Draft version December 18, 2022
Typeset using LaTeX **twocolumn** style in AASTeX631

Deep follow-up for gravitational-wave inference: a case study with GW151226

Avi Vajpeyi,[1,2] Rory Smith,[1,2] and Eric Thrane[1,2]

[1] *School of Physics and Astronomy, Monash University, VIC 3800, Australia*
[2] *OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Monash University, VIC 3800, Australia*

ABSTRACT

New analyses of gravitational wave events raise questions about the nature of some events. For example, LIGO–Virgo–KAGRA initially determined GW151226 to be a merger with a mass-ratio $q \approx 0.5$ and effective inspiral spin $\chi_{\rm eff} \approx 0.2$. However, recent works offer an alternative picture: GW151226 is a lower-mass-ratio event $q \approx 0.3$ with slightly higher spin $\chi_{\rm eff} \approx 0.3$. This discrepancy has been challenging to resolve as a wide range of differences are employed for each analysis. In this work, we introduce a "deep follow-up" framework to efficiently compute the posterior odds between two different peaks in parameter space. In doing so, we aim to help resolve disputes about the true nature of gravitational-wave events associated with conflicting astrophysical interpretations. Our proposal is not a replacement for standard inference techniques; rather, our method provides a diagnostic tool to better understand discrepancies between conflicting results. We demonstrate this method by studying three q-$\chi_{\rm eff}$ peaks proposed for GW151226. We find the $(q \sim 0.5, \chi_{\rm eff} \sim 0.2)$ interpretation is only slightly preferred over the $(q \sim 0.3, \chi_{\rm eff} \sim 0.3)$ hypothesis with a posterior odds of $\sim 1.7 \pm 0.4$, suggesting that neither of the two GW151226 can be ruled out. We discuss strategies to produce more reliable parameter estimation studies in gravitational-wave astronomy.

1. INTRODUCTION

The binary black hole merger GW151226 was the second gravitational-wave event detected (Abbott et al. 2016). The LIGO-Virgo-Kagra (LVK) collaboration (The LIGO Scientific Collaboration et al. 2015; Acernese et al. 2015) measured the event with a signal to noise ratio SNR ~ 13, source-frame masses of $(m_1, m_2) = (14.2^{+8.3}_{-3.7} M_\odot, 7.5^{+2.3}_{-2.3} M_\odot$, 90% credibility) and with the mass ratio $q \equiv m_2/m_1 \approx 0.5$ (Abbott et al. 2016).[1] The event was noteworthy at the time of detection because—unlike GW150914, the first gravitational-wave detection—GW151226 showed signs of black hole spin. The effective inspiral parameter,

$$\chi_{\rm eff} \equiv \frac{\chi_1 \cos\theta_1 + q\chi_2 \cos\theta_2}{1+q} \tag{1}$$

is a measure of the black hole spin projected onto the orbital angular momentum axis.[2] This parameter was found to be inconsistent with zero at high credibility $\chi_{\rm eff} = 0.18^{+0.25}_{-0.08}$, suggesting that at least one of the black holes in GW151226 was measurably spinning (Abbott et al. 2016). A subsequent analysis by Mateu-Lucena et al. (2021) draws a similar conclusion using a more sophisticated gravitational waveform (`IMRPhenomXPHM`, Pratten et al. 2021), which incorporates both higher multipoles and precession, than what was available for the initial analysis. These results suggest GW151226 is a fairly typical merger; most events in the LIGO–Virgo gravitational-wave transient catalog (The LIGO Scientific Collaboration et al. 2021) are consistent with $q > 0.3$ and approximately 20% exhibit signs of measurable spin (Roulet et al. 2021; Galaudage et al. 2021).

However, Chia et al. (2021) suggest this interpretation of GW151226 is incomplete. Using the same waveform as Mateu-Lucena *et al.*, Chia *et al.* reanalyzed GW151226 and found a low-q likelihood peak with source-frame $(m_1, m_2) = (19.1^{+12.1}_{-7.1} M_\odot, 5.8^{+2.9}_{-1.7} M_\odot)$ and $q = 0.30^{+0.22}_{-0.13}$. Due to the degeneracy between q and $\chi_{\rm eff}$, the low-q peak is associated with a larger effective inspiral spin $\chi_{\rm eff} = 0.30^{+0.17}_{-0.11}$. The low-$q$ interpretation makes GW151226 a more interesting event. With an asymmetric mass ratio $q \sim 3$, GW151226 is less like typical events in GWTC-3 (The LIGO Scientific Collaboration et al. 2021) and more like the mysterious GW190814 (Abbott

avi.vajpeyi@monash.edu

[1] The subscript 1 refers to the primary, more massive black hole while the 2 refers to the less massive secondary.

[2] $\chi_{1,2}$ are the dimensionless black-hole spins and $\theta_{1,2}$ are the tilt angles between the spin vectors and the orbital angular momentum vector.

et al. 2020a)—an event which is difficult to explain in the context of standard binary evolution. To simultaneously achieve such a large value of $\chi_{\rm eff}$ and a small value of q, the primary black hole must have non-zero spin (Qin et al. 2022). However, this is somewhat challenging to explain in the usual framework of tidal interactions, which serve to spin up only the second-born (typically less massive) black hole (Mandel & Fragos 2020).[3] Finally, Chia *et al.* suggest that GW151226 exhibits clear signs of orbital precession, making this the first system to exhibit this affect (Green et al. 2021; Abbott et al. 2021).

