abbott et al. 2023); therefore, events GW190424_180648, GW190514_065416, and GW190909_114149, which were used in the previous analysis, are now excluded. Second, we evaluate the prior at the posterior points analytically rather than by constructing a KDE from prior samples; for further discussion on this point, see Appendix A. Finally, as in Zevin et al. (2021a), we opt to exclude GW190521 and GW190814 from the analysis, as their posteriors extend significantly to regions of our model KDEs with little to no support. Furthermore, in the case of GW190814, it is uncertain whether this system is comprised of two black holes or a neutron star and a black hole. We find that the inclusion of GW190521 does not significantly affect our results; see Franciolini et al. (2022) for an analysis that includes GW190521. Overall, we do hierarchical inference on 68 BBHs, compared to the 45 BBHs considered in Zevin et al. (2021a). Unless otherwise specified, we report results as median and 90% symmetric credible intervals.

Figure 2 shows the posterior distributions on the underlying branching fraction β, including detections from O3b; the same plot but for the detectable branching fractions can be found in Appendix B. We find $\beta_{CE} = 90^{+5.0}_{-11}\%$, $\beta_{CHE} = 0.6^{+1.0}_{-0.5}\%$, $\beta_{GC} = 3.7^{+7.6}_{-3.4}\%$, $\beta_{NSC} = 1.1^{+1.7}_{-0.8}\%$, and $\beta_{SMT} = 4.0^{+6.5}_{-3.2}\%$, indicating strong support for the CE channel dominating the underlying astrophysical population in our set of models.

However, there is comparable contribution from all five channels to the *detectable* BBH population: $\beta^{det}_{CE} = 21^{+16}_{-10}\%$, $\beta^{det}_{CHE} = 6.9^{+8.5}_{-5.2}\%$, $\beta^{det}_{GC} = 24^{+26}_{-21}\%$, $\beta^{det}_{NSC} = 20^{+17}_{-13}\%$, and $\beta^{det}_{SMT} = 24^{+24}_{-19}\%$. We attribute this difference to the fact that, compared to other channels, the CE channel produces less massive black holes distributed at higher redshifts, which therefore are harder to detect; see Figure 1. With 90% (99%) credibility, no single formation channel contributes to more than 49% (61%) of the detectable BBH population. Additionally, over 98% of posterior samples have significant ($\beta^{det} > 10\%$) contributions from three or more different formation channels. Overall, we find that the CE channel contributes the most to the underlying BBH population; however, as we discuss in Section 4, the posterior for β_{CE} is also the most uncertain and prior dominated. We additionally find that a mix of formation channels contributes to the detectable BBH population. These results, as well as the overall shapes of the branching fraction posteriors, are consistent with the analysis with GWTC-2 data (Zevin et al. 2021a), which found $\beta_{CE} = 71^{+19}_{-60}\%$ and that no single channel contributed to more than 70% of the detectable BBH population with 99% confidence. No strong correlations are apparent upon examining a corner plot of the branching fractions.

Turning to the selection of physical prescription hyperparameters χ_b and α_{CE}, we find no posterior support for models with $\chi_b > 0.1$; there is no significant preference between the $\chi_b = 0.0$ and $\chi_b = 0.1$ models, with $\mathrm{BF}^{\chi_b=0.1}_{\chi_b=0.0} = 1.18$. We favor high common envelope efficiencies, with $\mathrm{BF}^{\alpha_{CE}=5.0}_{\alpha_{CE}=1.0} = 249$, and strongly disfavor low common envelope efficiencies, with
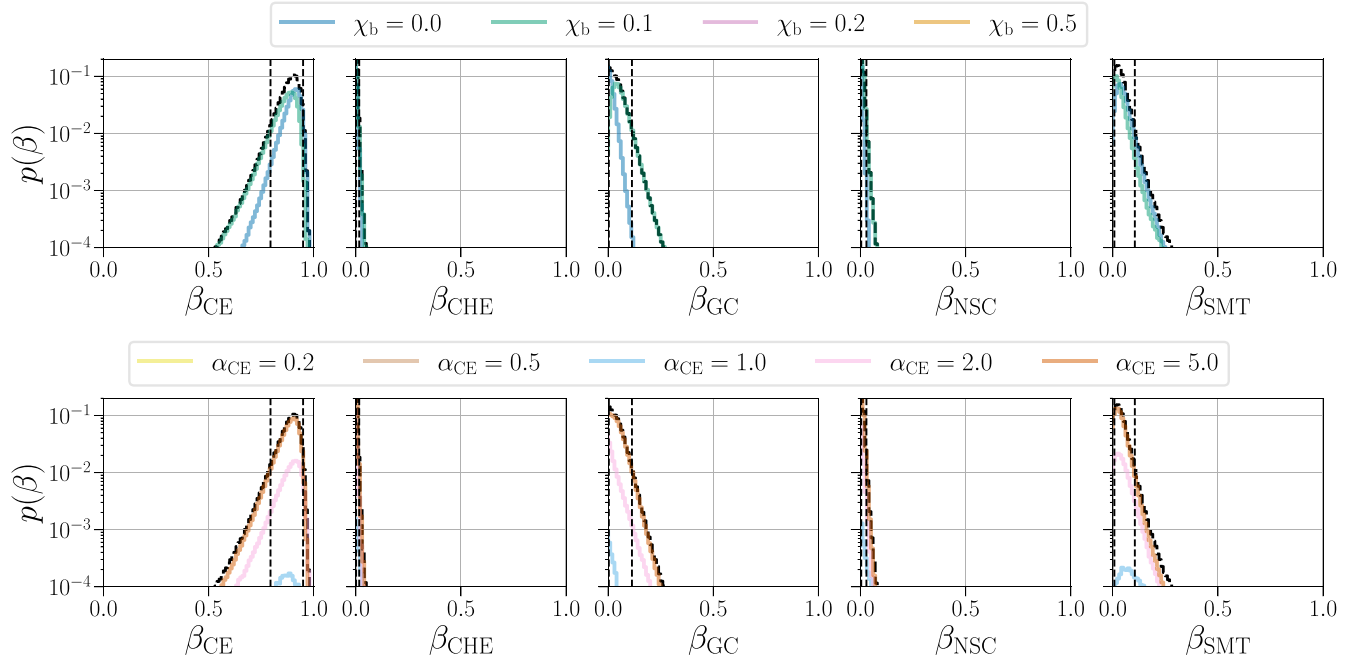
Figure 2. Marginalized branching fractions inferred from 68 BBHs from the cumulative GWTC-3 for each of the five formation channels. Colored curves show the contributions from different values of χ_b (top row) and α_{CE} (bottom row), marginalized over α_{CE} and χ_b, respectively; the black dotted curve shows the total contribution from all hypermodels. The log-scaling of the y-axis should be noted; the 90% symmetric credible intervals are marked by vertical lines.

$BF^{\alpha_{CE}=0.2}_{\alpha_{CE}=1.0} = 5 \times 10^{-5}$. These results are also consistent with Zevin et al. (2021a). The main difference is that we obtain stronger constraints, both in singling out preferred hypermodels and in the uncertainties (widths) of the branching fraction hyperposteriors. We attribute this effect to the increase in sample size from GWTC-2 to GWTC-3, as well as the difference in method in evaluating the prior at the posterior points during the inference; see Appendix A for further discussion.

3. Projections for Future Catalogs

3.1. Method

Next, we perform the same inference using simulated BBH observations in a universe where the BBH population exactly follows our models. The motivation for this analysis is to to quantify how uncertainties in the hyperposterior scale with the number of observed events.

We first create a mock "universe," i.e., a set of true values for our hyperparameters $\Lambda = [\beta, \chi_b, \alpha_{CE}]$. From each formation channel j, we draw $n_j = \beta_j n$ BBHs from the population model $p(\theta|\mu_j^{\chi_b, \alpha_{CE}})$, for a total of $n = 5 \times 10^4$ BBHs that form our underlying population. Next, we draw from this underlying population, assigning extrinsic parameters (sky location and inclination) from an isotropic distribution. For each BBH system, we calculate its optimal signal-to-noise ratio ρ_{opt} assuming a network consisting of LIGO-Hanford, LIGO-Livingston, and Virgo operating at O4 low (LIGO) and high (Virgo) sensitivities (Abbott et al. 2018; O'Reilly et al. 2022), and keep only the mock signals with $\rho_{opt} \geqslant 11$; we repeat this process until we have a mock catalog of N_{obs} detections. Then, we perform parameter estimation on the N_{obs} BBHs. For both the S/N calculation and parameter estimation, we use the Bayesian inference software `bilby` (Ashton et al. 2019) and the `IMRPhenomXP` waveform approximant (Pratten et al. 2021). Finally, we use these posterior samples for the

Table 1
Mock Universes

Sections	N_{obs}	β_{CE}	β_{CHE}	β_{GC}	β_{NSC}	β_{SMT}
3.2, 4.3	up to 250	40%	5%	25%	5%	25%
3.2, 4.2, 4.3	up to 250	20%	20%	20%	20%	20%
4.1	100	$\beta = 100\%$ for some channel				

Note. True values of the branching fractions for the different universes from which we generate our mock catalogs, along with the number of detections in each mock catalog used in the inference and the sections in which they are discussed. For all universes, we pick true values $\chi_b = 0.0$ and $\alpha_{CE} = 1.0$, except for the equal branching fraction universe (second row), where we use $\chi_b = 0.2$.

hierarchical inference analysis outlined in Section 2.1. For consistency, we use the same O4 sensitivities for the detection weighting during the inference as the above computation of mock signal S/Ns.

The various mock universes that we use in this paper, along with the sections in which they are discussed, are summarized in Table 1. For this analysis, we choose a fiducial unequal mixture of formation channels with underlying branching fractions $\beta_{CE} = 40\%$, $\beta_{CHE} = 5\%$, $\beta_{GC} = 25\%$, $\beta_{NSC} = 5\%$, and $\beta_{SMT} = 25\%$, quasi-isolated natal spins of $\chi_b = 0.0$, and a CE efficiency of $\alpha_{CE} = 1.0$.

To investigate the scaling of the hyperposterior with N_{obs}, for each universe we repeat the inference with $N_{obs} = 50$, $N_{obs} = 150$, and $N_{obs} = 250$. $N_{obs} = 250$ represents an estimate on the total number of BBH detections anticipated by the end of O4 (Weizmann Kiendrebeogo et al. 2023).

