Black Hole Coalescence: Observation and Model Validation

Jamee Elder

Published in *Working Toward Solutions in Fluid Dynamics and Astrophysics: What the Equations Don't Say.* Lydia Patton and Erik Curiel (ed.), Springer Briefs in History of Science and Technology. (2023) http://doi.org/10.1007/978-3-031-25686-8$_5$

Abstract This paper will discuss the recent LIGO-Virgo observations of gravitational waves and the binary black hole mergers that produce them. These observations rely on having prior knowledge of the dynamical behaviour of binary black hole systems, as governed by the Einstein Field Equations (EFEs). However, we currently lack any exact, analytic solutions to the EFEs describing such systems. In the absence of such solutions, a range of modelling approaches are used to mediate between the dynamical equations and the experimental data. Models based on post-Newtonian approximation, the effective one-body formalism, and numerical relativity simulations (and combinations of these) bridge the gap between theory and observations and make the LIGO-Virgo experiments possible. In particular, this paper will consider how such models are validated as accurate descriptions of real-world binary black hole mergers (and the resulting gravitational waves) in the face of an epistemic circularity problem: the validity of these models must be assumed to justify claims about gravitational wave sources, but this validity can only be established based on these same observations.

Jamee Elder
Black Hole Initiative, Harvard University, 20 Garden Street, 2nd Floor, Cambridge, MA 02138
e-mail: jelder@fas.harvard.edu

1 Introduction

On September 14, 2015 the "Laser Interferometer Gravitational-wave Observatory", (hereafter "LIGO"), comprising interferometers in Hanford, WA, and Livingston, LA, detected gravitational waves for the first time. This event, dubbed GW150914 (based on the date of detection), marked the beginning of a new epoch for the field of gravitational-wave astrophysics. The two Advanced LIGO interferometers, together with the Advanced Virgo observatory in Italy, form a global network of gravitational-wave interferometers capable of observing gravitational waves and (through these) the astrophysical sources that generate them.

The detection of gravitational waves has been hailed as a revolution for astrophysics. This is because the recent advent of gravitational-wave astrophysics gives us a new "window" through which to observe the universe, enabling us to observe events and objects that were previously invisible to us, such as collisions between two black holes.

Black holes and gravitational waves are gravitational phenomena predicted by Einstein's theory of general relativity. On the standard geometric interpretation of this theory, gravity is viewed as a manifestation of the curvature of spacetime. This curvature is encoded in the (Lorentzian) metric, $g_{\mu\nu}$, and the response of the metric to energy and momentum is governed by the Einstein field equations,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu} \tag{1}$$

The right hand side of this equation concerns the local energy and momentum within the spacetime, as expressed by the energy-momentum tensor (alternatively called the stress-energy tensor) $T_{\mu\nu}$. The left hand side concerns the curvature of spacetime. Here, $G_{\mu\nu}$ is the Einstein tensor, defined in terms of the Ricci tensor, $R_{\mu\nu}$, the metric, $g_{\mu\nu}$, and the Ricci scalar, R.

(1) can be viewed as an equation relating 4×4 matrices. The components of these matrices yield a total of sixteen non-linear partial differential equations for the metric $g_{\mu\nu}$. However, the symmetry of the metric tensor means that this reduces to only 10 independent equations.[1]

Unfortunately, exact solutions to the Einstein field equations are difficult to come by. While many such solutions now exist, they often describe very simple, idealised physical scenarios. Much of the usefulness of such solutions is in providing a starting point for approximation schemes, used to model more physically realistic systems (Kennefick 2007, 41). Such approximation schemes can even be used even when there are no exact solutions to the Einstein equations to use as a foundation. Instead, empirically successful descriptions provided by previous theories are used. For example, Einstein's famous calculation of Mercury's perihelion advance started from the Newtonian solution then added relativistic corrections in powers of $(v/c)^2$ (42–3). This "post-Newtonian" approximation scheme remains in use today. Indeed, compact binaries are well-modelled by post-Newtonian approximation for the early

[1] See Kennefick (2007, 46–7) for a very clear explanation of this point.

inspiral, where the characteristic velocities and the gravitational field strength remain small (in relativistic terms). Beyond the early inspiral the compact binary merger must be described in the "dynamical strong field regime", where velocities are high and the gravitational field is strong. Here, the lack of full analytic solutions becomes more problematic, as I will discuss in the remainder of this paper.

Despite the lack of exact analytic solutions, recent work has led to a range of approaches to modelling the late stages of the merger, notably including the effective one-body formalism and numerical relativity simulations. This paper concerns the validation of such models in the absence of exact analytic solutions.[2]

Interwoven with questions about the validity of these models are questions about the status of the LIGO-Virgo observations. These observations rely heavily on the use of such models for both gravitational wave detection (through matched filtering, see section 2.1) and inferences about the compact binary systems that produced them (through parameter estimation, see section 2.2). Indeed, advances in numerical relativity and other modelling approaches were crucial to the success of LIGO-Virgo experiments.

In this paper, I argue that modelling plays an essential role in connecting high-level theory, embodied in the Einstein field equations, with the LIGO-Virgo data. The models used in template-based searches for gravitational waves (section 2.1) and in parameter estimation (section 2.2) incorporate insights from a range of modelling approaches, allowing them to bridge the gap between theory and world. Thus, alongside technological advances, the modelling approaches discussed play a vital role in gaining empirical access to both gravitational waves and the binary black hole mergers that produce them.

However, validating these models—both as representative of the predictions of general relativity, and as accurate descriptions of the target systems—presents a challenge. This is because the model-dependence (or theory-ladenness) of the LIGO-Virgo observations creates a justificatory circularity: the accuracy of the models must be established using the LIGO-Virgo observations, but these observations *assume* the accuracy of the models. Additional features of the epistemic situation, such as the lack of independent access to these systems and the new physical regimes being probed, render this circularity difficult to break. However, the LIGO-Virgo do perform "tests of general relativity" that test specific assumptions in the experimental methodology. While this doesn't exactly break the circularity, it goes some way towards rendering it benign, or even virtuous.

Overall, this paper shows how the methodology of the LIGO-Virgo experiments is intimately bound up with models of binary black hole mergers and the gravitational waves that they produce. The success of these experiments rests on confidence in these models, which bridge the gap between theory and phenomena. The flip side of this is that much of the interesting work in validating the LIGO-Virgo results lies in validating the models themselves with respect to both the equations of general relativity and the physical systems being observed.

---

[2] There are distinctions between solutions that are "exact", "elementary", "algebraic" etc., which I do not delve into here. For discussion of these issues, see Fillion and Bangu (2015).