Given the implications of the Chia *et al.* results, it is important to understand why this result differs from the original LVK result (Abbott et al. 2016) and the more recent analysis by Mateu-Lucena et al. (2021) and Nitz et al. (2021a) which find posterior peaks at $(q, \chi_{\rm eff}) = (0.52^{+0.27}_{-0.21}, 0.20^{+0.10}_{-0.06})$ and $(0.53^{+0.24}_{-0.23},\ 0.22^{+0.10}_{-0.06})$ respectively.[4] In attempting to reconcile their results with Chia *et al.*, Mateu-Lucena *et al.* perform several cross-checks and suggest that the low-q mode observed by Chia *et al.* could be due to a too-low value of sampling frequency. The analysis by Nitz *et al.* failed to break the stalemate—obtaining results broadly consistent with Mateu-Lucena *et al.* while simultaneously finding weak support for a low-q peak. Since both Chia *et al.* and Mateu-Lucena *et al.* use the same gravitational waveform in their analyses (which include higher order modes and precession), Chia *et al.* speculate that the difference in posterior estimates is because they employ a uniform-in-$\chi_{\rm eff}$ prior, rather than an isotropic spin prior (which penalises sampling in the large $\chi_{\rm eff}$ regions).[5] This creates a predicament: the community has two conflicting interpretations of the GW151226 event.

In this work, we demonstrate a method to resolve the tension between two dissimilar sets of parameter estimation results. This method allows us to compute odds comparing two points in parameter space. These odds can be used to ascertain which of two or more inference results are more probable, useful for understanding discrepancies like those described above. We apply this technique to GW151226 to settle the debate on the relative importance of the low-q and high-q likelihood peaks. We compute the marginal likelihood of two points in the q-$\chi_{\rm eff}$ distribution, one from Chia *et al.*'s low-q likelihood peak and another from the Mateu-Lucena *et al.*'s high-q peak. By restricting the analysis to the two-dimensional q-$\chi_{\rm eff}$ space, we reduce the problem's dimensionality, permitting a "deep follow-up" investigation on the q-$\chi_{\rm eff}$ modes of GW151226 without the need for large computational resources. While we demonstrate this method on GW151226, the same principles can be applied to resolve other discrepancies in the gravitational-wave astronomy literature. The remainder of this letter is organized as follows. In Section 2, we describe the "deep follow-up" formalism used to compute our posterior odds. Then, in Section 3, we show results obtained from a deep follow-up of GW151226. We provide closing comments in Section 4.

2. METHODOLOGY

Different analyses may find different peaks in the marginal likelihood distribution. In this section, we provide a mathematical method to determine which peak is more probable than another. We label the two disparate likelihood peaks A and B and ask the question: what are the posterior odds of hypothesis A over hypothesis B? If this odds, denoted $\mathcal{O}^A_B$, is much larger (or much smaller) than unity ($\ln \mathcal{O} \gg 8$ or $\ln \mathcal{O} \ll -8$), then one hypothesis can be said to be strongly preferred over the other. When this happens, we may conclude that one hypothesis provides a better description of the posterior distribution.[6] On the other hand, if $\mathcal{O} \approx 1$, this implies that both hypotheses are approximately equally consistent with the data. When this happens, we may conclude that one or more analysis is incomplete insofar as they did not produce a posterior consistent with both A and B.

In the case of GW151226, we choose A and B from Mateu-Lucena *et al.* and Chia *et al.* peaks in the $(q, \chi_{\rm eff})$ plane because different peaks in this plane can lead to unique physical interpretations of the event. As we are only interested in this plane, the marginal likelihood distribution $\mathcal{Z}(d|q, \chi_{\rm eff})$ allows us to focus on just the 2 parameters at a time; the other parameters (e.g., total mass, the effective precession parameter, etc.) are integrated out. By fixing the parameter values associated with points A and B we reduce the dimensionality of our inference calculation, which improves convergence; for our GW151226 analysis, the dimensionality of the

[3] In some events, large primary spin can arise due to the phenomenon of "mass ratio reversal" (Olejak & Belczynski 2021), though, this seems unlikely to have occurred in a system with such a small mass ratio.

[4] Since this work began, Nitz *et al.* published updated posteriors for GW151226 in Nitz et al. (2021b), with a q-$\chi_{\rm eff}$ posterior peak at $(0.52^{-0.2}_{+0.26}, 0.21^{-0.06}_{+0.1})$.

[5] There are other subtle differences such as different stochastic samplers, different noise estimation techniques, different sampling frequencies, and different methods for likelihood estimation.

[6] More precisely, one hypothesis is preferred over the other given whatever assumptions we have made constructing our likelihood function.

marginalized likelihood is reduced by two. Moreover, we do not have to worry about our stochastic sampler failing to find peaks A and B—we force the sampler to study just these two points in parameter space. By setting up the problem this way, we give up trying to map the full $(q, \chi_{\rm eff})$ subspace, focusing our computational power so that we obtain reliable estimates of the marginal likelihood at just the two points A and B. If the inference calculations by Mateu-Lucena et al. (2021) and Chia et al. (2021) are analogous to all-sky searches, this analysis is analogous to a deep follow-up study.

The posterior odds are

$$\mathcal{O}^{\rm A}_{\rm B} = \frac{Z(d|A)}{Z(d|B)}\,\frac{\pi(A)}{\pi(B)}, \tag{2}$$

where $Z(d|...)$ refers to the marginal likelihood for hypothesis A or B and $\pi(...)$ refers to the prior odds for hypothesis A or B. The marginal likelihood for point A or B is given by

$$\mathcal{Z}(d|A,B) = \int \mathcal{L}(d|\theta)\,\pi(\theta|A,B), \tag{3}$$

where $\mathcal{L}(d|\theta)$ is the usual Whittle likelihood used in gravitational-wave astronomy (see, e.g., Thrane & Talbot (2019)) and $\pi(\theta|A,B)$ is the prior on θ conditioned upon hypothesis A, B. We first discuss how to calculate the marginal likelihood and then describe how to calculate the prior odds.