3.2. Results

In Figure 3, we plot for our chosen fiducial universe the overall posterior distribution on the underlying branching fractions for different values of N_{obs} (left column), as well as contributions (to
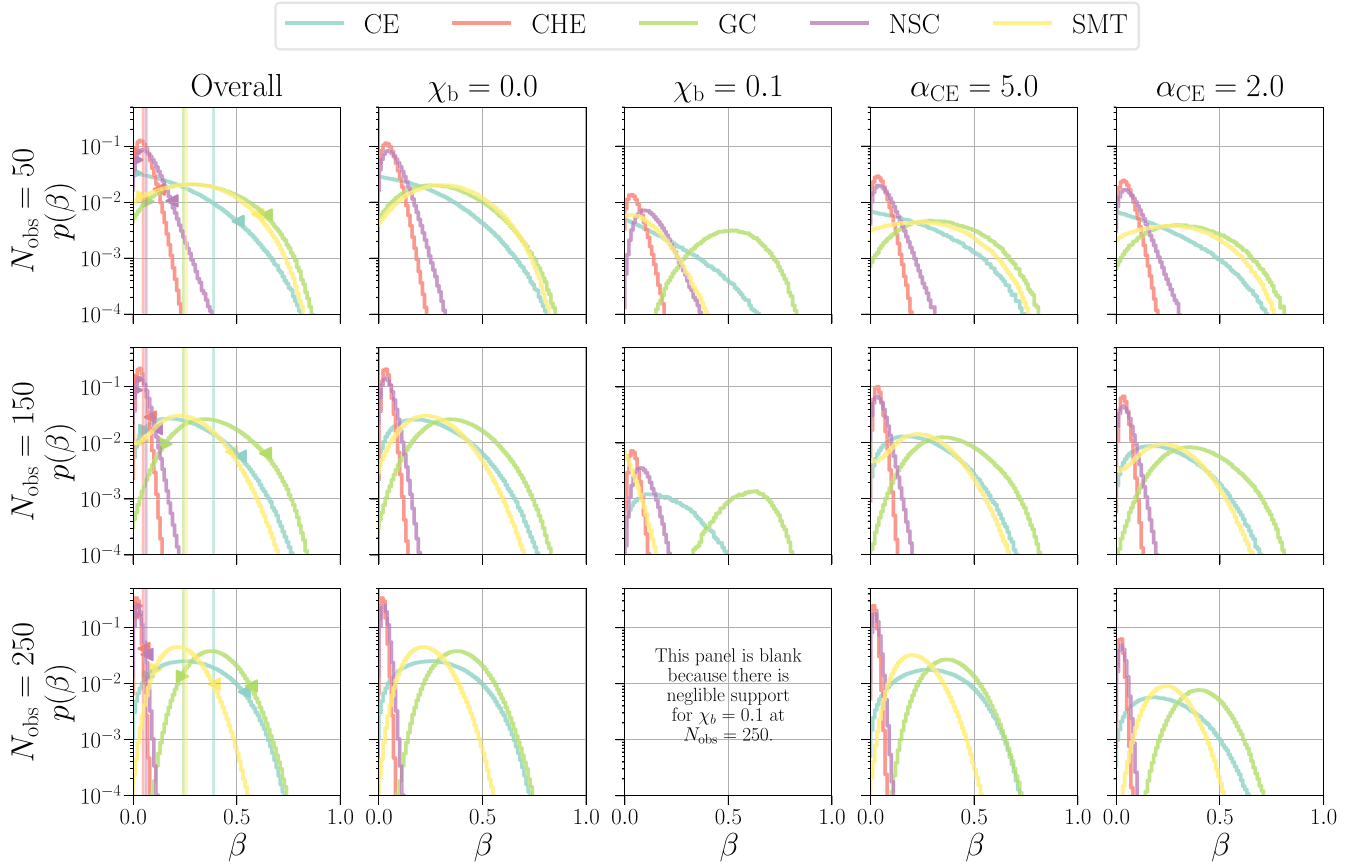
Figure 3. Marginalized branching fraction posteriors for 50 (first row), 150 (second row), and 250 (third row) mock detections. Here, branching fraction posteriors of different channels are plotted together as different colored lines, while contributions from different hypermodels are shown in different columns. The first column (from the left) shows the total marginalized branching fraction; the triangles mark the credible symmetric 90% intervals. The second and third columns show the contributions from the two values of χ_b (marginalized over α_{CE}) with the most support; the fourth and fifth columns are analogous for α_{CE}. It should be noted that the second through fifth columns do *not* represent normalized probability distributions; we plot them to scale with their contribution to the total marginalized posterior in the first column, just as with the colored curves in Figure 2. True values of the underlying branching fractions ($\beta_{CE} = 40\%$, $\beta_{CHE} = 5\%$, $\beta_{GC} = 25\%$, $\beta_{NSC} = 5\%$, and $\beta_{SMT} = 25\%$) are shown by the vertical colored lines in the "Overall" column, where β_{NSC} and β_{SMT} are given small artificial offsets, while $\chi_b = 0.0$ and $\alpha_{CE} = 1.0$.

scale) from the most favored values of χ_b and α_{CE} (columns 2 to 5). First, we do recover the true values of $\boldsymbol{\beta}$ and χ_b. For $N_{obs} = 250$, we find $\beta_{CE} = 28^{+26}_{-21}\%$, $\beta_{CHE} = 3.0^{+2.2}_{-1.5}\%$, $\beta_{GC} = 40^{+18}_{-16}\%$, $\beta_{NSC} = 3.6^{+3.1}_{-2.1}\%$, and $\beta_{SMT} = 24^{+16}_{-13}\%$; the true value of the branching fraction falls within the 90% symmetric credible interval of the posterior for all five channels. Furthermore, $\chi_b = 0$, the chosen true value for χ_b, is favored over the next best model, $\chi_b = 0.1$, by a Bayes factor of 3.5×10^5.

Unlike χ_b, the model with α_{CE} equal to the true value is not favored, with a marginal preference for higher values $\alpha_{CE} = 5.0$ and $\alpha_{CE} = 2.0$ over the true value $\alpha_{CE} = 1.0$ with Bayes factors $BF^{\alpha_{CE}=5.0}_{\alpha_{CE}=1.0} = 8.9$ and $BF^{\alpha_{CE}=2.0}_{\alpha_{CE}=1.0} = 2.5$. Because α_{CE} only affects the CE channel, it is not too surprising that the inference did not decisively favor any one value, in a universe where most detected BBHs do not come from the CE channel. We note two possible contributing factors to favoring higher values of α_{CE} over the true value. If BBHs formed in the CE channel disperse their envelopes more efficiently (i.e., have higher values of α_{CE}), the resulting BBHs are

(a) Less massive, leading to lower detection efficiencies and therefore larger measurement uncertainties for this channel. This is because models with larger α_{CE} have less low-mass binaries merging within the CE itself, leading to more low-mass BBHs being able to form and merge (Bavera et al. 2021).

(b) Lower spinning, due to the post-CE separations being wider for higher α_{CE} and therefore less susceptible to tidal spin-up (Qin et al. 2018; Zaldarriaga et al. 2018; Bavera et al. 2021); this leads to more detections with χ_{eff} closer to $\chi_b = 0.0$, the chosen fiducial value for this universe. In χ_{eff} space, this is a feature degenerate with dynamical channels such as the GC channel, which also produces BBHs with $\chi_{eff} \approx 0$ due to the BBHs produced having isotropic spin orientations.

This result suggests that our result in Section 2.2, that higher common envelope efficiencies are favored, should be viewed with some caution.

More notably, we see convergence of the hyperposteriors toward their true values as we increase the number of detections. First, we highlight the narrowing of the branching fraction posteriors from the first row to the third row of Figure 3. We quantify uncertainties in the branching fraction posteriors by the widths of the 90% symmetric credible intervals, and will hereafter use the two phrases interchangeably. These uncertainties decrease by up to 69% as we go from 50 to 250 mock events (see Table 2). The Bayes factor in favor of the correct χ_b increases by nearly five orders of magnitude, strongly selecting the true value $\chi_b = 0.0$. This is illustrated by the empty plot corresponding to the contribution from $\chi_b = 0.1$ for $N_{obs} = 250$ (third row, middle column), indicating

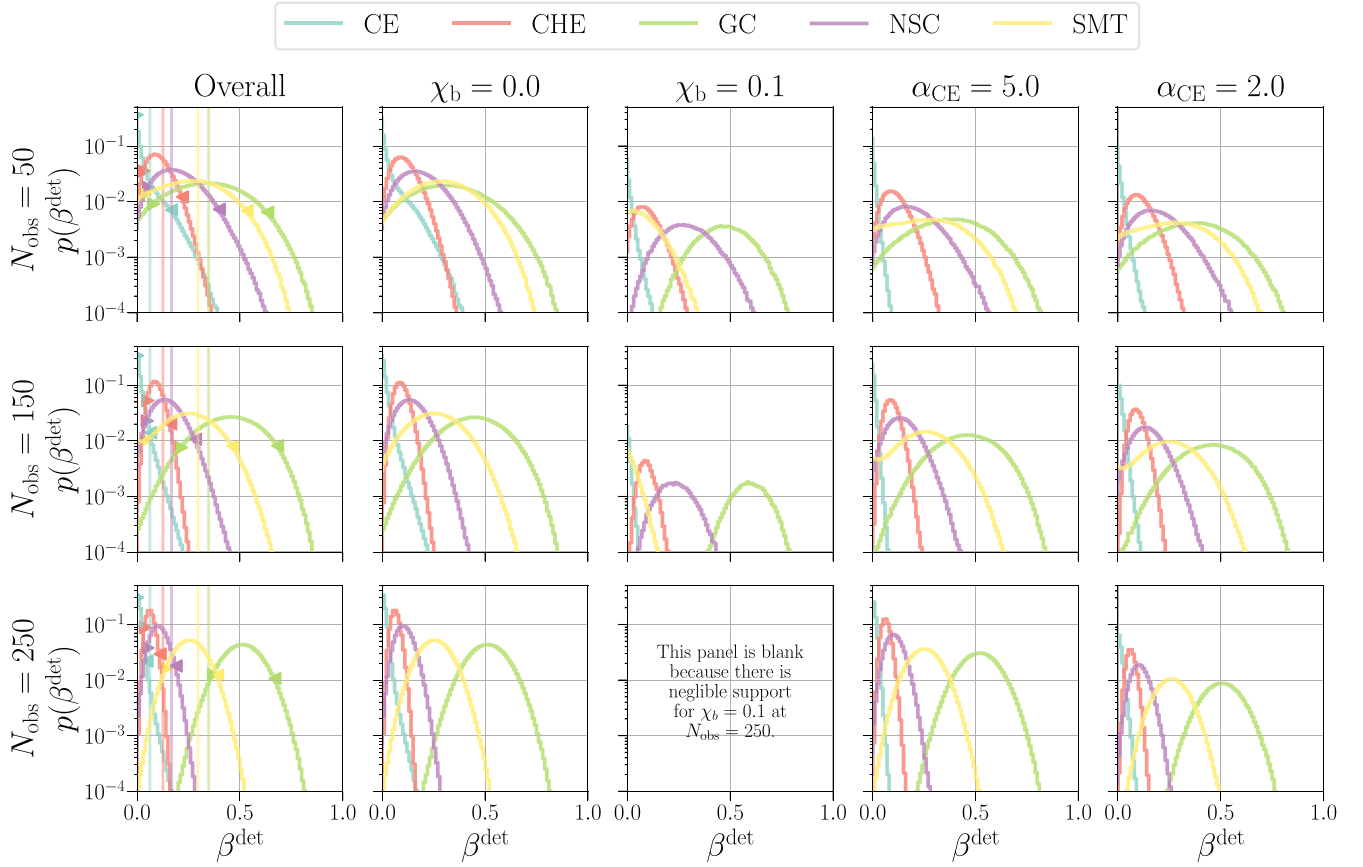**Figure 4.** The same as Figure 3, except with detectable branching fractions rather than underlying.

Table 2
Uncertainties in the Measurement of the Branching Fractions with N_{obs} Shown in Figures 3 and 4

N_{obs}	$\Delta\beta_{CE}$	$\Delta\beta_{CHE}$	$\Delta\beta_{SMT}$	$\Delta\beta_{GC}$	$\Delta\beta_{NSC}$	$\Delta\beta_{CE}^{det}$	$\Delta\beta_{CHE}^{det}$	$\Delta\beta_{SMT}^{det}$	$\Delta\beta_{GC}^{det}$	$\Delta\beta_{NSC}^{det}$	$BF_{\chi_b=0.1}^{\chi_b=0.0}$	$BF_{\alpha_{CE}=5.0}^{\alpha_{CE}=1.0}$
50	49%	11%	52%	17%	56%	17%	19%	55%	34%	49%	8.0	0.81
150	46%	6.2%	48%	9.7%	43%	6.4%	11%	47%	24%	42%	27	0.36
250	46%	3.7%	33%	5.2%	28%	4.6%	7.2%	30%	14%	25%	3.5×10^5	0.11
% change	5.9%	66%	36%	69%	49%	73%	62%	46%	60%	49%		

Notes. The rightmost two columns give the Bayes factors of the true value of χ_b and α_{CE} over the most favored competing value for 50, 150, and 250 mock detections (first, second, and third rows). The rest of the table shows the underlying and detectable branching fraction uncertainties, as well as their percent changes between the first and third rows. We note that these are credible 90% interval widths in percentage points, not percent uncertainties.

negligible support for competing values of χ_b. On the other hand, the inference has increasing support for $\alpha_{CE} = 5.0$, but even at 250 mock detections, we only weakly prefer it over the true model $\alpha_{CE} = 1.0$.