2 Modelling and Observing Binary Black Hole Mergers

The LIGO and Virgo interferometers produce gravitational-wave strain data as a time series, sampled more than 16,000 times per second. This data is very noisy, with terrestrial noise sources disguising even strong signals like GW150914. This means that it is not possible to see a signal in the unprocessed data. Instead, significant data-processing is needed both to search for gravitational wave signals in the data and to infer the properties of the systems that produced them.

2.1 Template-based Searches for Gravitational Waves

The LIGO-Virgo Collaboration has multiple independent search pipelines that are used to find gravitational waves in their data. This includes unmodelled "burst" searches using, for example, the coherent Waveburst ("cWB") algorithm. However, the most effective search methods are modelled searches, which rely on assumptions about the kinds of signals being sought. In this section, I'll mostly be concerned with reviewing the most important feature of these modelled searches: matched filtering. For more comprehensive descriptions of the data analysis techniques used, see, for example, Abbott et al. (2016a) and Abbott et al. (2020).

Matched filtering is a signal-processing technique that involves correlating a known signal, or "template", with an unknown signal, in order to detect the presence of the template within the unknown signal.[3] This technique allows for the extraction of a pre-determined signal from much larger noise. An optimal filtering maximises the "signal-to-noise ratio" (hereafter "SNR"), which (roughly) measures the ratio of the (amplitude of the) gravitational wave signal compared to the noise.

In order to extract a particular gravitational wave signal, the data must be searched using a closely matching template. Since it isn't known in advance which (if any) signal is present, the data must be searched for a range of signals, corresponding to the full range of gravitational waves that might be present. What is needed, then, is a library of template models corresponding to the range of possible signals. Templates are arrived at by considering the range of systems that we believe could produce measurable gravitational waves and determining the predictions of general relativity for the behaviour of such systems, including gravitational wave emission. Of course, this is no trivial task (to put it mildly).

Modelled searches using matched filtering appear to be a clear case of theory-laden observation. What can be "seen" in the data is determined by the models that are used to in the observation process. The theory- or model-ladenness of these observations naturally leads to concerns about whether the LIGO-Virgo Collaboration are observing all of and only the genuine gravitational wave signals in the data, and how accurate these observations are (i.e., how well the recovered signal reflects

[3] For a detailed introduction to matched-filtering in gravitational-wave astrophysics, see Maggiore (2008, section 7.3).

the gravitational waves passing through the interferometer). I discuss the potential pitfalls of the model-based searches in section 3.1.

2.2 Bayesian Parameter Estimation

Having detected a gravitational wave signal, it is then possible to make inferences about the properties of the source system that produced it. (a compact binary merger, usually a binary black hole merger). This involves estimating the values of a range of parameters characterising the system. Through parameter estimation, the detection of gravitational waves doubles as an observation of a compact binary merger (at least on permissive uses of the term "observation", see e.g., Shapere (1982) and Elder (2020, ch.2)). Indeed, GW150914 was called the *first* observation of a binary black hole merger.

There are 15 key parameters that determine the received signal.[4] These are:

- Luminosity distance to coalescence event (1 parameter)
- Angular location of event in sky (2 parameters)
- Orientation of the orbital plane relative to line of sight (2 parameters)
- Time of arrival (1 parameter)
- Orbital phase at time t (1 parameter)
- Masses of the component compact objects (2 parameters)
- Spin components of compact objects (6 parameters, 3 per object)

Of these, only the last two list items are "intrinsic parameters"; they concern the properties of the compact binary itself rather than its relationship to the interferometers used to observe it.

Parameter estimation is performed within a Bayesian framework. The basic idea is to calculate posterior probability distributions for the parameters describing the source system, using Bayes' theorem. To do this, we need the following: a model M that takes a set of system parameters and predicts the resulting signal; background or prior information I; and some data d. Bayes' theorem can then be written as:

$$p(\theta|d, M, I) = p(\theta|M, I)\frac{p(d|\theta, M, I)}{p(d|M, I)}, \tag{2}$$

where $\theta = \{\theta_1, \ldots \theta_N\}$ is a collection of parameters. This equation tells us how to calculate the posterior distributions for θ, given some fixed modelling approach M and other analysis assumptions I.[5]

On this approach, prior probability distributions must be specified for all 15 parameters. For some parameters this can be done based on symmetries of the parameter

[4] This assumes that the eccentricity of the orbit can be neglected, since the emission of gravitational radiation is expected to circularise the orbit by the time the emitted gravitational waves enter the bandwidth of the detector (Peters 1964).

[5] For a details, see Abbott et al. (2020, 36).

space. For example, for a redshift $z \ll 1$ in a Friedmann-Lemaître-Robertson-Walker cosmological model, equal numbers of coalescence events are expected to occur in equal co-moving volumes (Abbott et al. 2020, 37). For other parameters, the LIGO-Virgo approach is to choose simple priors such that the posteriors can be easily interpreted.

In addition to the parameters characterising properties of the binary, this analysis also includes $O(10)$ extra parameters per detector to model calibration uncertainties. Thus for a three-detector analysis around 45 parameters are being sampled. In order to efficiently sample this high-dimensional parameter space, the LIGO-Virgo Collaboration developed LALInference, a stochastic sampling library that uses two different algorithms to perform parameter estimation. The first is a parallel tempering Markov chain Monte Carlo algorithm, and the second is a nested sampling algorithm.[6] The end products of the LALInference analyses are posterior samples for all of the parameters.

With parameter estimation, we appear to have another clear case of theory- or model-laden observation. In this case (unlike the case of template-based searches for gravitational waves) there is no alternative to model-based inference. All such "observations" of compact binary mergers are based on what the theory has to say about the behaviour of such systems.

2.3 Modelling Black Hole Coalescence

Overall, we have seen that both the detection of gravitational waves, through matched filtering, and the observation of compact binaries, through Bayesian parameter estimation techniques, rely on having accurate models of these phenomena.

It was known in advance that the best candidates for detection by the LIGO and Virgo interferometers were gravitational waves produced by compact binary mergers. These include binary systems containing two black holes, two neutron stars, or one of each. It was thus important to determine what general relativity predicted about the behavior of such systems, including their emission of gravitational waves. However, the general relativistic two-body problem is notoriously hard and we do not have exact analytic solutions for spacetimes containing merging compact binaries.[7] Thus we must rely on approximations, models, and simulations of these spacetimes in order to predict the form that any resultant gravitational waves will take. These include (1) post-Newtonian approximations, (2) models generated using black hole perturbation theory, (3) numerical relativity simulations, and (4) models based on the effective one-body approach.