The marginal likelihood. Our first step is to define the priors conditioned on hypotheses A, B. In this study, $\pi(\theta|A,B)$ is the isotropic spin prior from Mateu-Lucena et al. (2021), conditioned on $(q, \chi_{\rm eff})_{A,B}$, the location of either Mateu-Lucena *et al.*'s q-$\chi_{\rm eff}$ peak A or Chia *et al.*'s q-$\chi_{\rm eff}$ peak B. The isotropic spin prior is uniform in the dimensionless spin parameters χ_1, χ_2, the azimuthal angle of the spin vectors ϕ_1, ϕ_2, and the cosine of the spin-vector zenith angle z_1, z_2. The prior distribution of black hole masses is uniform in (m_1, m_2) over the interval $(3M_\odot, 54M_\odot)$ subject to the constraint that $m_2 < m_1$. We employ a uniform-in-comoving volume prior for luminosity distance and assume standard priors for the other extrinsic parameters.

The next step is to condition the isotropic spin prior on $(q, \chi_{\rm eff})A, B$. Using Eq. 1, we can express our prior on $\chi_{\rm eff}$ as delta function:

$$\pi(\chi_{\rm eff}|\chi_1, z_1, z_2, \chi_2, q_{A,B}) = \delta\left(\chi_{\rm eff} - \frac{\chi_1 z_1 + q_{A,B}\chi_2 z_2}{1+q_{A,B}}\right). \tag{4}$$

Here, $q_{A,B}$ refer to the mass ratio associated with either hypothesis A or hypothesis B. By changing variables, we obtain a conditional prior on χ_2 for hypotheses A and B:

$$\pi(\chi_2|\chi_1, z_1, z_2)_{A,B} = \delta\left(\chi_2 - \frac{\chi_{\rm eff}^{A,B}(1+q_{A,B}) - \chi_1 z_1}{q_{A,B} z_2}\right). \tag{5}$$

Combining our isotropic spin prior with Eq. 5, we obtain the marginal prior distribution for χ_1 conditioned on A, B[7]

$$\begin{aligned}\pi(\chi_1)_{A,B} &= \int dz_1\, dz_2\, d\chi_2\, \pi(\chi_1, z_1, z_2, \chi_2)_{A,B} \\ &= \int dz_1\, dz_2\, d\chi_2\, \pi(\chi_1, z_1, z_2)\, \pi(\chi_2|\chi_1, z_1, z_2)_{A,B} \\ &\propto \frac{1}{n}\sum_k^n \mathcal{I}\left(0 < \frac{\chi_{\rm eff}^{A,B}(1+q_{A,B}) - \chi_1 z_1^k}{q_{A,B} z_2^k} < 1\right)\end{aligned} \tag{6}$$

To obtain the final expression, we wrote the integrals over z_1, z_2 as a sum over prior samples drawn from the $\pi(z_1)$ and $\pi(z_2)$ distributions. The symbol $\mathcal{I}$ denotes an indicator function, which returns zero when its argument is false and one when its argument is true. Equation 6 can be used to generate random samples of χ_1 with inverse transform sampling conditioned on either hypothesis A or hypothesis B. We denote these samples $\{\chi_1^j\}_{A,B}$

The next step is to add values of z_1 to the χ_1 samples. The marginal prior distribution for z_1 conditioned on χ_1^j and hypothesis A or B is

$$\begin{aligned}&\pi(z_1|\chi_1^j)_{A,B} \\ &\quad\propto \frac{1}{n}\sum_k^n \mathcal{I}\left(0 < \frac{\chi_{\rm eff}^{A,B}(1+q_{A,B}) - \chi_1^j z_1}{q z_2^k} < 1\right).\end{aligned} \tag{7}$$

This time, the sum over k is a sum over prior samples for z_2. We use this distribution to generate random samples z_1^j given χ_1^j (see Fig. 1).[8] Using the same logic, the prior distribution for z_2 conditioned on χ_1^j, z_1^j and hypothesis A or B is

$$\pi(z_2|\chi_1^j, z_1^j)_{A,B} \propto \mathcal{I}\left(0 < \frac{\chi_{\rm eff}^{A,B}(1+q_{A,B}) - \chi_1^j z_1^j}{q z_2} < 1\right). \tag{8}$$

This distribution allows us to generate a random sample z_2^j. For each sample, we now have $\chi_{\rm eff}, q, \chi_1^j, z_1^j, z_2^j$, so the value for χ_2 is completely determined by Eq. 1.

[7] Probability densities in equations 6,7, 8 and 10 are normalized manually.

[8] We compute the distribution in Eq. 7 on a two-dimensional grid in (z_1, χ_1). We perform inverse-transform sampling using columns from this grid.

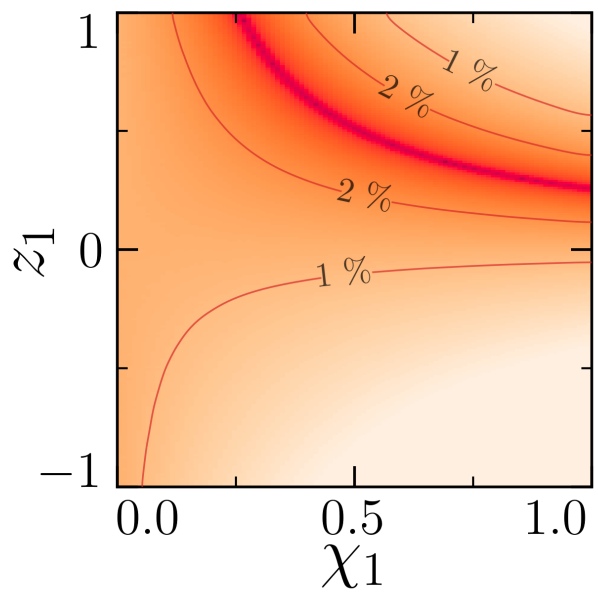

Figure 1. Heatmap of $\pi(z_1, \chi_1|q, \chi_{\rm eff})$. Contours are drawn at the 1% and 2% percentiles. Here, $(q, \chi_{\rm eff}) = (0.68, 0.15)$.