It is worth noting that the scaling of the uncertainty with N_{obs} varies significantly between formation channels. Some channels have more distinctive features in parameter space (see Figure 1) that make them easier to distinguish during the inference. Additionally, differences in the detection efficiencies of different formation channels likely play a role as well. Figure 5 shows the percent decrease in the uncertainty of branching fraction posteriors from $N_{obs} = 50$ to $N_{obs} = 250$ against the detection efficiency of its channel $\xi^{\chi_b,\alpha_{CE}}$ for our fiducial values $\chi_b = 0.0$ and $\alpha_{CE} = 1.0$. Channels with lower detection efficiencies, most notably the CE channel, appear to scale much more poorly with N_{obs} than channels with higher

detection efficiencies. Indeed, the CE channel tends to produce lower-mass black holes at higher redshifts (due to typically shorter delay), leading to fewer of them being detectable (see Figure 1). We emphasize the effect of the low detection efficiency of the CE channel: despite the fact that 40% of the mock underlying population originates from this channel, only 2, 5, and 17 CE BBHs end up in the mock observations, out of the 50, 150, and 250 total detections, respectively. Not only does the β_{CE} posterior have the largest uncertainty of all formation channels, it also barely decreases in uncertainty as we increase N_{obs}. If we instead examine how the *detectable* branching fractions scale with N_{obs} in Figure 4, we can see that all detectable branching fractions narrow at similar rates as we increase N_{obs} from 50 (first row) to 250 (third row). Indeed, we can see in Figure 5 that the detectable branching fractions (dotted line) do not exhibit the dependence of the convergence
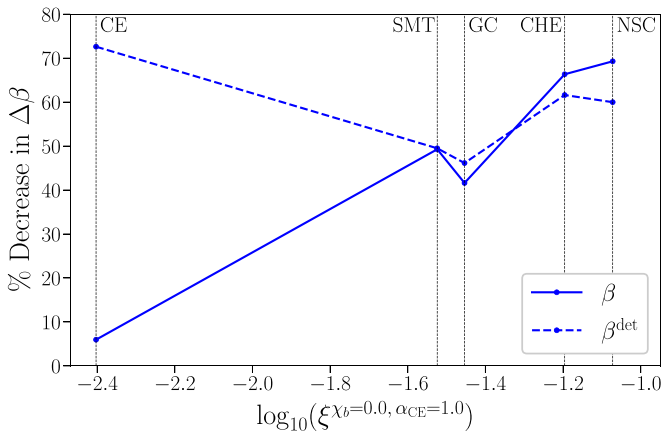
Figure 5. The percent decrease in the branching fraction posterior 90% symmetric credible interval width from 50 to 250 mock detections plotted against the channel's detection efficiency. The black dotted vertical lines mark which points correspond to which formation channel. While the underlying branching fractions (solid line) converge faster with N_{obs} in channels with higher detection efficiencies, no such correlation is seen in the detectable branching fractions (dashed line).

rate on detection efficiency that the underlying branching fractions (solid line) have.

Already, we can see some of the difficulties that arise from hierarchical Bayesian inference in the face of large measurement uncertainties, selection effects, and degenerate features in parameter space. This is a common theme, and we will explore these problems in further detail in Section 4.

Finally, we repeat this analysis for a different set of hyperparameters consisting of equal branching fractions between all formation channels ($\beta_j = 0.2$ for all j), $\chi_b = 0.2$, and $\alpha_{CE} = 1.0$ (see the second row of Table 1). We find similar results in the convergence with N_{obs}. Uncertainties in the underlying branching fraction decrease by up to 47% from $N_{obs} = 50$ to $N_{obs} = 250$, and consistently with our previous results, the uncertainty in β_{CE} does not decrease. We find also that the shrinking of the detectable branching fraction (β^{det}) posterior uncertainty is more consistent across different channels than the underlying branching fraction. Increasing N_{obs} causes a strong preference for the true value of $\chi_b = 0.2$; while we have $BF^{\chi_b=0.2}_{\chi_b=0.1} = 1.0$ for $N_{obs} = 50$, there is no posterior support for $\chi_b \neq 0.2$ at $N_{obs} = 150$ and $N_{obs} = 250$. Similarly to the previous example, there is no strong preference for the true value of α_{CE}, although it is slightly favored with $BF^{\alpha_{CE}=1.0}_{\alpha_{CE}=0.5} = 1.7$. Figures showing the marginalized posteriors on β and β^{det} for this set of hyperparameters can be found in Appendix B.

4. Biases of Population Inference

Finally, we investigate the biases that may arise when performing hierarchical inference with the methods described above in Section 2.1. We again refer to Table 1 for the chosen true values of the hyperparameters that we present in subsequent sections; we use the same method for hierarchical inference and simulated BBH detections as outlined in Sections 2.1 and 3.1, respectively. In Section 4.1, we perform hierarchical inference on sources originating exclusively from each individual formation channel and examine differences in the recovered posteriors. In Section 4.2, we isolate the effect of the natal black hole spin hyperparameter by comparing the posteriors of universes with different choices of χ_b. In Section 4.3, we

explore the consequences of doing hierarchical inference with incomplete information (i.e., excluding a channel from the inference) with detections from a mixture of formation channels. Finally, we summarize and discuss our results in Section 4.4.

4.1. Inference in Single-channel Dominated Universes

For each channel j, we perform hierarchical inference on $N_{obs} = 100$ mock detections in a universe where the entire underlying BBH population originates from channel j (i.e., $\beta_j = 1$). We choose for our true values of the physical prescription $\chi_b = 0.0$ and $\alpha_{CE} = 1.0$, although the latter choice only affects the CE-dominated universe. Figure 6 shows the branching fraction posteriors and support for different values of χ_b for each single-channel dominated universe.

In general, none of the branching fraction posteriors have significant support for $\beta = 1$, even for the channel that is actually producing the entirety of the BBHs. This happens because we use a flat, symmetric Dirichlet distribution for our prior, which results in a prior preference for a mixture of channels rather than a single dominating channel; this is illustrated by the gray curves in Figure 6.

The fifth percentiles of the dominating-channel branching fraction posteriors for the CE, CHE, GC, NSC, and SMT-dominated universes are $\beta_{CE} = 94\%$, $\beta_{CHE} = 62\%$, $\beta_{GC} = 41\%$, $\beta_{NSC} = 31\%$, and $\beta_{SMT} = 51\%$, respectively. The degree to which we underestimate the contribution from the dominating channel varies significantly depending on the channel. We again highlight the effect of detection efficiency. As we saw in Figure 5, the CE, GC, and SMT channels have the lowest detection efficiencies, especially the CE channel. Across the different universes under consideration, β_{CE}, β_{GC}, and β_{SMT} (first, middle, and last columns, respectively) have larger uncertainties that compete with and take away from the branching fraction posterior of the dominating channel; because of the lower detection efficiencies of these channels, it is difficult to discern whether this channel is nonexistent or if its contribution to the full set of *detected* observations is minor. The case of β_{CE} is particularly severe: as it does not strongly deviate from the prior, the 95th percentile for β_{CE} is greater than 32% in all universes for which there is no CE contribution to the underlying population. The CE-dominated universe (first row) does not suffer from this effect; as a result, only in that universe do we recover with narrow precision that the dominating channel is indeed dominating. In the opposite case, the NSC channel has the highest detection efficiency. In the NSC-dominated universe (fourth row), the median value of the β_{NSC} posterior is less than 0.5, with large contributions to the BBH population from the channels with lower detection efficiencies (β_{CE}, β_{GC}, and β_{SMT}).

There are also several interesting features of the spin model selection. First, in the CHE-dominated universe (second row), the inference is not able to select the true natal black hole spin of $\chi_b = 0.0$, and instead gives approximately equal support to $\chi_b = 0.0$, 0.1 and 0.2, as illustrated by the similar heights of the different colored curves. Let us recall that, while the aligned-spin CE and SMT field channels mostly produce BBHs with $\chi_{eff} \approx \chi_b$ (although CE has a tail for higher spins from tidal spin-up) and the isotropic-spin GC and NSC dynamical channels produce BBHs scattered around $\chi_{eff} \approx 0$, the CHE channel uniquely produces BBHs with $\chi_{eff} \gtrsim 0.2$ *irrespective of χ_b*, due to strong tidal spin-up effects (see Figure 1). As a result, the inference is not able to discern between values of χ_b
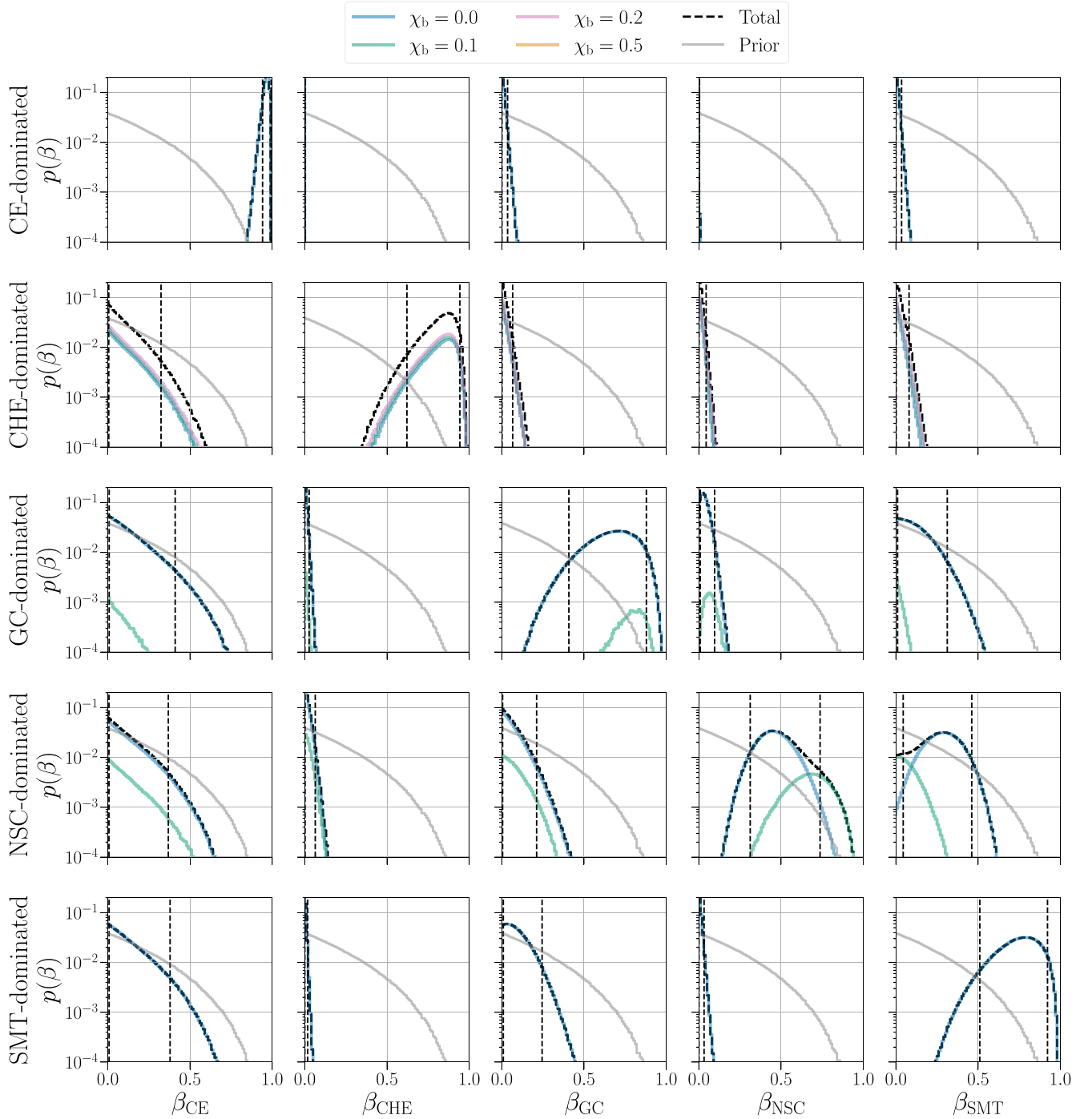
Figure 6. Marginalized branching fraction posteriors inferred from a catalog of 100 mock detections for a universe where the only BBH formation channel is (top to bottom) CE, CHE, GC, NSC, or SMT. As in the top row of Figure 2, different colored curves represent support for different values of χ_b marginalized over α_{CE}, while the dotted black curve gives the total marginalized posterior. We have also included the prior in gray. The vertical lines mark the 90% symmetric credible interval.