(1) Post-Newtonian ("PN") theory is a well-established method for modelling systems where motions are slow compared to the speed of light and the gravitational

[6] For details about LALInference analyses, see Veitch et al. (2015).

[7] See, for example, Kennefick (2007, Ch.8) and Havas (1989, 1993) for the history of this problem.

field is weak.[8] PN models take Newtonian solutions as the starting point, then add general relativistic corrections to the equations of motion and the radiation field order by order in powers of v^2/c^2 (where v is the characteristic orbital velocity of the compact binary). Several different approaches to PN theory have been explored with the result that the PN equations of motion for two spinless black holes are now known up to 4PN order.[9] The gravitational radiation can be extracted through the multipolar post-Minkowskian wave generation formalism (Blanchet and Damour 1986; Blanchet 1987, 1998) or through the "direct integration of the relaxed Einstein equation" approach (Will and Wiseman 1996; Pati and Will 2000). PN theory is thought to model a compact binary and the resulting radiation well for the early inspiral. However, it fails to accurately model the late inspiral onwards, as the separation between the compact objects decreases and the orbital velocity becomes large.

(2) Black hole perturbation theory ("BHP") is a useful tool for modelling compact binaries with extreme mass ratios, where $m_2/m_1 \ll 1$.[10] For such mass ratios, the motion of the smaller object can be modelled by introducing perturbations in the background metric of the larger object. Leading order gravitational wave emission (related to the dissipative component of the self-force) is described by the Regge-Wheeler-Zerilli equations for a Schwarzschild background, and by the Teukolsky equation for a Kerr background.

(3) Numerical Relativity ("NR") waveform models are models produced through simulations on supercomputers. These simulations solve the exact Einstein equations numerically, enabling the calculation of the form of the gravitational waveform that would emanate from a binary merger with specific parameter values. After several decades of work to overcome a range of formidable technical challenges (e.g., formulations, gauge conditions, stable evolutions, black hole excision, boundary conditions, and wave extraction), numerical relativity saw a major breakthrough in 2005, with the first stable simulations of the final orbits, plunge, merger, and ringdown of a binary black hole merger. These original successes are reported in Pretorius (2005), Baker et al. (2006), and Campanelli et al. (2006).[11] Numerical relativity simulations are thought to be extremely accurate, even for the plunge and merger phases.[12] However, they are extremely computationally expensive, so it is not feasible to generate a 250,000-waveform template library from NR simulations alone.

(4) The effective one-body ("EOB") models of binary black holes are quasi-analytic models produced using the effective one-body formalism. This approach, developed by Alessandra Buonanno and Thibault Damour in the late 1990s, recasts the two-body problem as an effective field theory for a single particle (see Buonanno and Damour (1999) and Buonanno and Damour (2000)). The basic idea behind

[8] For a detailed discussion of the post-Newtonian approach, see Maggiore (2008, Ch. 5).

[9] Since the nPN order refers to inclusions of terms $O(1/c^{2n})$, 4PN results include terms $O(1/c^8)$

[10] Here, by convention, $m_1 > m_2$.

[11] For detailed reviews of the history of numerical relativity simulations, see e.g., Holst et al. (2016) and Sperhake (2015).

[12] I discuss reasons for this confidence in Section 3.3.

this approach is to transform the conservative dynamics of two compact objects (masses m_1 and m_2, spins S_1 and S_2) into the dynamics of an effective particle ("mass" $\mu = m_1 m_2/(m_1 + m_2)$, and "spin" S^*) moving in a deformed Kerr metric ($M = m_1 + m_2$, S_{Kerr}). The EOB models build on the PN approach (by taking high-order PN results as input) in order to produce accurate analytic results for the entire process.

Each of these modelling approaches have different (but overlapping) domains of application. As mentioned above, PN models are valid for the early inspiral, but become increasingly inaccurate as the compactness parameter increases.[13] For BHP theory, the limiting factor is the mass ratio; the approximation becomes inaccurate as the mass ratio increases.[14] The domains of validity for PN and BHP do not have sharp cutoffs. Rather, the cutoff used depends on the acceptable level of error for a given calculation (Le Tiec 2014, 4). In principle, the domain of applicability for NR spans the whole parameter space. However, in practice, NR simulations are limited by available computing resources, with simulations involving large separations r or large mass ratios m_2/m_1 involving long and therefore costly calculations. In contrast, the EOB approach is supposed to be valid throughout the parameter space without being computationally expensive.

The models derived from each of these approaches are related in a range of ways. For example, EOB models and NR simulations take PN approximations as input; EOB models are calibrated ("tuned") against NR simulations; NR simulations are tested against PN approximations; and (calibrated) EOB models are tested against NR simulations.

In fact, the waveform models that are actually used by the LIGO-Virgo collaboration combine PN, EOB, and NR results. Two important families of hybrid models are the "EOBNR" and "IMRPhenom" models. (Abbott et al. 2016c, 5). The EOBNR models are EOB models whose (unknown) higher order PN terms have been tuned to NR results in order to improve the Hamiltonian. They are thus hybrids in the sense that they are EOB models that are tuned to incorporate insights from both PN and NR results. Essentially, the EOB models have free parameters that can be modelled by comparison to NR simulations, producing the EOBNR models. The IMRPhenom models are hybrids in a more direct sense, in that they fit together inspiral, merger, and ringdown models derived from the PN, EOB, and NR approaches. These models are constructed by extending frequency-domain PN results then creating hybrids of PN and EOB models with NR waveforms. In particular, those used for parameter estimation in Abbott et al. (2016c, 5) are made by fitting untuned EOB waveforms and NR simulations.

---

[13] Here the compactness parameter is defined as as the ratio M/r where $0 < M/r \lesssim 1$, and $M = m_1 + m_2$.

[14] The mass ratio is defined as m_2/m_1 where, by convention, m_1 is always the larger mass and thus $0 < m_2/m_1 \lesssim 1$.

2.4 Model-dependent Observation

The LIGO-Virgo Collaboration described GW150914 as the first "direct detection" of gravitational waves and the first "direct observation" of a binary black hole merger.[15] However, these descriptors have the potential to obscure the fact that these observations are also *indirect* in the sense that they are mediated by models, such as those in the EOBNR and IMRPhenom modelling families. Generating models that act as "mediating instruments" (Morgan and Morrison 1999) between theory and world was a vital contribution to the theoretical and technological advances that made the LIGO-Virgo experiments possible. As we have seen, "observations" in gravitational-wave astrophysics depend on the availability of these models.