Using these equations (6-8), we define priors for our stochastic sampler that are conditioned on hypotheses A and B.[9] We perform two parameter estimation runs, one for hypothesis A and one for B. The evidence from each run yields the marginal likelihood defined in Eq. 3. The ratio of these two evidence values gives us the Bayes factor

$$\mathrm{BF} = \frac{\mathcal{Z}(d|A)}{\mathcal{Z}(d|B)}, \tag{9}$$

which represents the relative probability for hypothesis A over B given even prior odds. However, the isotropic spin prior described above does not assign even prior odds; small values of $\chi_{\rm eff}$ are more likely than larger ones. Thus, to quantify the posterior odds, we now move on to the prior odds: $\pi_{A,B} = \pi(q_{A,B}, \chi_{\rm eff}^{A,B})$.[10]

The prior odds. The marginal prior probability density for $(q, \chi_{\rm eff})$ can be calculated like so:

$$\pi(q, \chi_{\rm eff}) \propto \frac{2}{n} \sum_k^n \left| \frac{1 + q\chi_{\rm eff}}{\chi_1^k} \right| \left(\frac{\chi_1^k + q\chi_2^k z_2^k}{1+q} \right). \tag{10}$$

[9] We check the calculations by generating corner plots of our prior samples conditioned on A, B and comparing them to corner plots created with the isotropic spin prior, but keeping only samples in the vicinity of $(q_{A,B}, \chi_{\rm eff}^{A,B})$.

[10] For the time being, we duck the question of whether the isotropic spin prior is a good (astrophysically-motivated) prior. Our short-term goal is to compute the posterior odds given that choice of prior.

Here, the sum over k is a sum over isotropic spin prior samples for (χ_1, χ_2, z_2). What follows is a derivation of Eq. 10. To begin, we note that

$$\pi(q, \chi_{\rm eff}) = \pi(\chi_{\rm eff}|q)\, \pi(q). \tag{11}$$

We rewrite the first term in terms of marginalisation integrals to obtain

$$\pi(q, \chi_{\rm eff}) = \int d\chi_1\, d\chi_2\, dz_2\, \pi(\chi_{\rm eff}|q, \chi_1, \chi_2, z_2)\, \pi(q, \chi_1, \chi_2, z_2). \tag{12}$$

We have a closed-form expression for the isotropic spin prior $\pi(q, \chi_1, \chi_2, z_2)$. The remaining term in Eq. 12 is related to the fiducial distribution for z_1. Since z_1 is uniformly distributed, $\chi_{\rm eff}$ is uniformly distributed with limits depending on (χ_1, q, χ_2, z_2):

$$\begin{aligned} &\pi(\chi_{\rm eff}|q, \chi_1, \chi_2, z_2) \\ &\quad = \left| \frac{\partial z_1}{\partial \chi_{\rm eff}} \right| U\left[\frac{-\chi_1 + q\chi_2 z_2}{1+q}, \frac{\chi_1 + q\chi_2 z_2}{1+q} \right] \\ &\quad = \left| \frac{1 + q\chi_{\rm eff}}{\chi_1} \right| U\left[\frac{-\chi_1 + q\chi_2 z_2}{1+q}, \frac{\chi_1 + q\chi_2 z_2}{1+q} \right]. \end{aligned} \tag{13}$$

Plugging Eq. 13 into Eq. 12 and replacing the marginalisation integrals with a sum over prior samples, we obtain Eq. 10.

Sampling. For hypothesis A (the high-q peak), we set $(q, \chi_{\rm eff}) = (0.52, 0.2)$. For hypothesis B (the low-q peak), we set $(q, \chi_{\rm eff}) = (0.3, 0.3)$. These values are the median q-$\chi_{\rm eff}$ values from the Mateu-Lucena *et al.* and Chia *et al.* posteriors and are plotted in Figure 2. We compute the marginalised likelihood for each hypothesis using `Parallel Bilby` (Ashton et al. 2019; Smith et al. 2020), employing `Dynesty` (Speagle 2020) as our nested sampler (Skilling 2004, 2006). We use 2,000 live points, a data sampling rate of 4096 Hz and the `IMRPhenomXPHM` waveform. The configuration files for our analyses are located on this project's GitHub repository (Vajpeyi 2021).

Validation. To validate this method, we compute the q-$\chi_{\rm eff}$ deep follow-up posterior odds $\mathcal{O}_{\rm B}^{\rm A}$ for two additional cases: a GW151226-like signal injected in Gaussian noise and GW150914, displayed in Table 1. We apply a kernel density estimate (KDE) of the two signals' posterior samples in the $(q, \chi_{\rm eff})$ plane. The $\mathcal{O}_{\rm B}^{\rm A}$ fall within the uncertainty of the ${}^{K}\mathcal{O}_{\rm B}^{\rm A}$, demonstrating that the deep follow-up method yields consistent posterior odds with KDE posterior odds for the signals. Details of the injection and analyses can be found in the same repository (Vajpeyi 2021).

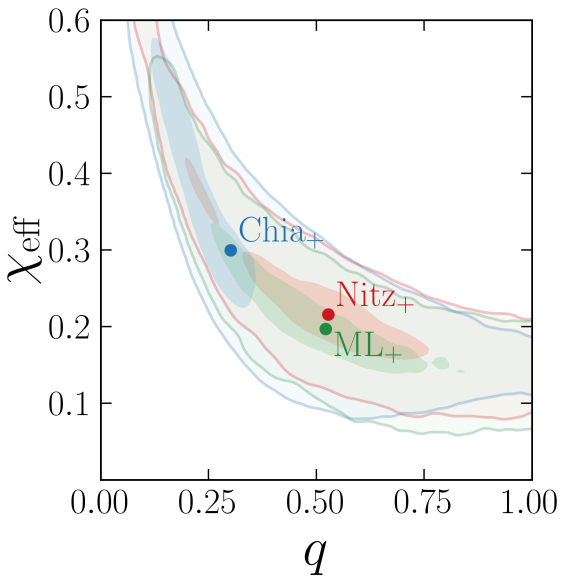

Figure 2. GW151226 q-$\chi_{\rm eff}$ posteriors from Mateu-Lucena et al. (2021) ML$_+$, Chia et al. (2021) Chia$_+$, Nitz et al. (2021a) Nitz$_+$ analyses. Contours are drawn at the 1,3-σ regions. The points mark the median q-$\chi_{\rm eff}$ values of each analysis.