between 0 and the true value 0.2. In the universes dominated by dynamical formation channels (GC, third row, and NSC, fourth row), the selection of the true value of $\chi_b = 0.0$ is less strong, with non-negligible support for $\chi_b = 0.1$ as shown by the green curves. In these universes, most detections have $\chi_{eff} \approx 0$, and χ_b only affects the width of this distribution, which is a weaker feature to detect. Additionally, the subpopulation of hierarchical mergers in these channels help to drive the mild support for $\chi_b > 0$; at lower values of χ_b, hierarchical mergers occur more

readily, due to weaker gravitational recoil kicks. Hierarchical mergers can have $|\chi_{eff}|$ significantly greater than zero (Pretorius 2005; González et al. 2007; Buonanno et al. 2008; Kimball et al. 2020), as illustrated by the wings of the marginalized χ_{eff} distribution for the GC and NSC channels in Figure 1. Because hierarchical mergers form a larger subpopulation in NSC, due to their deeper potential wells and ability to retain post-merger black holes (Miller & Lauburg 2009; O'Leary et al. 2009; Hong & Lee 2015;
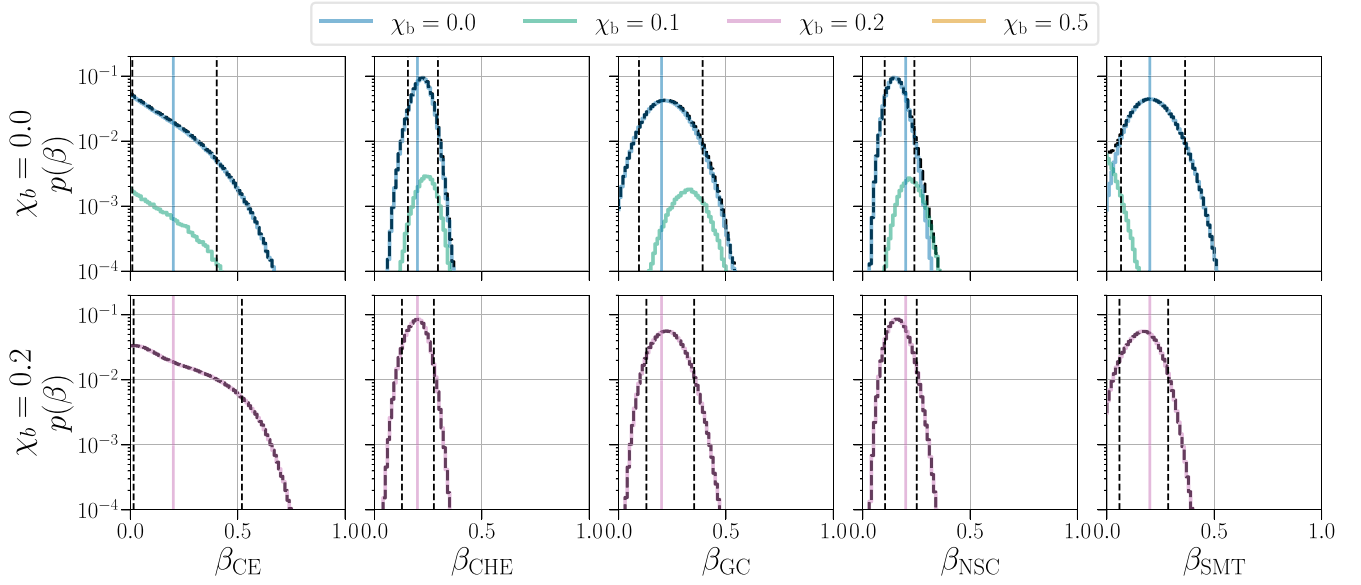
Figure 7. Marginalized branching fraction posteriors in the same format as Figure 6, except here the rows show two universes (each with 250 mock observations) with different true values of χ_b (0.0, top; and 0.2, bottom). The true values of the branching fractions ($\beta_j = 20\%$ for all channels j) are given by the solid vertical lines, color-coded by the true value of χ_b.

Antonini & Rasio 2016), this effect is greater for the NSC-dominated universe. Additionally, in the NSC-dominated universe, the $\beta_{\rm NSC}$ posterior for $\chi_b = 0.1$ (green curve) peaks at a higher value than $\chi_b = 0.0$ (blue curve). This is because, when $\chi_b = 0.1$, field channels produce fewer of the $\chi_{\rm eff} \approx 0$ BBHs that make up the bulk of the population.

Finally, although it is not shown in Figure 6, we comment on the inference on $\alpha_{\rm CE}$. First, we do favor the true value of $\alpha_{\rm CE} = 1.0$ for the CE-dominated channel, preferring it over the next most favored model with $\mathrm{BF}^{\alpha_{\rm CE}=1.0}_{\alpha_{\rm CE}=2.0} = 72$, as expected from a universe where all detections are from the CE channel. We also note consistency with the results of Section 3.2 in the recovery of $\alpha_{\rm CE}$ for universes where there is no contribution from the CE channel; there is a small bias toward higher values of $\alpha_{\rm CE}$. We find $\mathrm{BF}^{\alpha_{\rm CE}>1.0}_{\alpha_{\rm CE}<1.0} = 1.52$, 1.54, 1.68, and 1.58 for the CHE, GC, NSC, and SMT-dominated universes, respectively.[9]

4.2. Biases in Spin Inference

To examine the effect of χ_b on the hyperposterior, we perform hierarchical inference on different universes with the same underlying branching fractions but different true values of χ_b, using 250 mock detections. To isolate the effect of χ_b, we choose an equal mixture of formation channels in the underlying population ($\beta_j = 0.2$ for all channels j). We choose $\alpha_{\rm CE} = 1.0$.

Figure 7 shows the marginalized branching fraction posteriors for universes with $\chi_b = 0.0$ and $\chi_b = 0.2$. Consistently with our results in Section 3.2, we recover the true values of the branching fractions as well as the true value of χ_b for both universes. Here, we can see that the selection of χ_b is nonlinear: it is harder to distinguish between lower black hole spins (i.e., $\chi_b = 0.0$ versus $\chi_b = 0.1$) and higher ones. While there is no support for other values of χ_b in the posterior of the $\chi_b = 0.2$ universe, there is still non-negligible support for

---

[9] To be precise, the quantity we calculate is

$$(\mathrm{BF}^{\alpha_{\rm CE}=2.0}_{\alpha_{\rm CE}=1.0} + \mathrm{BF}^{\alpha_{\rm CE}=5.0}_{\alpha_{\rm CE}=1.0})/(\mathrm{BF}^{\alpha_{\rm CE}=0.2}_{\alpha_{\rm CE}=1.0} + \mathrm{BF}^{\alpha_{\rm CE}=0.5}_{\alpha_{\rm CE}=1.0}).$$
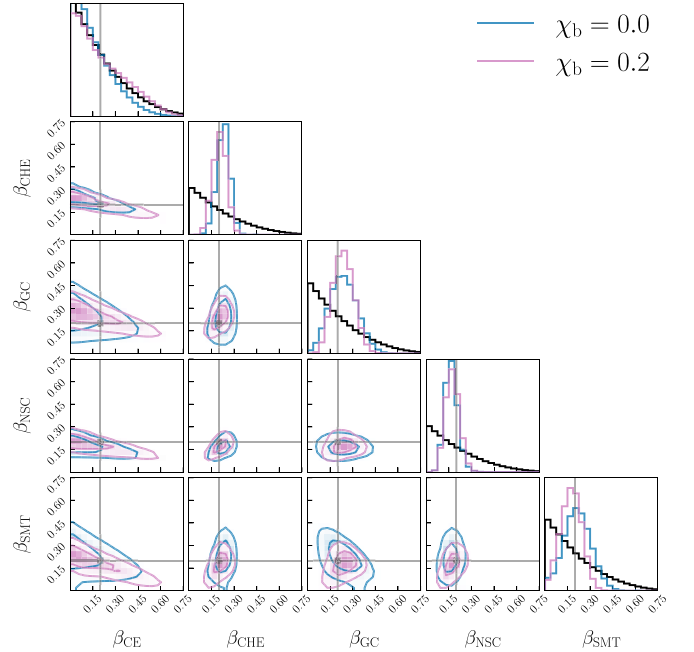

Figure 8. Corner plot of the branching fraction posteriors for an equal-mixture universe, with $\alpha_{\rm CE} = 1.0$ and $\chi_b = 0.0$ (blue) and 0.2 (pink). Contours correspond to 50% and 90% credible intervals. The true values of the branching fractions are marked by the gray lines, while the prior is plotted in black only in the 1D histograms, for clarity.

$\chi_b = 0.1$ in the $\chi_b = 0.0$ universe, with $\mathrm{BF}^{\chi_b=0.0}_{\chi_b=0.1} = 30$. This is as expected; it is difficult to discern between slowly spinning and nonspinning populations, due to the inherent measurement uncertainty of GW observations (Baird et al. 2013; Vitale et al. 2014; Pürrer et al. 2016; Vitale et al. 2017a).

Figure 8 shows a corner plot of the branching fraction posteriors for both universes. Here, we can see the same effect: despite having the same number of mock detections, the data are more informative (yielding a posterior more different from the prior) in the universe with nonzero χ_b.

4.3. Inference with an Incomplete Set of Populations

Finally, we perform hierarchical inference with one formation channel excluded, such that, while all five channels are contributing sources, the inference is only performed with four channels. This analysis is motivated by the fact that any population analysis done on real BBH data likely does its inference with an incomplete set of formation channels; we most likely do not know the totality of all possible BBH formation channels, nor can we model them all accurately and self-consistently. We again consider an equal-mixture branching fraction universe ($\beta = 0.2$ between all formation channels) and true values for our physical prescription of $\alpha_{CE} = 1.0$ and $\chi_b = 0.2$. Figure 9 shows the marginalized branching fraction posteriors and support for different values of χ_b for the full inference as well as for inferences with one channel excluded. We defer discussion of biases in α_{CE} selection to Appendix B and instead focus on χ_b and the formation channel branching fractions in this section.

By examining which channels receive more or less posterior support as a result of the inference's incomplete knowledge of formation channels, we can infer the correlations between the branching fractions of different formation channels. For example, when the CE channel is excluded (second row), the increase in branching fraction is spread approximately equally over the other channels, suggesting that β_{CE} is negatively correlated with the other branching fractions, as is the prior. On the other hand, when the SMT channel is excluded from the inference (last row), the β_{CE} posterior shifts toward higher values, while support for the other channels *decreases*.

Such correlations are consistent with the branching fraction corner plot from the full inference (Figure 8); the pink contours are relevant to the set of mock detections discussed in this section. The uncertainties in β_{CE} are the greatest of all channels, and the posterior the least constrained; as such, the marginalized β_{CE} posterior closely follows the prior. As a result, β_{CE} is negatively correlated with all channels, which leads to systematic overestimation when doing inference with a channel excluded. In general, we can see that some branching fraction posteriors (i.e., β_{NSC} and β_{CHE}) are constrained much better than others and are affected the least by the choice of prior. As seen in Section 4.1, the prior can cause a bias toward higher values of β for channels with lower detection efficiencies and greater uncertainties.

Next, we highlight the effect of incomplete formation channel knowledge on the selection of the natal black hole spin χ_b. We see two cases in which the wrong value of χ_b is selected. When the CHE channel is excluded (third row), we infer a high natal black hole spin of $\chi_b = 0.5$, as indicated by the yellow curve, with a Bayes factor over the true value $\chi_b = 0.2$ of $BF^{\chi_b=0.5}_{\chi_b=0.2} = 373$. This is due to the high-χ_{eff} BBHs that the CHE channel produces. When the inference does not account for the CHE channel, it tries to explain these highly spinning CHE black holes with other field-channel BBHs spinning at a higher χ_b. Then, in order to still account for the lower χ_{eff} non-CHE BBHs, the branching fractions for the GC and SMT channels are increased and decreased, respectively. We remark that no such adjustment is seen for the NSC and CE channels, despite them having features in χ_{eff} space similar to those of the GC and SMT channels, respectively.