The EOBNR and IMRPhenom models are a hodgepodge of other modelling approaches. Far from being a problem, this is what enables them to successfully bridge the gap between theory and data in the context of the LIGO-Virgo experiments. On the one hand, post-Newtonian approximation has a long history of bringing the theory of general relativity into empirical contact with observed systems, including the good agreement with observations of the Hulse-Taylor pulsars to high PN order. On the other hand, numerical relativity simulations provide the closest contact with the full Einstein field equations, since these simulations are as close as we have to exact solutions for the systems of interest. Thus determining what the theory of relativity tells us about the world is not a matter of analytically solving a system of differential equations. Rather, it involves heterogeneous modelling in a middle ground between theory and data, constrained by interfaces with existing theoretical and empirical results.

The EOBNR and IMRPhenom modelling frameworks are both thought to accurately describe gravitational wave signals for systems like the source of GW150914.[16] However, justifying confidence in these models, especially for the late stages of a binary black hole merger, is far from straightforward. This is largely due to the lack of either analytic solutions or previous empirical constraints in this regime. In the remainder of this paper, I critically examine the justification for confidence in models based on these approaches—both as faithful predictions derived from the Einstein field equations, and as accurate descriptions of real physical systems.

3 Model Validation and Circularity

When we turn to considering the justification for the models employed in the LIGO-Virgo experiments, the theory-ladenness or model-dependence of the observations

[15] In other work, including Elder (2020) and Elder (2021b), I provide an account of what is meant by these descriptors, drawing connections to recent work in the philosophy of measurement (e.g., Tal (2012, 2013) and Parker (2017)).

[16] For the purposes of this paper, I will mainly focus on the models used for this initial detection. However, modelling these systems is an active area of research, with new iterations of these modelling approaches incorporating more physical effects.

becomes a source of potential concern. As with other instances of theory-ladenness, the general concern here is that any inaccuracy in the models could be systematically biasing observations and hence conclusions about the world. In sections 3.1 and 3.2, I examine how this general concern plays out for the observation of gravitational waves and binary black hole mergers, respectively.[17]

3.1 Worry 1: Theory-laden Observation of Gravitational Waves

The use of matched filtering in the observation of gravitational waves appears to be a clear case of theory-laden observation, since this technique relies on advance knowledge of what the form of the gravitational wave signal will be. Traditional concerns about the theory-ladenness of observation are to do with more traditional notions of observation—the observations made by individual people—and how these are influenced by the conceptual schemes or "paradigms" of those individuals. Here the concern applies to a broader notion of observation that encompasses the outputs of experimental procedures. Franklin (2015) calls this the "theory-ladenness of experiment". The observation of gravitational waves via matched filtering offers a particularly clear case of such theory-ladenness. However, such theory-ladenness also appears to be a generic feature of experiment. With this in mind, it might seem that the reliance on models in making observations can hardly be taken to be a special concern for LIGO-Virgo, but rather a general concern about all experiment-based observations. It is beyond the scope of this paper to address broader concerns about the theory-ladenness of either observations or experiments in general. (Indeed, I am inclined to doubt that such a general treatment would offer much insight into the particular concerns I address here.) Instead, my focus will be on the details of the present case and three potential problems that could arise due to the model-dependence of matched filtering: (1) missed signals, (2) false signals, and (3) sub-optimal signal extraction.

First, we might worry about gravitational wave signals being missed due to a lack of any corresponding templates. Such signals might lack corresponding templates because they originate from different types of sources than those represented in the template library, or because they exhibit behavior that deviates from the predictions of those templates (e.g., due to deviations from general relativity in the strong field regime). In either case, such signals would be a very interesting discovery and failure to detect them could be a major scientific loss. Missing such signals could also lead to a biased sampling of the kinds of events that produce gravitational waves detectable by LIGO and Virgo, leading to inaccurate conclusions about the populations of such events. However, there are some considerations that cut against this worry. For one thing, it is worth noting that matched filtering using general relativity-based

[17] My focus here is on binary black hole mergers, in part because I am focused on the case of GW150914 specifically, and in part because the case is arguably slightly different for binary neutron star mergers. I discuss how the epistemic situation is changed for "multi-messenger" sources like GW170817 in Elder (2020, Ch.4) and Elder (2021a).

templates should still be able to detect gravitational waves from a binary black hole merger unless the deviations from general relativistic descriptions are quite large (Yunes and Pretorius 2009, 1–2). It is also worth remembering that the data does not just disappear. This means data can be searched later. Indeed, work done by Abedi, Dykaar, and Afshordi (2017) to detect "echoes" in the LIGO-Virgo data provides an example of such a search. Finally, and perhaps most importantly, the LIGO-Virgo Collaboration also have unmodelled ("burst") search pipelines, which are able to detect gravitational wave signals with minimal assumptions about the form of the signal (Abbott et al. 2016b). These are less effective than the matched filtering search pipelines, reporting lower SNR and correspondingly lower statistical significance for the same gravitational wave signal. Furthermore, there have not been any confirmed detections from the unmodelled searches that were not also detected with modelled searches.[18] This makes it hard to determine the extent to which the unmodelled search pipelines help alleviate the concern that signals are being missed. However, the fact that these unmodelled searches are sensitive enough to detect the same events as the modelled searches (even if with lower SNR) at least demonstrates a reduced reliance on matched filtering for successful detection of gravitational waves.

Second, we might worry about whether the matched filtering procedure ever recovers false signals. After all, it seems that if we search for enough templates, through enough data, it is only a matter of time before the noise just happens to correlate with a template with a reported SNR above the designated threshold. If such false signals were being interpreted as genuine gravitational wave signals, there would indeed be something problematic about the data analysis procedures. However, this is the kind of problem that the LIGO methodology is best designed to avoid. Indeed, BICEP2's retraction was fresh in the minds of the LIGO-Virgo scientists, leading to an even higher level of caution about any detection claims (Collins 2017, 72).[19] The standards for detection implemented by the LIGO-Virgo Collaboration prioritize the avoidance of such false-positives, reflecting a high level of caution about any discovery claims. For example, a high SNR in one detector is not sufficient for an event to be classified as a detection. The LIGO-Virgo detection procedure also requires that there be coincident events in at least two detectors. For the modelled searches, these must be coincident triggers with the same template. Such a coincidence is unlikely to be mimicked by noise. Indeed, the LIGO-Virgo methods for analyzing the significance of the detection, using time slides, are designed to quantify the likelihood that this coincidence could have been generated by chance due to detector noise. For GW150914, the false alarm rate was found to be less than one event per 203,000 years. In other words, it is thought to be extremely unlikely that the events detected were products of processes that contribute to the noise background.

---

[18] O3 has recently reported one burst candidate, given the preliminary name "S200114f", but this candidate has not yet been confirmed.