Table 1. Comparisons of the deep follow-up posterior odds $\mathcal{O}^{\rm A}_{\rm B}$ to kernel density estimate posterior odds ${}^{K}\mathcal{O}^{\rm A}_{\rm B}$ for a GW151226-like simulated signal and GW150914.

	Simulated	GW150914
$A\,(q, \chi_{\rm eff})$	$0.36, 0.17$	$0.7, 0.02$
$B\,(q, \chi_{\rm eff})$	$0.39, 0.25$	$0.78, -0.04$
$\mathcal{O}^{\rm A}_{\rm B}$	0.47 ± 0.11	0.24 ± 0.1
${}^{K}\mathcal{O}^{\rm A}_{\rm B}$	0.47 ± 0.02	0.24 ± 0.01

3. RESULTS

Our deep follow-up findings for GW151226 are summarized in Table 2. Assigning the high-q, Mateu-Lucena *et al.* q-$\chi_{\rm eff}$ peak, to hypothesis A and the low-q, Chia *et al.* q-$\chi_{\rm eff}$ peak, to hypothesis B, we obtain a Bayes factor of 1.0 ± 0.2, implying that—if we assign even prior odds to hypotheses A and B—both peaks are equally preferred. However, the isotropic spin prior odds are 1.68 ± 0.01 in favour of the hypothesis A. Thus, we obtain a posterior odds $\mathcal{O}^{\rm A}_{\rm B} = 1.7 \pm 0.4$, implying that high-$q$ peak is only slightly preferred over low-q peak. Comparing the Nitz *et al.* and Chia *et al.* q-$\chi_{\rm eff}$ peaks yeilds similar results: a slight preference of the high-q peak over the low-q peak after incorporating the isotropic spin prior odds. To compare the deep follow-up odds with the Mateu-Lucena *et al.*'s q-$\chi_{\rm eff}$ posterior, we apply a kernel density estimate (KDE) of their posterior samples in the $(q, \chi_{\rm eff})$ plane. The KDE yields a posterior odds of ${}^{K}\mathcal{O}^{\rm A}_{\rm B} = 3.45 \pm 0.01$, a value twice the deep follow-up $\mathcal{O}^{\rm A}_{\rm B}$. This suggests that—for whatever reason—the Mateu-Lucena *et al.* analysis underestimates the posterior support for the Chia *et al.* q-$\chi_{\rm eff}$ peak (by a factor of ≈ 2). At the same time, using a KDE of the Chia *et al.* samples, we find the ${}^{K}\mathcal{O}^{\rm A}_{\rm B}$ between the two points is ${}^{K}\mathcal{O}^{\rm A}_{\rm B} = 0.33 \pm 0.01$, a value one-fifth the $\mathcal{O}^{\rm A}_{\rm B}$. This suggests that the deep follow-up analysis finds less support for the Chia *et al.* q-$\chi_{\rm eff}$ peak than reported by Chia *et al.* (by a factor of $\approx 1/5$).

Table 2. The median q-$\chi_{\rm eff}$ values from Mateu-Lucena et al. (2021) ML$_+$, Chia et al. (2021) Chia$_+$, and Nitz et al. (2021a) Nitz$_+$ analyses, along with the prior odds $\pi(A)/\pi(B)$, Bayes-Factors BF, "deep follow-up" posterior odds $\mathcal{O}^{\rm A}_{\rm B}$, and kernel density estimate (KDE) posterior odds ${}^{K}\mathcal{O}^{\rm A}_{\rm B}$. The three rows of KDE posteriors odds are computed using the ML$_+$, Chia$_+$, and Nitz$_+$ q-$\chi_{\rm eff}$ samples.

	ML$_+$vs Chia$_+$	ML$_+$vs Nitz$_+$	Nitz$_+$vs Chia$_+$
$A\,(q, \chi_{\rm eff})$	$0.52, 0.20$	$0.52, 0.20$	$0.53, 0.22$
$B\,(q, \chi_{\rm eff})$	$0.30, 0.30$	$0.53, 0.22$	$0.30, 0.30$
$\pi(A)/\pi(B)$	1.68 ± 0.01	1.08 ± 0.01	1.56 ± 0.01
BF	1.0 ± 0.2	1.3 ± 0.2	0.8 ± 0.2
$\mathcal{O}^{\rm A}_{\rm B}$	1.7 ± 0.4	1.4 ± 0.3	1.2 ± 0.3
ML$_+$ ${}^{K}\mathcal{O}^{\rm A}_{\rm B}$	3.45 ± 0.01	1.46 ± 0.04	2.38 ± 0.1
Chia$_+$ ${}^{K}\mathcal{O}^{\rm A}_{\rm B}$	0.33 ± 0.01	0.93 ± 0.01	0.36 ± 0.01
Nitz$_+$ ${}^{K}\mathcal{O}^{\rm A}_{\rm B}$	1.72 ± 0.01	0.87 ± 0.01	1.97 ± 0.01

4. DISCUSSION

Increasingly, analyses of gravitational-wave events have contradicting results (e.g. GW151226, GW190425, GW190521 The LIGO Scientific Collaboration et al. 2021; Olsen et al. 2022; Nitz et al. 2021a), leaving astrophysicists to ponder: which analysis best describes the data? Faulty inferences of events can provide misleading interpretations of the population. In some cases, these analyses differ in software, sampler settings, data, waveforms, and noise models, making the determination of the "correct" analysis challenging. The method to compute posterior odds demonstrated in this work can help us understand contradictory results, allowing the community to determine which analysis best represents the posterior distribution.