An opposite effect is seen when excluding the NSC channel (fifth row), which, as noted above, produces BBHs scattered around $\chi_{eff} = 0$ with tails that extend to more positive and negative values, due to the presence of hierarchical mergers. The inference has constructed two competing explanations in order to explain these lower-spinning BBHs: low-χ_b CE BBHs (represented by the green and blue curves with posterior support for high β_{CE} and low β_{GC}) and higher-χ_b GC BBHs (represented by the pink curve with support for low β_{CE} and high β_{GC}). This highlights the degeneracy between low-spinning field channels and high-spinning dynamical channels in producing similar features in the population χ_{eff} distribution. This case is especially remarkable because the green and blue curves ($\chi_b = 0.0$ and $\chi_b = 0.1$) for the β_{CE} posterior bear a striking resemblance to the β_{CE} posterior inferred from GWTC-3 data (Figure 2), even though these features are purely an artifact of the inference neglecting a single BBH formation channel. We again remark on the difference between the β_{CE} and β_{SMT} posteriors: despite the fact that both CE and SMT are field channels with the similar features in χ_{eff} space, support for β_{SMT} does not increase (and rather decreases) in the low-χ_b model as a result of the exclusion of the NSC channel. One reason why this may occur is that SMT BBHs cannot go through tidal spin-up, unlike CE BBHs, and hence they account for a narrower range of χ_{eff} concentrated around χ_b. Therefore, β_{SMT} can be strongly affected by the exclusion of a channel with strong features in χ_{eff} space, especially when the wrong model of χ_b is inferred.

As mentioned in the previous paragraph, we also find it noteworthy that the inference favors two separate competing explanations when NSC is excluded. The GC and NSC marginalized χ_{eff} KDEs are similar due to the isotropy of spin orientations in these two channels; upon examination of the marginalized KDEs of the other three BBH parameters ($\mathcal{M}_c$, q, z), the GC channel appears still to have the closest resemblance to the NSC channel. Despite this, the inference does not simply overcompensate for the exclusion of the NSC channel by correspondingly increasing β_{GC}, suggesting the influence of higher-dimensional features in parameter space. On the other hand, no such effect is seen when excluding the GC channel, whose posterior has a simple overcompensation in β_{NSC} and β_{CE}.

Indeed, although we have been able to broadly interpret these posteriors by focusing on different channels' features in χ_{eff} space, there must be other subtle effects at play. We have shown in multiple ways the differences between the β_{CE} and β_{SMT} posteriors and the β_{GC} and β_{NSC} posteriors, despite having them similar features in χ_{eff} space. Although the exclusion of each channel has its own unique and interesting consequences, there appears to be a bias for the CE and GC channels, systematically underestimating the CHE and NSC channels. We point again to detection efficiency: of the field and dynamical channels, respectively, the CE and GC channels have by far the lowest detection efficiencies. With the uniform branching fraction spread in our current set of hyperparameters, only 2% of detections are expected to be from the CE channel (versus 30% and 16% from the CHE and SMT channels) and 16% from the GC channel (versus 36% from the NSC channel).

Finally, we have repeated this analysis for the set of hyperparameters used in the convergence analysis in Section 3, with an unequal mixture of channels and $\chi_b = 0.0$. In this universe, the bias toward overestimating β_{GC} is more severe due to the lack of spin information from our choice of χ_b. Consistently with our previous discussion, there again seems to
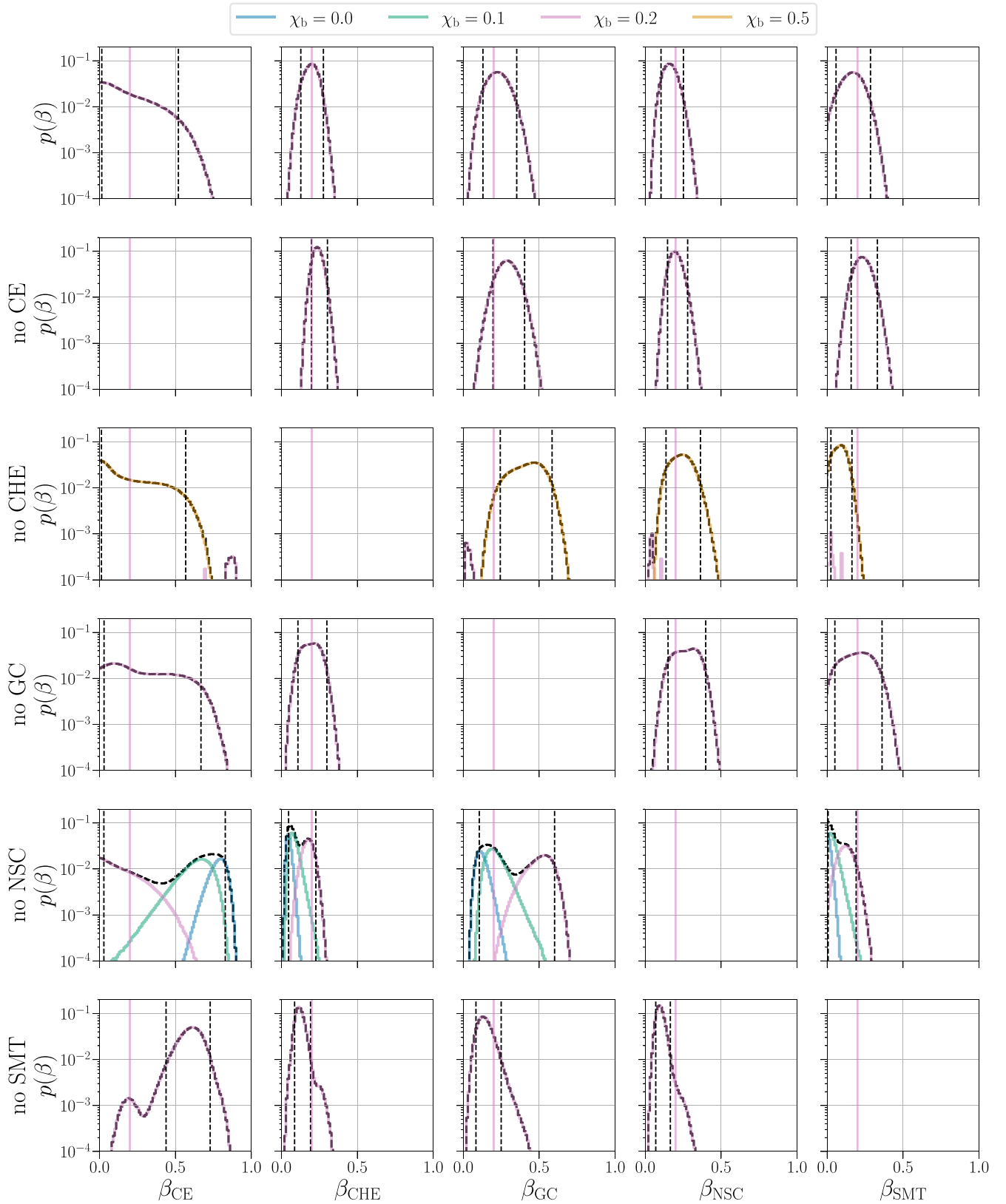
Figure 9. Marginalized branching fraction posteriors. The first row shows the results from the full inference in our $\chi_b = 0.2$, $\alpha_{CE} = 1.0$ equal branching fraction universe with $N_{obs} = 250$, identical to the second row of Figure 7, while each of the remaining rows shows the result of inference performed without knowledge of one of the formation channels. The format of this plot is the same as Figures 6 and 7.

be a bias toward the CE and GC channels when one channel is excluded. Due to the CHE and NSC channels' high detection efficiencies, the β_{CHE} and β_{NSC} posteriors support their low

true values (5% for both channels) with small uncertainties. The corresponding plot of the branching fraction posteriors for this universe can be found in Appendix B.

4.4. Discussion

In this section, we conducted several investigations of different biases that arise in hierarchical Bayesian inference based on astrophysical formation models of BBHs. We summarize the key takeaways as follows:

1. Single-channel dominated universes (Section 4.1)
   (a) Even when our models of the underlying formation channels are perfectly accurate, at $N_{\rm obs} = 100$ the data are still relatively uninformative due to information loss from parameter estimation and low detection rates. Thus, results of inference are still influenced by the choice of a flat prior and exhibit bias toward a mixture of formation channels. This puts a caveat on our result from Section 2.2 that no single channel dominates the underlying BBH population, from the inference on GWTC-3 data.
   (b) It is difficult to precisely infer the true value of $\alpha_{\rm CE}$, as it only affects the CE channel, which produces very few detectable BBHs, due to its low detection efficiency relative to the other channels considered. Only with mock catalogs where the CE channel dominates the detections can we recover the true value of $\alpha_{\rm CE}$, as expected.
2. Biases in spin inference (Section 4.2)
   (a) It is easier to infer higher values of $\chi_{\rm b}$ than lower values. When $\chi_{\rm b}$ is low (0.0 or 0.1), one has less spin information, and uncertainties in both the branching fraction posteriors and the selection of the true value of $\chi_{\rm b}$ are greater.
3. Inference with incomplete populations (Section 4.3)
   (a) Both the branching fraction posteriors and the inferred value of $\chi_{\rm b}$ can be heavily affected if the inference is performed without knowledge of all contributing formation channels. Some branching fractions are overestimated or underestimated by a factor of ~ 3 or more from the exclusion of a formation channel, and ignoring channels that produce particularly high or low $\chi_{\rm eff}$ BBHs can cause the inference to strongly select an incorrect value of $\chi_{\rm b}$.
   (b) Degeneracies exist between different sets of hyper-parameters that can make it difficult for the inference to discriminate between them. For example, it can be difficult to distinguish between field BBHs with low $\chi_{\rm b}$ and dynamical BBHs with higher $\chi_{\rm b}$, due to $\chi_{\rm eff}$ being the only spin information used in the inference (which in turn is due to the fact that $\chi_{\rm eff}$ is arguably the only spin parameter that can be measured for all BBHs with advanced detectors). There are likely more subtle degeneracies and correlations in higher-dimensional parameter space that are difficult to explain from the marginalized distributions but nonetheless play a role in the inference.
   (c) Inference on the underlying branching fractions can be biased due to the varying detection efficiencies of different channels. In particular, the CE channel has a detection efficiency nearly an order of magnitude below the other channels, causing large measurement uncertainties in $\beta_{\rm CE}$ and a relatively uninformed (i.e., close to the prior) posterior for $\beta_{\rm CE}$. As a result, the exclusion of a channel from the inference usually results in the $\beta_{\rm CE}$ posterior support extending to

higher values; similar effects can be seen in other low detection efficiency channels, such as the GC and SMT channels.

5. Conclusions

Understanding the physical processes and formation environments of compact binary mergers is one of the most pressing questions in GW astrophysics. In this paper, we pair the most recent catalog of BBH mergers provided by the LVK with an expansive, self-consistent suite of astrophysical models to investigate the origins of BBH mergers. Consistent with Zevin et al. (2021a), we find that, given our set of astrophysical models, multiple formation channels are likely contributing to the observed population (though see Section 4.1 for a caveat). We demonstrate both the predictive power of our inference methodology and its scaling with future detections by generated mock observations with realistic measurement uncertainties from synthetic universes with known branching fractions and physical prescriptions. Perhaps most importantly, we also demonstrate the pitfalls of this type of inference, in particular how an incomplete census of formation models or incorrect physical assumptions can lead to significant biases in inference. This work should be treated as a cautionary tale for those attempting to understand relevant physical processes leading to compact binary mergers and formation environments of compact binary progenitors, as inference can be severely compromised if models suffer from inaccuracies of incompleteness.