[19] "BICEP" stands for "Background Imaging of Cosmic Extragalactic Polarization". They initially reported the observation of signatures of primordial gravitational waves in 2014 but were forced to retract this claim, instead attributing the observation to cosmic dust (Collins 2017, 72).

Third, there is a kind of intermediate worry to be considered: the matched filtering procedure could extract a signal, but do so in such a way that the measured signal does not accurately represent the gravitational waves. This doesn't present a problem when it comes to detection, except perhaps in cases where the SNR is near the threshold value and imperfect extraction could be the difference between an event being classified as a detection or not. However, imperfect extraction of a signal does have implications for the inferences that we wish to make about it. For example, underestimating the amplitude of the wave could lead us to false conclusions about the distance to the compact binary, or the inclination of the orbital plane relative to the line of sight. More generally, the possibility of imperfectly filtered signals leads to worries about parameter estimation, and the further inferences we make based on biased "observations" of the compact binary systems. I turn to these issues in the following section.

3.2 Worry 2: Theory-laden Observation of Binary Black Hole Mergers

The theory-ladenness, or model-dependence, of the LIGO-Virgo methodology becomes more problematic once we consider how the interferometers are used to "observe" compact binary coalescences. In particular, there is an apparently problematic circularity in the validation of these models, with epistemic implications for parameter estimation, theory testing, and further inferences about the population of such events. In this section, I will try to answer two important questions: First, what justification do we have for thinking that the models used by LIGO are good ones?; And second, what are the potential consequences of inaccuracies in these models?

There are two main sources of error to consider here: error due to modelling practices (PN, EOB, NR, etc.) and error due to the underlying theory (GR). Yunes and Pretorius (2009) classify these as "modelling bias" and "fundamental bias," respectively. More generally, we can characterise these sources of bias as being due to the inaccuracy of the model with respect to the underlying theory, in the case of modelling bias, and the inaccuracy of the theory with respect to the target system, in the case of fundamental bias. Some related distinctions are made elsewhere, with respect to the validity of models and simulations.

Verification and validation are prominent notions in the philosophical literature concerning computer simulations. These notions are derived from the corresponding scientific literature, since practitioners—especially in engineering and climate science contexts—have produced a large body of work dealing with these notions (Winsberg 2019). Verification and validation are two categories of processes for checking the accuracy or reliability of a simulation.

Verification is the process of determining whether the simulation output is a good approximation to the actual solution of the equations of the relevant model. This can be broken up into two parts: checking that the computer code successfully implements the intended algorithm ("code verification") and checking that the outputs of this algorithm approximate the (analytic) solutions to the equations being solved via

simulation ("solution verification") (Winsberg 2019). The latter, solution verification, sometimes takes the form of comparing simulation outputs to known analytic solutions that act as "benchmarks." In cases where no known benchmarks are on offer (as in the case of compact binary mergers) different approaches are needed. For example, it can be checked whether successive simulations converge on a stable solution as the resolution of the simulation is increased. In other words, this checks whether the result of the simulation stays approximately the same as the physical situation is modelled in an increasingly fine-grained manner.

Validation is the process of checking whether the given model is a good enough representation of the target system. Thus validation procedures require empirical tests of whether the simulation outputs provide a good enough description of the behaviour of the physical system of interest.

Verification concerns modelling bias. In the LIGO-Virgo case, verification procedures would be those processes designed to ensure that the outputs of numerical relativity simulations were true solutions to the Einstein field equations. Theoretical bias of the kind described by Yunes and Pretorius (2009) comes into play in the context of validation, since it is here that we are trying to establish whether the outputs of a simulation are accurate with respect to the physical world. Validation procedures must contend with any theoretical bias introduced into the broader testing procedure due the reliance on models that assume the accuracy of general relativity.

Some related concepts are those of internal and external validity.[20] The classic definitions of these notions are from Campbell and Staley (1966) and Cook (1979). According to these authors, *internal validity* concerns the validity of an inference to there being some kind of causal relationship that is captured by the experimental data. *External validity* concerns the validity of inferences about the generalisability of this causal relationship. In other words, an experiment is internally valid if we can gain genuine knowledge about the experimental system ("object") based on the data and it is externally valid if this knowledge allows us to make justified inferences about a system or class of systems that we want to learn about ("target").

Internal validity and external validity are related to modelling bias and fundamental theoretical bias through the roles models play in making inferences about both the object and target systems. Experiments involve a hierarchy of models that go between the theory and the experimental data (Suppes 1962). Morgan and Morrison (1999) discuss the many roles that models play in mediating between theory and data, arguing that they can be understood as technologies for learning—both about the theory and about the world. Thus, for example, we use template waveforms as a tool for learning about the signal passing through the object system (through matched-filtering, as discussed above). We can also use these models for learning about the binary black hole system that produced the signal using sophisticated Bayesian inference software. Establishing internal and external validity in this case involves a series of model-based inferences for which both modelling bias and fundamental theoretical bias need to be taken into account.

[20] Other notions of validity can be added here. For example, see Cronbach and Meehl (1955) for a discussion of the notion of "construct validity".

3.3 Modelling Bias

Even if we can assume that general relativity provides an accurate description of systems like binary black hole mergers (an assumption I discuss below), there may still be reason to be concerned that the models used in detecting and reasoning about these systems are inaccurate. Given that we lack full analytic solutions against which to check our models, it is worth considering how these models are justified. In particular, how could we have enough confidence in these models that we were willing to rely on them in gravitational-wave astrophysics?

3.3.1 Validating EOBNR

As discussed above, PN, BHP, and NR have limited domains of applicability (though for different reasons). In contrast, the EOB formalism is supposed to offer a single framework that generates valid models throughout the parameter space. However, untuned EOB models are inadequate; it is EOBNR models, which have been tuned to NR simulations, that are needed for gravitational wave detection and compact binary parameter estimation.[21]

The EOBNR models are supposed to incorporate the results of the PN, EOB, NR and BHP approaches. Thus while the EOB models have some original physical motivation, confidence in the validity of these models is largely derived from confidence in other modelling approaches that are considered more secure. In particular, our confidence in these models in the strong dynamical regime (where PN models are no longer valid) is derived from our confidence in NR simulations.

NR simulations are believed to be extremely accurate. Sometimes these simulations are even referred to as "solutions" of the full Einstein field equations. For example, Abbott et al. (2016c, 3), states:

> As the BHs get closer to each other and their velocities increase, the accuracy of the PN expansion degrades, and eventually the full solution of Einstein's equations is needed to accurately describe the binary evolution. This is accomplished using numerical relativity (NR) which, after the initial breakthrough, has been improved continuously to achieve the sophistication of modelling needed for our purposes.