In this work, we investigate the contradicting interpretations for GW151226. Chia et al. (2021) suggest that the cause for the discrepancy was due to Abbott

et al. (2016) using a waveform without both higher order multipoles and precession, and using an isotropic spin prior (leading to under-sampling in the large $\chi_{\rm EFF}$ regions of parameter space). Although some argue that the uniform-in-$\chi_{\rm eff}$ is unphysical , past work utilising uniform-in-$\chi_{\rm eff}$ spin priors has demonstrated to support regions of parameter space previously unexplored when using isotropic spin priors (Olsen et al. 2021; Nitz & Capano 2021). This would not be surprising given that uniform-in-chirp-mass priors provide more efficient sampling than uniform-in-component mass (Romero-Shaw et al. 2020).

In addition to the difference in priors used by Abbott et al. (2016) and Chia et al. (2021), the two analyses differ in the method for estimating noise power spectral density (PSD), the method for estimating the likelihood, the data sampling rates, and the data cleaning methods. Our study's output can help study the impact of the different techniques on this event. For example, one can study the impact of an alternative PSD by re-weighting the posterior samples for points A and B (publicly accessible at Vajpeyi (2021)) with a different PSD. The steps for this are described in Appendix A.

Our analysis demonstrates that the high-q and low-q likelihood peaks have comparable posterior odds. The high-q mode corresponds to a binary black hole system similar to others reported in GWTC-3 (The LIGO Scientific Collaboration et al. 2021). However, the low-q mode corresponds to a highly spinning, unequal mass ratio event with non-negligible precession (see Fig. 3). If the low-q mode describes the binary, this event would be an outlier amongst the other events in the current gravitational wave catalog.

Similar to GW151226, there are other events (e.g., GW190425 and GW190521) with alternative interpretations (Nitz et al. 2020, 2021a; Olsen et al. 2021). Nitz et al.'s analysis of GW190425 resulted in a $\chi_{\rm eff}$ slightly higher than that reported by Abbott et al. (2020). However, this is likely due to the differences in the spin priors used in the analyses—when Nitz et al. re-weighted their posteriors with a "low-spin" prior, they obtained posteriors comparable with those of Abbott et al. (2020). In the case of GW190521, while Abbott et al. (2020b), Olsen et al. (2021) and Nitz et al. (2021a) report the event to have source frame total mass of $\sim 150\,{\rm M}_\odot$, they also report different values for the event's mass ratio and spins. Abbott et al. (2020b) report the event to have an almost equal mass ratio $q \sim 0.79^{+0.19}_{-0.29}$ and negligible spin, while Nitz et al. (2021a) and Olsen et al. (2021) report the event to have a bi-modal q-$\chi_{\rm eff}$ distribution, with a low-q peak ($q < 0.3, \chi_{\rm eff} < 0$) and a high-$q$ peak ($q > 0.3, \chi_{\rm eff} \sim 0$). These bi-modal posteriors may be due to the different mass and spin priors used in the Abbott et al. (2020b) analysis; by using source-frame mass priors and uniform-in-$\chi_{\rm eff}$ prior instead of using detector-frame mass priors and isotropic spin prior. Determining if there are other events with multi-modal structures and peaks that have not been accounted for is an area worthy of future study.

An area of future research is to repeat this analysis using a Fisher expansion to estimate the evidence associated with each peak. (This would increase the computational cost by a factor of five as we would need to calculate the likelihood in several points in the vicinity of points A and B to calculate the second derivatives of the likelihood function in the q and $\chi_{\rm eff}$ directions, but this would not be prohibitive.) This may yield additional/complementary information beyond what we can tell from comparing the likelihood at just two points.

5. ACKNOWLEDGMENTS

The authors gratefully thank Chia *et al.*, Mateu-Lucena *et al.* and Nitz *et al.* for sharing their q-$\chi_{\rm eff}$ GW151226 posterior samples. We thank Shanika Galaudage for their technical assistance. The authors also thank Horng Sheng Chia and David Keitel for helpful discussions. Finally, the authors credit the referee for their Fisher expansion idea presented in Section 4.

We gratefully acknowledge the Swinburne Supercomputing OzSTAR Facility for computational resources. All analyses (including test and failed analyses) performed for this study used 37.6K core hours on OzSTAR. This would have amounted to a carbon footprint of $\sim 2.3\,{\rm t}\ {\rm CO_2}$ (Australian Government - Department of the Environment and Energy 2021a,b). However, as OzSTAR is powered by wind energy from Iberdrola Australia, the electricity for computations produces negligible carbon waste.

This material is based upon work supported by NSF's LIGO Laboratory, a major facility fully funded by the National Science Foundation. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European