Though the suite of BBH formation channel models used in this work are state-of-the-art and apply self-consistent physical treatments where possible, they in many ways can be treated as exemplary. Given the numerous uncertainties in massive-star evolution, binary physics, compact object formation, and environmental effects, it is currently impossible to construct models with complete physical accuracy or to fully explore all the uncertainties that impact the source property predictions of population synthesis. Regardless, the biases demonstrated in this analysis are a generic concern when performing inference based on an incomplete or inaccurate set of astrophysical models. We do not suggest that such studies have no utility; compared to population inference that rely on heuristic or flexible models, studies such as these have the benefit of translating directly to physical constraints, albeit requiring proper caveats. Despite the potential issues with such analyses, we anticipate that, given the diversity of BBH properties observed to date, the key result of multiple formation channels contributing to the detected population of BBHs remains robust.

A potential concern one might have when considering multiple formation channels for the production of BBH mergers is how the Universe could conspire to have multiple distinct formation pathways, governed by unique physics, to contribute to the population of merging BBHs at a similar rate. Occam's razor would suggest that this is an unlikely scenario. However, astrophysical transients have been shown in many instances not to obey this principle (Maccarone 2021). Many channels of BBH formation have predicted rates within the same order of magnitude (Rodriguez et al. 2021; Broekgaarden et al. 2022; see Mandel & Broekgaarden 2022 for a review), and the selection effects inherent to GW detection are certainly capable of causing sources from intrinsically rare channels to

be heavily represented in the detected population. Future observations and improved population synthesis routines may help to more robustly disentangle the relative rates of various compact binary formation channels, and thereby have the capability of placing constraints on underlying physical processes. Nonetheless, for the time being, we show that it is important to consider the potential biases that can accumulate when accounting for an incomplete picture of compact binary mergers in the universe.

Multiple avenues can be used in tandem with the analyses presented in this work to help expedite the ability of placing robust constraints on compact binary formation channels. In addition to analyses of the full population of BBH mergers, observational signatures from single events that are unique to one or a subset of formation channels will help to place constraints on the relative contribution of various formation channels (e.g., Zevin et al. 2021b). Observational constraints from other probes of compact binary formation outside of GW astronomy, such as electromagnetic surveys of BBH stellar progenitors, astrometric observations of compact object binaries, identification and host association of gamma-ray bursts and kilonovae, and characterization of pulsar binaries in the Milky Way can all help complement and improve constraints that rely solely on GW observations. Incorporating such information into astrophysical inference will help population analyses using astrophysical simulations remain pertinent and scale with the rapidly growing catalog of compact binary merger observations.

The posterior samples in the analyses presented in this work, the code for calculating the prior at the posterior points (see Appendix A), and all figures, along with additional figures and the accompanying Jupyter notebook, are available on Zenodo (Cheng 2023).

Appendix A
Calculating the Prior at the Posterior Points

During hierarchical inference, the goal is to calculate the hyperposterior for our hyperparameters $\mathbf{\Lambda} = [\beta, \chi_b, \alpha_{CE}]$

$$p(\mathbf{\Lambda}|\mathbf{x}) = \frac{\pi(\mathbf{\Lambda})p(\mathbf{x}|\mathbf{\Lambda})}{p(\mathbf{x})}, \qquad (A1)$$

as given by Bayes' theorem, where $\mathbf{x} = \{\mathbf{x}_i\}_i^{N_{obs}}$ is the set of BBH detections and

$$p(\mathbf{x}|\mathbf{\Lambda}) = \prod_{i=1}^{N_{obs}} \frac{p(\mathbf{x}_i)}{\int p(\boldsymbol{\theta}|\mathbf{\Lambda})P_{det}(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$
$$\times \int \frac{p(\boldsymbol{\theta}_i|\mathbf{x}_i)p(\boldsymbol{\theta}_i|\mathbf{\Lambda})}{\pi(\boldsymbol{\theta}_i)}\, d\boldsymbol{\theta} \qquad (A2)$$

is the hyperlikelihood (see Appendix D of Zevin et al. 2021a, as well as Farr et al. 2015, Mandel et al. 2019, and Vitale et al. 2022b for reviews). Here, we divide out the parameter estimation prior $\pi(\theta)$ evaluated at each point $\boldsymbol{\theta}_i$ as we integrate over the space of BBH parameters $\boldsymbol{\theta} = [\mathcal{M}_c, q, \chi_{eff}, z]$. We approximate this integral via a Monte Carlo discrete sum over the posterior samples. Therefore, it is necessary to calculate the prior at each posterior sample point.

Zevin et al. (2021a) did this by constructing a four-dimensional Gaussian-kernel KDE with the prior samples in the GWTC-1 and GWTC-2 data releases. Due to the potentially prohibitive behavior of high-dimensional KDEs with insufficient training samples, we choose instead to evaluate the prior at each posterior sample analytically by using the analytical priors from the GWTC-2.1 and GWTC-3 data releases and applying the appropriate Jacobians (see Callister 2021, for a review).

Figure 10 shows the marginalized branching fraction posteriors inferred from GWTC-2.1 and GWTC-3 data, but with the prior $\pi(\theta_i^k)$ for each event i evaluated at each posterior sample k calculated via a four-dimensional Gaussian KDE constructed from the prior samples provided from LVK data releases, as in Zevin et al. (2021a). Comparing with Figure 2, determining $\pi(\theta_i^k)$ in this way gives rise to some noisy features in the posterior, such as nontrivial support for $\chi_b = 0.2$ and $\chi_b = 0.5$ (pink and yellow curves, respectively). While there is no support for $\chi_b > 0.1$ with analytical evaluation of the prior, with the KDE method we have $BF_{\chi_b>0.1}^{\chi_b\leqslant 0.1} = 4.00$. Similarly, preference between different values of α_{CE} is also weaker, with $BF_{\alpha_{CE}=1.0}^{\alpha_{CE}=5.0} = 4.24$, as opposed to 249. The primary notable features in the posterior as discussed in Section 2.2, however, still remain robust.
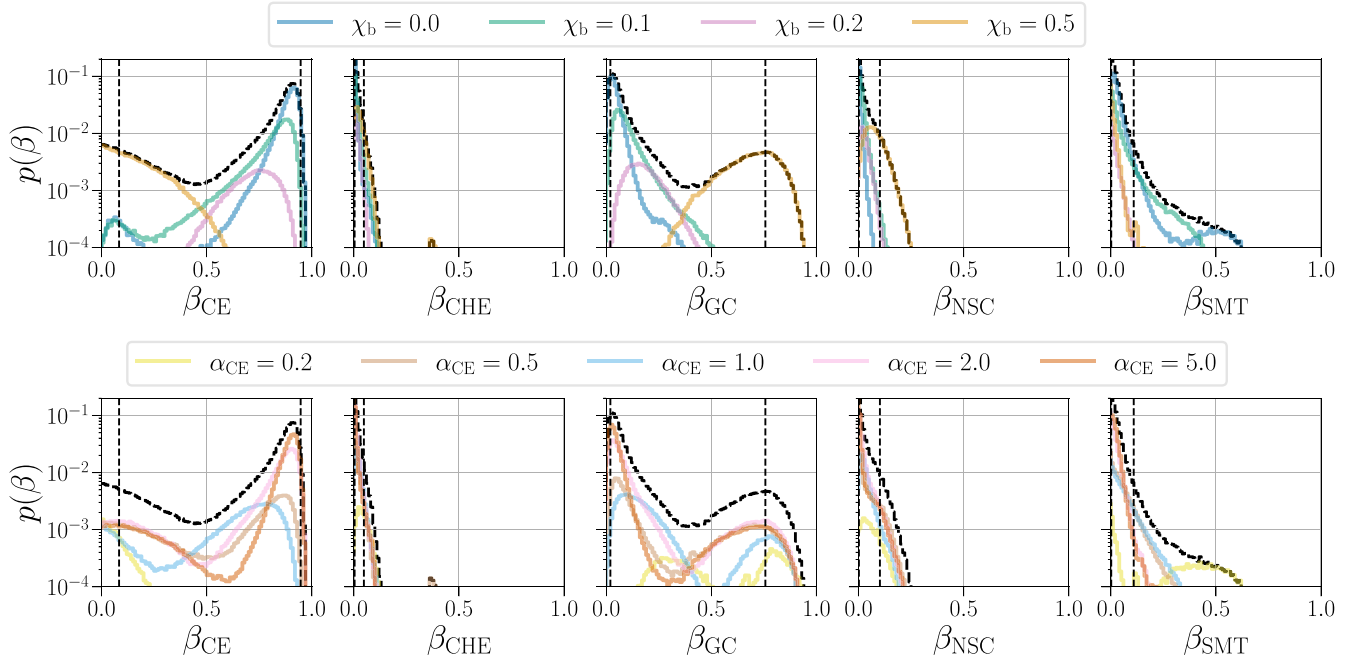
Figure 10. The same as Figure 2, showing the marginalized branching fractions inferred from 68 events in the cumulative GWTC-3 catalog, but with the prior at the posterior points evaluated with a KDE, rather than analytically.

Appendix B
Additional Figures

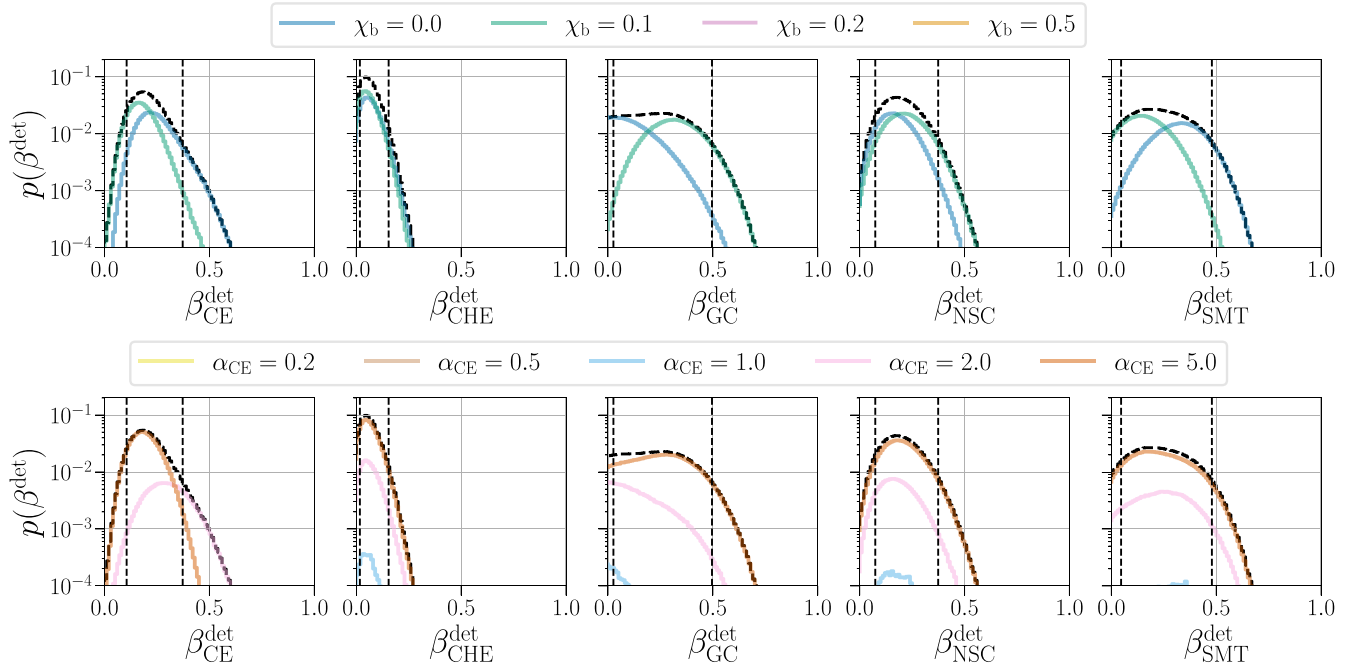In this appendix, we show additional plots (Figures 11–14). Details are given in each figure's caption.