What this means is that NR simulations are thought to be well verified with respect to general relativity. As with other simulations, this involves both code verification and solution verification.

However, solution verification is challenging in this context due to the lack of any full analytic solutions to use as benchmarks. Indeed, NR simulations are needed to act as benchmarks for other modelling approaches. To some degree, consistency with other modelling approaches—within their domains of applicability—can play a similar role. However, what is really needed is a way to verify NR simulations in the

[21] For the purposes of this paper I will not explicitly discuss the IMRPhenom models. However, given that these hybrid models use the same ingredients as EOBNR models (PN, EOB, and NR), most if not all of what I say about EOBNR should also apply to IMRPhenom.

strong dynamical regime, where PN results are no longer adequate. Two important methods are used: convergence studies, and code comparison studies.

Convergence studies involve successive simulations of the same system with greater and greater resolution.[22] This is a valuable tool for establishing the accuracy of a simulation—in particular, for establishing that the result does not depend on the limited resolution used. If a study can show convergence to a stable solution, it can be reasonably inferred that this solution would remain valid if the resolution were increased. In other words, convergence studies are used to demonstrate that the results of a simulation are not an artifact of course-graining. In one of the breakthrough papers, Campanelli et al. (2006) perform such a study, demonstrating a high level of stability in the solution as the resolution is increased. Since NR simulations are extremely computationally expensive, it is important to avoid wasting computational resources when the gain in accuracy is insignificant. Developing an understanding of the relationships between resolution, convergence, and hence accuracy is thus important for practical reasons as well as for verification purposes.

Solution verification is also achieved by comparing the results from simulations conducted using different codes. It is worth noting that some of these codes employ very different methods. For example, the three breakthrough simulations of 2005 used two different approaches for dealing with the interiors of the black holes and the attendant singularities. Pretorius used a process called *excision*, according to which the black hole interior is simply not involved (and it need not be, since the region inside the event horizon cannot affect the region outside of it). The other two simulations used what has come to be called the *moving puncture* approach, according to which singularities are allowed to develop in the interior region, but are rendered sufficiently benign by an appropriate choice of gauge. The first code comparison was performed on codes used by each of these three groups: the LazEv code developed at Brownsville and Rochester Institute of Technology, the Hahndol code developed at NASA's Goddard Space Flight Center (GSFC), and Pretorius' original code. This study showed "exceptional agreement for the final burst of radiation, with some differences attributable to small spins on the black holes in one case" (Baker et al. 2007, S25).

Since then, more extensive comparisons have been performed. One important example is the Samurai project (Hannam et al. 2009) described below:

> For the Samurai project, comparisons were made between the SpEC ["Spectral Einstein Code"] code developed by the SXS ["Simulating eXtreme Spacetimes"] collaboration, the Hahndol code, the MayaKranc code developed at Penn State/Georgia Tech, the CCATI code developed at the Albert Einstein Institute, and the BAM ["Bi-functional Adaptive Mesh"] code developed at the University of Jena. One of the major differences between the Samurai project and [the earlier comparison] was the use of simulated LIGO noise data to determine if the differences between the waveforms generated by the various codes is, in practice, detectable.[23] (Duez and Zlochower 2018, 19)

---

[22] In general, convergence studies need not always involve increasing the resolution. For example, for Monte Carlo simulation convergence studies are performed for increasing sample sizes.

[23] For more on these codes, see Duez and Zlochower (2018), especially section II, and the references therein).

Additional comparisons were conducted in the process of confirming the status of GW150914. The SXS Collaboration, which uses the SpeC code, and the Rochester Institute of Technology (RIT) group, which uses LazEv, compared the results of their attempts to model the source of GW150914 in Lovelace et al. (2016). The two codes used different initial data generation techniques, evolution techniques, and waveform extraction techniques, and shared no common routines. Despite these methodological differences, they found that the dominant modes produced by the two codes agreed to better than 99.9% (Duez and Zlochower 2018, 20). The level of agreement in comparisons such as this, combined with the simulators' confidence in their own codes, is the main source of confidence in the results of NR simulations of compact binary mergers.[24]

NR simulations of GW150914-like events are generally considered to be well-verified for the purposes of both detection and parameter estimation, despite the lack of analytic benchmarks. Indeed, the agreement between waveforms derived from independent simulations is such that the waveforms are indistinguishable below an SNR (Hannam et al. 2009). However, they are too computationally expensive to generate all the templates needed for matched-filter-based detection. Instead, the NR simulations are used as a kind of benchmark for EOB models, which require minimal computational resources. The EOB models contain a number of free parameters that can be tuned to NR results. The entire parameter space can then be filled by interpolation between known NR results. While the EOB models do have some independent physical motivation, much of the confidence in templates based on this approach—that is, EOBNR models—is derived from confidence in NR.

The EOBNR models are effective mediating instruments for use in gravitational wave detection and compact binary observation. They combine insights from a range of modelling approaches, and are hence well verified through their relationships to these other models. However, producing models that incorporate all the relevant physical effects for the full range of possible binary systems is still a significant challenge. At the time of the O1 observing run, there were no models that could take account of all such effects (e.g., eccentricity and higher order modes in the presence of spin) for the full range of possible binary systems (Abbott et al. 2016c, 4). However, this is an area of active research; since GW150914, simulation studies have continued to explore new regions of the parameter space. For a summary of progress in numerical relativity simulations of compact binaries in the twenty-first century, see Duez and Zlochower (2018).

Ultimately, one major source of modelling error for the templates comes from the practical limitations on NR simulations. NR simulations can be used as benchmarks, to tune the EOB models, but they cannot be used exclusively to generate a whole

[24] The agreement across variations in the simulation methods forms the basis for a robustness argument: the simulation outputs are considered to be robust (and hence reliable) due to the agreement across independent methods. As with other robustness arguments, such reasoning relies on the methods being genuinely independent (see e.g., Staley (2004) and Dethier (2020)). Note that while robustness arguments of this kind have been considered controversial in the context of climate modelling, the present case appears to be one where robustness arguments are considered to be uncontroversial (if fallible). However, a comparative study of these cases is beyond the scope of this paper.

template library. This means that the majority of templates have not been directly compared to full NR simulations (although they are based on extrapolations from such comparisons). Following detection and parameter estimation, new NR simulations are performed using the physical parameters that have been estimated from the source to test for agreement with the measured signal.

3.4 Theoretical Bias

Yunes and Pretorius (2009) discuss what they call a "fundamental bias" in the methodology of gravitational-wave astrophysics: the assumption of the validity of general relativity, and the Einstein field equations.