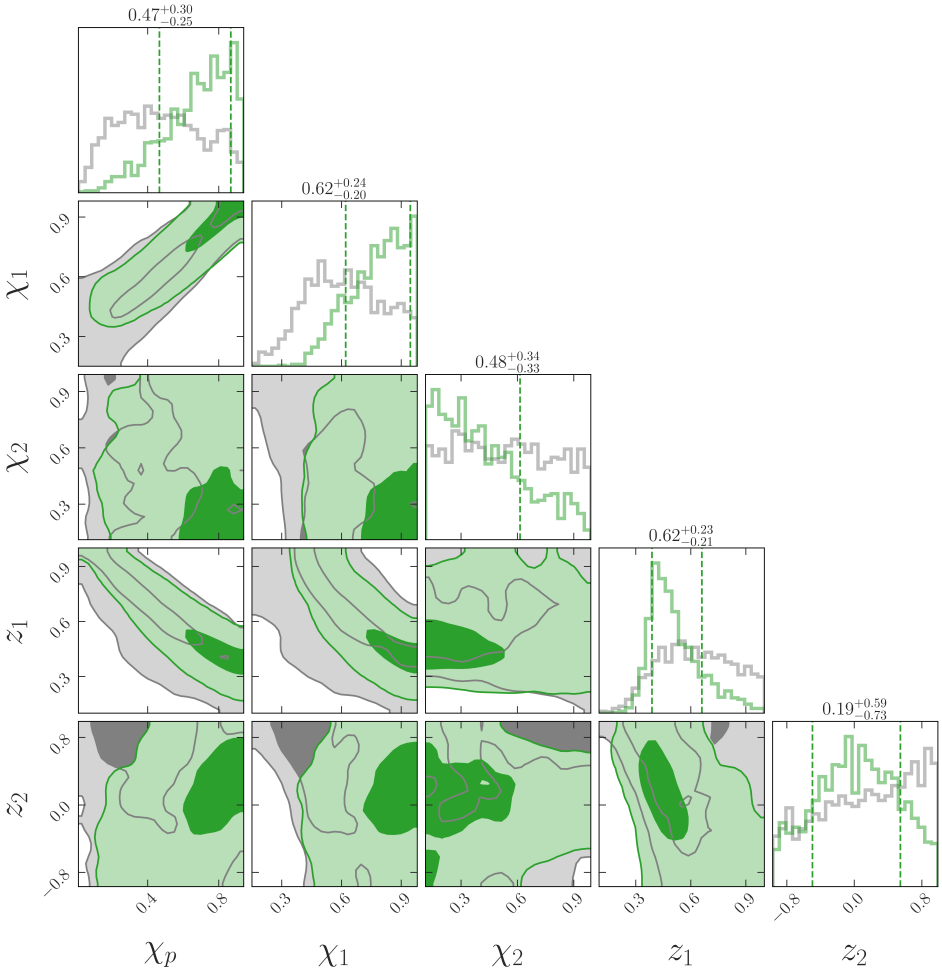

Figure 3. Posterior (green) and prior (gray) samples from the low-q (Chia *et al.*) deep follow-up analysis. The contours are drawn along the 1,3-σ regions. The vertical dashed lines in the 1D marginal distributions are plotted at the 1-σ credible regions for the posterior. The median and 1-σ credible values for posterior's marginal distributions are included above the 1D plots.

Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. The construction and operation of KAGRA are funded by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS), the National Research Foundation (NRF) and the Ministry of Science and ICT (MSIT) in Korea, Academia Sinica (AS) and the Ministry of Science and Technology (MoST) in Taiwan.

This work is supported by the Australian Research Council (ARC) Centre of Excellence CE170100004.

REFERENCES

Abbott, B. P., Abbott, R., Abbott, T. D., Abraham, S., & et al. 2020, ApJ, 892, L3, doi: 10.3847/2041-8213/ab75f5

Abbott, B. P., et al. 2016, Phys. Rev. Lett., 116, 241103

Abbott, R., Abbott, T. D., Abraham, S., Acernese, F., & et al. 2021, ApJ, 913, L7, doi: 10.3847/2041-8213/abe949

Abbott, R., et al. 2020a, Astrophys. J. Lett., 896, L44, doi: 10.3847/2041-8213/ab960f

—. 2020b, Phys. Rev. Lett., 125, 101102, doi: 10.1103/PhysRevLett.125.101102

Acernese, F., et al. 2015, Classical Quantum Gravity, 32, 024001, doi: 10.1088/0264-9381/32/2/024001

Ashton, G., Hübner, M., Lasky, P., & Talbot, C. 2019, Bilby: A User-Friendly Bayesian Inference Library, 0.4.0, Zenodo, doi: 10.5281/zenodo.2602178

Australian Government - Department of the Environment and Energy. 2021a, National Greenhouse Gas Inventory: Quarterly updates, Department of the Environment and Energy. https://www.industry.gov.au/data-and-publications/national-greenhouse-gas-inventory-quarterly-updates

—. 2021b, Carbon calculator, Department of the Environment and Energy. https://www.powershop.com.au/carbon-calculator/

Biscoveanu, S., Haster, C.-J., Vitale, S., & Davies, J. 2020, Phys. Rev. D, 102, 023008, doi: 10.1103/PhysRevD.102.023008

Chatziioannou, K., Cornish, N., Wijngaarden, M., & Littenberg, T. B. 2021, Phys. Rev. D, 103, 044013

Chatziioannou, K., Haster, C.-J., Littenberg, T. B., et al. 2019, Phys. Rev. D, 100, 104004, doi: 10.1103/PhysRevD.100.104004

Chia, H. S., Olsen, S., Roulet, J., Dai, L., & et al. 2021, arXiv e-prints, arXiv:2105.06486. https://arxiv.org/abs/2105.06486

Galaudage, S., Talbot, C., Nagar, T., et al. 2021, Astrophys. J. Lett., 921, L15

Green, R., Hoy, C., Fairhurst, S., Hannam, M., & et al. 2021, Phys. Rev. D, 103, 124023, doi: 10.1103/PhysRevD.103.124023

Littenberg, T. B., & Cornish, N. J. 2015, Phys. Rev. D, 91, 084034, doi: 10.1103/PhysRevD.91.084034

Mandel, I., & Fragos, T. 2020, ApJ, 895, L28, doi: 10.3847/2041-8213/ab8e41

Mateu-Lucena, M., Husa, S., Colleoni, M., Estellés, H., & et al. 2021, arXiv e-prints, arXiv:2105.05960. https://arxiv.org/abs/2105.05960

Nitz, A. H., & Capano, C. D. 2021, ApJ, 907, L9, doi: 10.3847/2041-8213/abccc5

Nitz, A. H., Capano, C. D., Kumar, S., et al. 2021a, ApJ, 922, 76, doi: 10.3847/1538-4357/ac1c03

Nitz, A. H., Kumar, S., Wang, Y.-F., et al. 2021b, arXiv e-prints, arXiv:2112.06878. https://arxiv.org/abs/2112.06878

Nitz, A. H., Dent, T., Davies, G. S., et al. 2020, ApJ, 891, 123, doi: 10.3847/1538-4357/ab733f

Olejak, A., & Belczynski, K. 2021, ApJ, 921, L2, doi: 10.3847/2041-8213/ac2f48

Olsen, S., Roulet, J., Chia, H. S., et al. 2021, Physical Review D, 104, doi: 10.1103/physrevd.104.083036