Figure 11. The same as Figure 2, except here we show detectable branching fractions inferred from GWTC-2.1 and GWTC-3 data rather than underlying branching fractions.
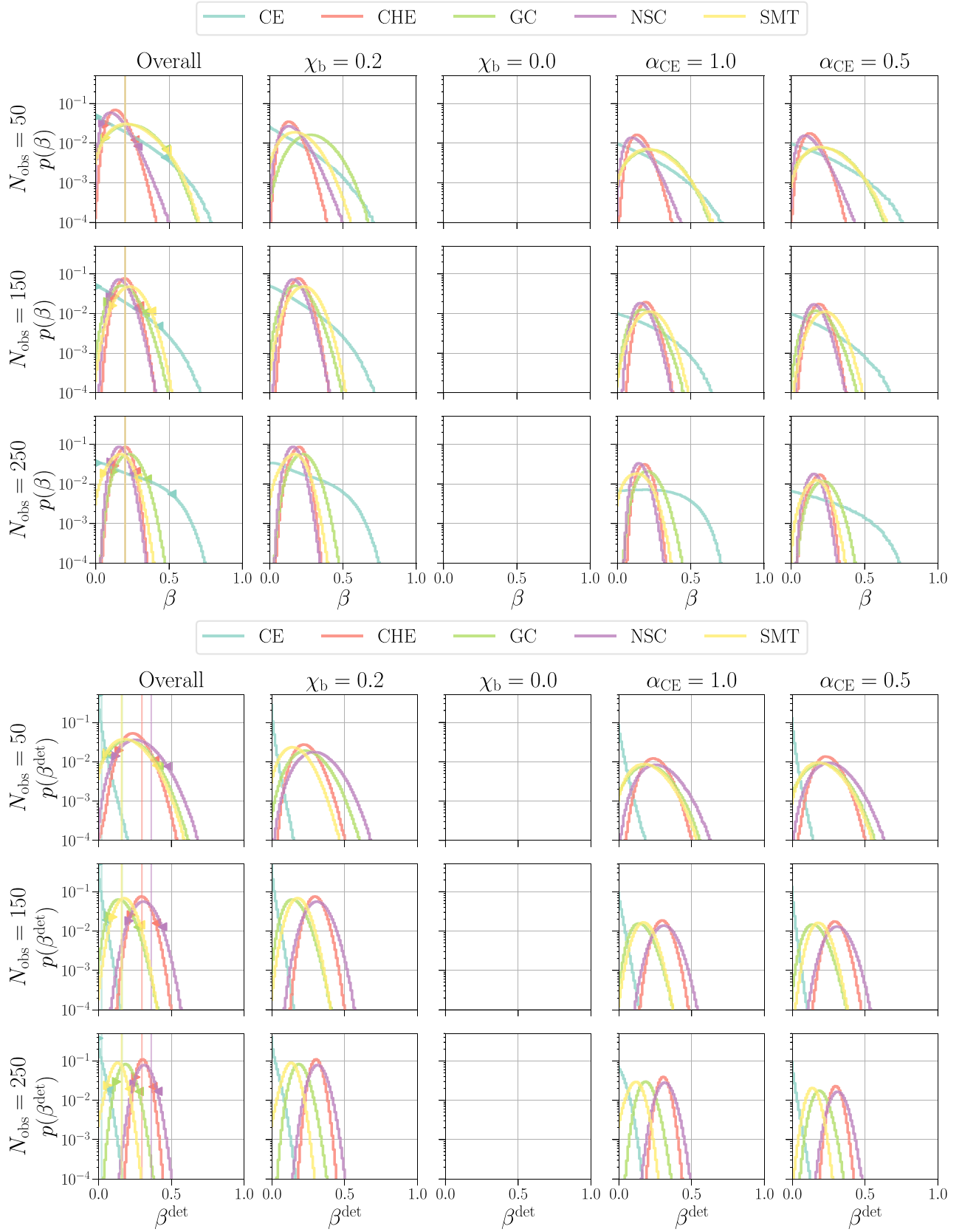
Figure 12. The same as Figure 3 (top, underlying branching fractions) and Figure 4 (bottom, detectable branching fractions), showing the convergence of the posteriors with $N_{\rm obs}$, except in this universe we have $\beta = 20\%$ across all channels, as well as $\chi_{\rm b} = 0.2$ and $\alpha_{\rm CE} = 1.0$. As in Figures 3 and 4, the empty plots (middle column) represent the lack of support for the $\chi_{\rm b} = 0.0$ model, while the populated plots corresponding to $\alpha_{\rm CE} = 0.5$ (right column) show that the inference does not significantly favor the true value of $\alpha_{\rm CE} = 1.0$.
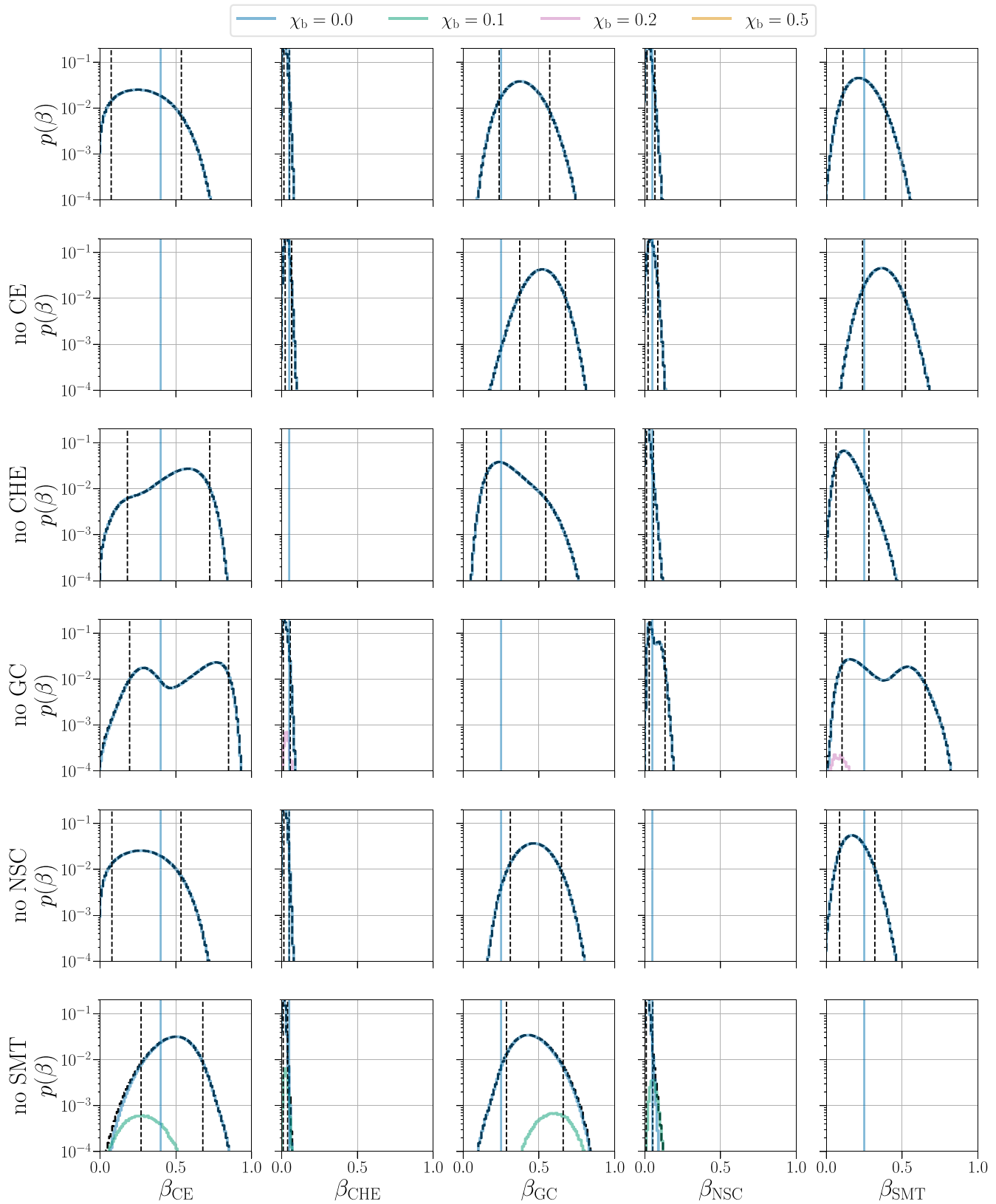
Figure 13. The same as Figure 9, showing the marginalized branching fraction posteriors as a single channel is excluded from the inference, except in this universe we have unequal branching fractions, $\chi_{\rm b} = 0.0$, and $\alpha_{\rm CE} = 1.0$. We also use $N_{\rm obs} = 250$ for this analysis.
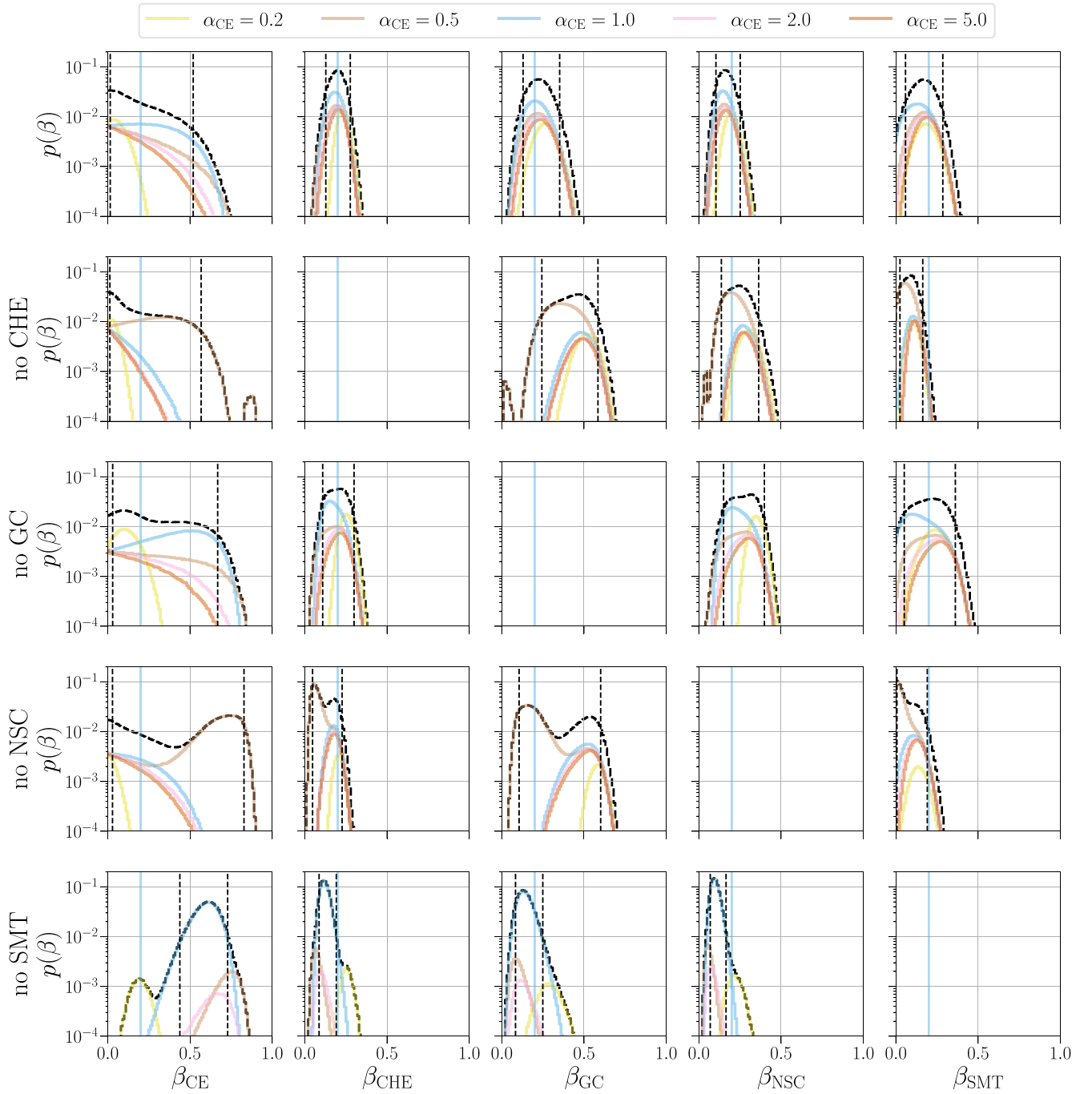
Figure 14. The same as Figure 9, showing the marginalized branching fraction posteriors for each channel in an equal branching fraction universe, except here we show with the colored curves support for different values of α_{CE}, marginalized over χ_b. The blue vertical lines show the true values of the branching fractions and of $\alpha_{CE} = 1.0$. We note that no value of α_{CE} is inferred when CE is not included in the inference, so we do not show it in this plot. There are two scenarios in which the wrong value of α_{CE} is strongly preferred: when CHE is excluded (second row) and when NSC is excluded (fourth row). The selection of a lower value of α_{CE} when CHE is excluded is likely due to the necessity of tidally spun-up CE BBHs to explain the high-χ_{eff} CHE BBHs. Additionally, the posteriors of the branching fractions (especially β_{CE}) can have varying support depending on the value of α_{CE}. It is possible that the flexibility of the CE channel from being able to take on different values of α_{CE} can cause bias toward higher support for β_{CE} when marginalizing over α_{CE}, but future investigations involving the exclusion of α_{CE} models from the inference are necessary to confirm this effect.