It is important to note at the outset that we do have high confidence in the theory of general relativity *within the regimes in which it has been tested*. So far, general relativity has stood up to every empirical test we have thrown at it, from Einstein's successful retrodiction of the precession of the perihelion of Mercury, to the prediction of the orbital decay of the Hulse-Taylor pulsars.[25] However, these tests alone cannot justify extrapolation to the extreme conditions present when two black holes coalesce.

The success of general relativity under previously-probed conditions provides no guarantee of its success under the extreme conditions present in a binary black hole merger. Binary black hole mergers involve both high velocities and strong gravity, placing them firmly in the dynamical strong field regime. For all we know, another theory of gravity (or quantum gravity) might distinguish itself as the better theory in such regimes. In advance of empirical investigations of such regimes, we cannot assume that general relativity provides an adequate description of merger dynamics.

These concerns about the theory in turn give some reason to be concerned about the descriptions of specific systems provided by our models of binary black hole mergers. After all, if the conditions present in binary black hole mergers turn out to be beyond the domain of applicability of general relativity, then any models based on this theory may also be inaccurate under these conditions. We do have good reason to think that the part of the waveform—the early inspiral—that is based on post-Newtonian approximations will hold up. After all, these approximations are within the regime that has been tested already by observation of Hulse-Taylor binaries. However, numerical relativity simulations of the plunge and merger may well be inaccurate if general relativity fails to be an accurate description of the system when we reach the dynamical strong field regime. Since these simulations are used as a benchmark for ensuring the accuracy of other models in the template bank, this inaccuracy would likely infect the other models too. Insofar as current models *are* good models of the system according to general relativity, any deviations

[25] Of course, from the perspective of proponents of (relativistic extensions of) Modified Newtonian Dynamics (MOND), the empirical discrepancies that are usually attributed to dark matter in order to save general relativity are instead a motivation to modify general relativity. On this view, of course, general relativity has not stood up to all the empirical tests it has faced.

of the dynamics of the system from those predicted by general relativity will render the models inaccurate descriptions of the target systems that they are supposed to represent.

The possibility of deviations from general relativity in the strong regime leads to a fundamental bias in our inferences about these systems. This could have an impact both on the observation of gravitational waves (through matched-filtering) and on the inferences we make about the source (through parameter estimation).

First, theoretical bias could lead to non-optimal filtering. A possible example of this, given by Yunes and Pretorius (2009, 3), is that deviations from general relativity due to scalar radiation during late stages of the merger could lead to late time de-phasing with general relativity templates (due to inspiral occurring faster). This could lead to a systematically smaller SNR for detections, and thus systematic overestimation of the distance to the source. On a population level, we may end up concluding that such events occur more often farther away (i.e. further in the past).

Second, theoretical bias can be introduced at the level of parameter estimation, where the assumption of the accuracy of general relativity leads us to the inaccurate conclusions about the systems being observed. Yunes and Pretorius (2009, 3) provide the following example of this:

> For a second hypothetical example, consider an extreme mass ratio merger, where a small compact object spirals into a supermassive BH [black hole]. Suppose that a Chern-Simons (CS)-like correction is present, altering the near-horizon geometry of the BH [. . .] To leading order, the CS correction reduces the effective gravitomagnetic force exerted by the BH on the compact object; in other words, the GW emission would be similar to a compact object spiraling into a GR Kerr BH, but with smaller spin parameter a. Suppose further that near-extremal ($a \approx 1$) BHs are common (how rapidly astrophysical BHs can spin is an interesting and open question). Observation of a population of CS-modified Kerr BHs using GR templates would systematically underestimate the BH spin, leading to the erroneous conclusion that near-extremal BHs are uncommon, which could further lead to incorrect inferences about astrophysical BH formation and growth mechanisms.

Thus the assumption that general relativity provides an accurate description of black hole coalescence may bias parameter estimation and any subsequent inferences.

4 Model Validation with the LIGO-Virgo Observations

We have now seen that there are a range of reasons that the models used by the LIGO-Virgo Collaboration might provide inaccurate descriptions of binary black hole mergers, at least in the late stages of these events. Any inaccuracies could lead to systematic biases in the inferences that we make about such systems. In advance of any empirical testing, it is impossible to be sure that these models are accurate. *Prima facie* it seems possible that the LIGO-Virgo measurements themselves could be used to validate the models of the systems that they are observing. Thus they could be used to demonstrate that general relativity provides a good description of such events. However, this is called into question by the model-dependence of the observations, in terms of both the matched filtering needed to optimally retrieve

gravitational wave signals, and the Bayesian parameter estimation used to determine the properties of the system being observed.

The basic problem is this: testing the validity of the models using the LIGO-Virgo observations relies on a parameter estimation process that presupposes the validity of the models being used—models such as EOBNR and IMRPhenom. Thus any empirical test seems to implicate us in a circular justification scheme. We can only test the predictions of general relativity if we know the properties (mass, spin, etc.) of the objects we are observing, but we can only estimate these properties by assuming that our general relativistic models of these objects are accurate. Essentially, this is because we can only test general relativity insofar as we test its predictions about the dynamical behaviour of known objects (and consequences of this for gravitational wave emission, etc.). However, for binary black hole mergers, our only way of learning about these objects is via models that presuppose the accuracy of general relativity within the dynamical strong field regime.

Clearly this circularity problem has some connections with (Collins 1985)'s "Experimenter's Regress," according to which we do not know whether we have made a good measuring device until we have one that gives us the right results, but we do not know what the right results are until we know that we have a good measuring device (that is producing those results). Aside from describing this problem as a regress, Collins sometimes describes this as a circularity between the measurement device and the measurement result; the validity of each depends on the validity of the other. According to Collins, using an experiment as a test requires finding a way 'to break into the circle' (84). Controversially, Collins thinks that this circularity is broken by social negotiation rather than rational arguments, while others, such as Franklin (1994) argue that epistemic criteria are sufficient to break the circle.

In the case of the circularity problem I have described for LIGO-Virgo, it is not (primarily) the detectors themselves that are in question. In this sense, it is a distinct problem to the one concerning Collins, which is best understood as applying to gravitational wave detection.[26] *Even if* we think that we can 'break the circle' described by Collins (i.e., we are confident that the interferometers have been successfully used to detect gravitational waves), the circularity I describe with respect to the observation of black holes presents a further problem to be resolved.