Olsen, S., Venumadhav, T., Mushkin, J., et al. 2022, arXiv e-prints, arXiv:2201.02252. https://arxiv.org/abs/2201.02252

Pankow, C., et al. 2018, Phys. Rev. D, 98, 084016

Pratten, G., García-Quirós, C., Colleoni, M., et al. 2021, Phys. Rev. D, 103, 104056

Qin, Y., Wang, Y.-Z., Wu, D.-H., Meynet, G., & et al. 2022, ApJ, 924, 129, doi: 10.3847/1538-4357/ac3982

Robert, C. P., & Casella, G. 2004, Monte Carlo Statistical Methods, 2nd edn. (New York: Springer)

Romero-Shaw, I. M., Talbot, C., Biscoveanu, S., et al. 2020, MNRAS, 499, 3295, doi: 10.1093/mnras/staa2850

Roulet, J., Chia, H.-S., Olsen, S., et al. 2021, Phys. Rev. D, 104, 083010

Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405, doi: 10.1063/1.1835238

Skilling, J. 2006, Bayesian Analysis, 1, 833, doi: 10.1214/06-BA127

Smith, R. J. E., Ashton, G., Vajpeyi, A., & Talbot, C. 2020, MNRAS, 498, 4492, doi: 10.1093/mnras/staa2483

Speagle, J. S. 2020, MNRAS, 493, 3132, doi: 10.1093/mnras/staa278

Talbot, C., Thrane, E., Biscoveanu, S., & **R. Smith**. 2021, Phys. Rev. Res., 3, 043049

The LIGO Scientific Collaboration, Aasi, J., Abbott, B. P., et al. 2015, Classical and Quantum Gravity, 32, 074001, doi: 10.1088/0264-9381/32/7/074001

The LIGO Scientific Collaboration, the Virgo Collaboration, the KAGRA Collaboration, et al. 2021, arXiv e-prints, arXiv:2111.03606. https://arxiv.org/abs/2111.03606

Thrane, E., & Talbot, C. 2019, PASA, 36, e010, doi: 10.1017/pasa.2019.2

Vajpeyi, A. 2021, Deep Followup Source Code, https://github.com/avivajpeyi/deep_gw_pe_followup, GitHub

APPENDIX

A. TRACKING DOWN THE SOURCE OF DISCREPANCIES

Gravitational-wave astronomers use Bayesian inference to derive posterior distributions of astrophysical parameters given data. These inferences are affected by several assumptions and data analysis choices. Key ingredients include the sampling frequency, the noise power spectral density (PSD) (Littenberg & Cornish 2015; Chatziioannou et al. 2019; Talbot et al. 2021; Biscoveanu et al. 2020), data-cleaning procedures (Pankow et al. 2018; Chatziioannou et al. 2021), the prior distribution, and the sampler (and sampler settings) employed to probe the parameter space. All these factors can affect the final posterior distributions. Thus, given two discrepant posterior distributions obtained with the same data, the source of the disagreement may not be immediately clear. However, the deep-followup method presented here can assist us in determining which ingredient(s) are responsible for the difference.

In this paper, we investigate the high-q and low-q posteriors peaks for GW151226. We conduct our deep follow-up with the ingredients utilised by the Mateu-Lucena *et al.* analysis and find slightly more support than Mateu-Lucena *et al.* for the low-q peak, but not nearly as much as Chia *et al.*. Since we observe a modest difference from Mateu-Lucena *et al.* using the same noise PSD, the same analysis software, etc., it seems plausible—by process of elimination—that the difference is attributable to the ability (or inability) of the sampler to probe the low-q peak with the uniform-in-$\chi_{\rm eff}$ prior. As part of ongoing work, we are attempting to test this hypothesis by reanalyzing GW151226 using a uniform-in-$\chi_{\rm eff}$ prior, which Olsen et al. (2021); Nitz & Capano (2021) suggest probes the parameter space of black hole spins more efficiently than the isotropic spin prior.

However, our findings could have been different; we could have obtained results consistent with Mateu-Lucena *et al.*, which would have ruled out the choice of prior as the source of the discrepancy with Chia *et al.*. It is instructive to consider what next steps we would take in this hypothetical scenario to understand better the discrepancy between the Chia *et al.* and Mateu-Lucena *et al.* analyses. First, to investigate the difference in noise PSD, we would reweight our posteriors samples to see how the results change using the noise PSD from Chia *et al.*. The weight for sample i is given by

$$w_i = \frac{\mathcal{L}(d|\theta_i, \mathrm{PSD}_{\text{Chia } et\ al.})}{\mathcal{L}(d|\theta_i, \mathrm{PSD}_{\text{Mateu-Lucena } et\ al.})}, \tag{A1}$$

where $\mathcal{L}$ is the likelihood function of the data d given astrophysical parameters θ_i, and conditioned on one of two different noise models.[11] If the reweighted samples reproduced key features from Chia *et al.*, we could attribute the difference to Chia *et al.*'s choice of noise PSD.

However, let us imagine that a different choice of noise PSD yields results broadly consistent with Mateu-Lucena *et al.*. In this scenario, the next step is to repeat the importance sampling procedure but now focusing on a different ingredient, such as the data cleaning procedure. Once again, we calculate weights, but this time, the target distribution would use the Chia *et al.* data-cleaning procedure while the proposal distribution would use the data-cleaning procedure (if any) from Mateu-Lucena *et al.*. We would continue down the list of data-analysis ingredients until either we find the culprit or run out of plausible explanations for the discrepancy. If we can find nothing that alters the Mateu-Lucena *et al.* results to reproduce the salient features from Chia *et al.*, we may be left to speculate that there is simply a bug in one or more analyses.

[11] This is an example of importance sampling; see, e.g., Robert & Casella (2004). The Mateu-Lucena *et al.* noise model serves as the proposal distribution, while the Chia *et al.* noise model is what is known as the target distribution.