ORCID iDs

April Qiu Cheng ⬤ https://orcid.org/0009-0007-8996-0735
Michael Zevin ⬤ https://orcid.org/0000-0002-0147-0835
Salvatore Vitale ⬤ https://orcid.org/0000-0003-2700-0767

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, LRR, 21, 3
Abbott, R., Abbott, T. D., Acernese, F., et al. 2023, PhRvX, 13, 011048
Acernese, F., Agathos, M., Agatsuma, K., et al. 2015, CQGra, 32, 024001

Adamcewicz, C., & Thrane, E. 2022, MNRAS, 517, 3928
Antonini, F., & Rasio, F. A. 2016, ApJ, 831, 187
Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27
Askar, A., Szkudlarek, M., Gondek-Rosińska, D., Giersz, M., & Bulik, T. 2017, MNRAS, 464, L36
Baibhav, V., Doctor, Z., & Kalogera, V. 2023, ApJ, 946, 50
Baird, E., Fairhurst, S., Hannam, M., & Murphy, P. 2013, PhRvD, 87, 024035
Bavera, S. S., Fragos, T., Zevin, M., et al. 2021, A&A, 647, A153
Belczynski, K., Kalogera, V., & Bulik, T. 2002, ApJ, 572, 407
Belczynski, K., Repetto, S., Holz, D. E., et al. 2016, ApJ, 819, 108
Bethe, H. A., & Brown, G. E. 1998, ApJ, 506, 780
Biscoveanu, S., Callister, T. A., Haster, C.-J., et al. 2022, ApJL, 932, L19
Broekgaarden, F. S., Berger, E., Stevenson, S., et al. 2022, MNRAS, 516, 5737
Buonanno, A., Kidder, L. E., & Lehner, L. 2008, PhRvD, 77, 026004
Callister, T. A. 2021, arXiv:2104.09508
Callister, T. A., & Farr, W. M. 2023, arXiv:2302.07289
Callister, T. A., Haster, C.-J., Ng, K. K. Y., Vitale, S., & Farr, W. M. 2021, ApJL, 922, L5
Callister, T. A., Miller, S. J., Chatziioannou, K., & Farr, W. M. 2022, ApJL, 937, L13
Cheng, A. Q. 2023, Use and Abuse of Astrophysical Models in Gravitational-wave Population Analyse, v1, Zenodo, doi:10.5281/zenodo.8117875
de Mink, S. E., & Mandel, I. 2016, MNRAS, 460, 3545
Dominik, M., Belczynski, K., Fryer, C., et al. 2012, ApJ, 759, 52
Downing, J. M. B., Benacquista, M. J., Giersz, M., & Spurzem, R. 2010, MNRAS, 407, 1946
Edelman, B., Doctor, Z., Godfrey, J., & Farr, B. 2022, ApJ, 924, 101
Edelman, B., Farr, B., & Doctor, Z. 2023, ApJ, 946, 16
Eldridge, J. J., & Stanway, E. R. 2016, MNRAS, 462, 3302
Farah, A. M., Edelman, B., Zevin, M., et al. 2023, arXiv:2301.00834
Farr, W. M., Gair, J. R., Mandel, I., & Cutler, C. 2015, PhRvD, 91, 023005
Franciolini, G., Baibhav, V., De Luca, V., et al. 2022, PhRvD, 105, 083526
Fregeau, J. M., & Rasio, F. A. 2007, ApJ, 658, 1047
Galaudage, S., Talbot, C., Nagar, T., et al. 2021, ApJL, 921, L15
Gallegos-Garcia, M., Berry, C. P. L., Marchant, P., & Kalogera, V. 2021, ApJ, 922, 110
Giacobbo, N., & Mapelli, M. 2018, MNRAS, 480, 2011
Golomb, J., & Talbot, C. 2022, arXiv:2210.12287
González, J. A., Sperhake, U., Brügmann, B., Hannam, M., & Husa, S. 2007, PhRvL, 98, 091101
Gültekin, K., Miller, M. C., & Hamilton, D. P. 2006, ApJ, 640, 156
Hong, J., & Lee, H. M. 2015, MNRAS, 448, 754
Hut, P., McMillan, S., Goodman, J., et al. 1992, PASP, 104, 981
Ivanova, N., Justham, S., Chen, X., et al. 2013, A&ARv, 21, 59
Kass, R. E., & Raftery, A. E. 1995, JASA, 90, 773, http://www.jstor.org/stable/2291091
Kimball, C., Berry, C., & Kalogera, V. 2020, RNAAS, 4, 2
Kremer, K., Ye, C. S., Rui, N. Z., et al. 2020, ApJS, 247, 48
Kroupa, P. 2001, MNRAS, 322, 231
Lightman, A. P., & Shapiro, S. L. 1978, RvMP, 50, 437
LIGO Scientific Collaboration, Aasi, J., Abbott, B. P., et al. 2015, CQGra, 32, 074001
Maccarone, T. J. 2021, JAAVSO, 49, 83
Mandel, I., & Broekgaarden, F. S. 2022, LRR, 25, 1
Mandel, I., & de Mink, S. E. 2016, MNRAS, 458, 2634
Mandel, I., & Farmer, A. 2022, PhR, 955, 1
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
Mapelli, M. 2021, Handbook of Gravitational Wave Astronomy (Berlin: Springer), 16
Marchant, P., Langer, N., Podsiadlowski, P., Tauris, T. M., & Moriya, T. J. 2016, A&A, 588, A50
McMillan, S., Hut, P., & Makino, J. 1991, ApJ, 372, 111

Miller, M. C., & Hamilton, D. P. 2002, ApJ, 576, 894
Miller, M. C., & Lauburg, V. M. 2009, ApJ, 692, 917
Neijssel, C. J., Vigna-Gómez, A., Stevenson, S., et al. 2019, MNRAS, 490, 3740
Nitz, A. H., Kumar, S., Wang, Y.-F., et al. 2023, ApJ, 946, 59
O'Leary, R. M., Kocsis, B., & Loeb, A. 2009, MNRAS, 395, 2127
O'Leary, R. M., Rasio, F. A., Fregeau, J. M., Ivanova, N., & O'Shaughnessy, R. 2006, ApJ, 637, 937
Olsen, S., Venumadhav, T., Mushkin, J., et al. 2022, PhRvD, 106, 043009
O'Reilly, B. O., Branchesi, M., Haino, S., et al. 2022, Noise curves used for Simulations in the update of the Observing Scenarios Paper, https://dcc.ligo.org/LIGO-T2000012/public
Paczynski, B. 1976, in IAU Symp. 73, Structure and Evolution of Close Binary Systems, ed. P. Eggleton, S. Mitton, & J. Whelan (Berlin: Springer), 75
Portegies Zwart, S. F., & McMillan, S. L. W. 2000, ApJL, 528, L17
Pratten, G., García-Quirós, C., Colleoni, M., et al. 2021, PhRvD, 103, 104056
Pretorius, F. 2005, PhRvL, 95, 121101
Pürrer, M., Hannam, M., & Ohme, F. 2016, PhRvD, 93, 084042
Qin, Y., Fragos, T., Meynet, G., et al. 2018, A&A, 616, A28
Rinaldi, S., & Del Pozzo, W. 2022, MNRAS, 509, 5454
Rodriguez, C. L., Chatterjee, S., & Rasio, F. A. 2016, PhRvD, 93, 084029
Rodriguez, C. L., Kremer, K., Chatterjee, S., et al. 2021, RNAAS, 5, 19
Rodriguez, C. L., Morscher, M., Pattabiraman, B., et al. 2015, PhRvL, 115, 051101
Roulet, J., Chia, H. S., Olsen, S., et al. 2021, PhRvD, 104, 083010
Samsing, J., MacLeod, M., & Ramirez-Ruiz, E. 2014, ApJ, 784, 71
Samsing, J., & Ramirez-Ruiz, E. 2017, ApJL, 840, L14
Sigurdsson, S., & Hernquist, L. 1993, Natur, 364, 423
Sigurdsson, S., & Phinney, E. S. 1993, ApJ, 415, 631
Stevenson, S., Vigna-Gómez, A., Mandel, I., et al. 2017, NatCo, 8, 14906
Talbot, C., & Thrane, E. 2017, PhRvD, 96, 023012
Talbot, C., & Thrane, E. 2018, ApJ, 856, 173
The LIGO Scientific Collaborationthe Virgo Collaboration, Abbott, R., et al. 2021a, arXiv:2108.01045
The LIGO Scientific Collaborationthe Virgo Collaborationthe KAGRA Collaboration, et al. 2021b, arXiv:2111.03606
Thrane, E., & Talbot, C. 2019, PASA, 36, e010
Tiwari, V. 2021, CQGra, 38, 155007
Tiwari, V. 2022, ApJ, 928, 155
Tutukov, A. V., & Yungelson, L. R. 1993, MNRAS, 260, 675
van den Heuvel, E. P. J. 1976, in IAU Symp. 73, Structure and Evolution of Close Binary Systems, ed. P. Eggleton, S. Mitton, & J. Whelan (Berlin: Springer), 35
van den Heuvel, E. P. J., Portegies Zwart, S. F., & de Mink, S. E. 2017, MNRAS, 471, 4256
van Son, L. A. C., de Mink, S. E., Renzo, M., et al. 2022, ApJ, 940, 184
Veske, D., Bartos, I., Márka, Z., & Márka, S. 2021, ApJ, 922, 258
Vitale, S., Biscoveanu, S., & Talbot, C. 2022a, A&A, 668, L2
Vitale, S., Farr, W. M., Ng, K. K. Y., & Rodriguez, C. L. 2019, ApJL, 886, L1
Vitale, S., Gerosa, D., Farr, W. M., & Taylor, S. R. 2022b, Handbook of Gravitational Wave Astronomy (Berlin: Springer), 45
Vitale, S., Lynch, R., Raymond, V., et al. 2017a, PhRvD, 95, 064053
Vitale, S., Lynch, R., Sturani, R., & Graff, P. 2017b, CQGra, 34, 03LT01
Vitale, S., Lynch, R., Veitch, J., Raymond, V., & Sturani, R. 2014, PhRvL, 112, 251101
Weizmann Kiendrebeogo, R., Farah, A. M., Foley, E. M., et al. 2023, arXiv:2306.09234
Zaldarriaga, M., Kushnir, D., & Kollmeier, J. A. 2018, MNRAS, 473, 4174
Zevin, M., Bavera, S. S., Berry, C. P. L., et al. 2021a, ApJ, 910, 152
Zevin, M., Romero-Shaw, I. M., Kremer, K., Thrane, E., & Lasky, P. D. 2021b, ApJL, 921, L43
Ziosi, B. M., Mapelli, M., Branchesi, M., & Tormen, G. 2014, MNRAS, 441, 3703