When justifying the LIGO-Virgo results qua detection of gravitational waves, confidence in the detector plays an important role (Elder 2020, 2021b). Here, the kinds of rational arguments described by Franklin feature prominently. For example, confidence in understanding the response of the detector is established through calibration procedures, using lasers to push the interferometer test masses, mimicking the effect of a passing gravitational wave. Without dismissing the social dimensions of scientific collaboration and discovery, I think it is fair to say that these epistemic considerations play a persuasive justificatory role in the the LIGO-Virgo detection claim. However, when it comes to the observation of binary black hole mergers, the

[26] Nonetheless, insofar as the models are embedded in the experimental methodology, there is a sense in which this could be understood as a kind of experimenter's regress, spelled out in terms of models rather than detectors. There are also clear similarities here to the related "simulationist's regress" (Gelfert 2012; Meskhidze 2017).

circularity I have described presents a further (and more difficult) problem because these usual avenues for breaking the circle are unavailable.

Beyond the circularity itself, other features of the epistemic situation make the problem of model validation a particularly challenging one. First, the binary black hole systems being observed are distant astrophysical systems, meaning that it is not possible to manipulate or intervene on them in any way. Thus the analogue of calibration is not possible, since we cannot test the interferometer response to a binary black hole merger with known properties. Second, we have no independent access to these systems, given that black holes emit no electromagnetic radiation. This rules out using consilience, or coherence testing to improve confidence in the LIGO-Virgo results (and methods).[27] Third, the mergers themselves occur in regimes that have never been probed before—the dynamical strong field regime. This means that the previous success of general relativity provides no guarantee that the theory will continue to provide accurate descriptions in the regimes being probed by LIGO-Virgo. The lack of interventions or independent empirical access combined with the novel regimes being probed renders the problem of theoretical bias particularly acute, bringing the circularity problem beyond more generic issues of theory- or model-ladenness in empirical science. This circularity threatens to mask any bias implicit in the models.

Nonetheless, the LIGO-Virgo Collaboration does perform a number of tests of general relativity. Such tests seem to offer some empirical validation of the general relativity-based models used to observe binary black hole mergers. Indeed, taken together, they seem to constitute a kind of methodological bootstrapping (Glymour 1975). Each of the tests probes different assumptions that go into the observation of binary black hole mergers—despite making some assumptions based on general relativity in the course of the tests. In particular, these tests involve looking for evidence that the model-dependent methodology—and circularity—could be biasing observations.

A detailed examination of theory-testing with LIGO-Virgo will be the subject of a future paper. However, a brief consideration of two tests helps illustrate how the LIGO-Virgo Collaboration is able to empirically validate models like EOBNR in the face of the circularity problem.[28]

First, the "residuals test" considered in Abbott et al. (2016d) tests the consistency of the residual data with noise. This involves subtracting the best-fit waveform from the GW150914 data and then comparing the residual with detector noise (for time periods where no gravitational waves have been detected). The idea here is to check whether the waveform has successfully removed the entire gravitational-wave signal from the data, or whether some of the signal remains. This process places constraints on the residual signal, and hence on the deviations from the best-fit waveform that could be present in the data. However, this doesn't constrain deviations from general relativity *simpliciter*, due to the possibility that the best-

[27] See Bokulich (2020) for discussion of the distinction between consilience and coherence testing.

[28] See also Patton (2020) for an excellent discussion of a different test, based on the parameterised post-Einsteinian framework developed by Yunes and Pretorius, including how this connects to the issue of "fundamental theoretical bias".

fit general relativity waveform is degenerate with non-general relativity waveforms for events characterized by different parameters. That is, the same waveform could be generated by a compact binary merger (described by parameters different from those that we think describe the GW150914 merger) with dynamics that deviate from general relativistic dynamics. In this case, we could be looking at different compact objects than we think we are, behaving differently than we think they are, but nonetheless producing very similar gravitational wave signatures. This is stated (though not fully explained) in the following passage:

> We use this estimated level [of residual] to bound GR violations *which are not degenerate with changes in the parameters of the binary* (2, emphasis mine).

This test could potentially show inconsistency with general relativity, but not all deviations from general relativity will be detectable in this way. Thus the test shows that GW150914 is consistent with general relativity, but the methodology of this test could also be masking such deviations due to the fundamental bias associated with assuming general relativity for the purposes of parameter estimation.

Second, the "IMR consistency test" considered in Abbott et al. (2016d) considers the consistency of the (early) low-frequency part of the gravitational-wave signal with the (later) high-frequency part. This test proceeds as follows. First, the masses and spins of the two compact objects are estimated from the inspiral (low-frequency), using LALInference. This gives posterior distributions for component masses and spins. Then, using formulas derived from numerical relativity, posterior distributions for the remnant, post-merger object are computed. Finally, posterior distributions are also calculated directly from the measured post-inspiral (high-frequency) signal, and the two distributions are compared. These are also compared to the posterior distributions computed from the inspiral-merger-ringdown waveform as a whole. If there are any deviations from general relativity to be found, these are expected to occur in the late part of the signal, where the full non-linear Einstein field equations are needed and approximations are known to become invalid. In contrast, previous empirical constraints give us reason to doubt that such deviations will be significant for the early inspiral. In the presence of high-frequency deviations, parameter estimation based on general relativity models will deviate from the values of a system that is well-described by general relativity. Hence (in such cases) we can expect the parameter values estimated from the low frequency part of the signal to show discrepancies with the parameter values estimated from the high frequency part of the signal. This leaves the test open to detection of subtle deviations from general relativity; if the parameters associated with the two waveforms are different, this could suggest some deviation from general relativity (Abbott et al. 2016d). Interestingly, it does so despite assuming the validity of general-relativistic descriptions at each step in the process.

Although these two tests are just consistency tests, they place constraints on ways that the errors in the models could be undermining the accuracy of observations. Taken together, they constrain both the accuracy of the extracted gravitational waveform and the consistency of this waveform with general relativity. Further tests performed by the LIGO-Virgo place further constraints on loopholes in their

methods—that is to say, these tests place constraints on the extent to which particular aspects of the LIGO-Virgo methodology could be biased by inadequate modelling of the target system. In doing so, they can be understood as a response to the circularity problem that I have described in this paper.[29]

5 Conclusion

In this paper, I have argued that modelling plays an essential role in connecting high-level theory, embodied in the Einstein field equations, with the LIGO-Virgo data. The models used in template-based searches for gravitational waves and in parameter estimation incorporate insights from a range of modelling approaches, allowing us to gain empirical access to binary black hole mergers.

However, I have also argued that the model-dependent methods used by the LIGO-Virgo Collaboration to observe binary black hole mergers lead to some epistemic challenges; the potential bias introduced through the use of general relativity-based models leads to a circularity problem for the validation of these models in the regimes probed by the LIGO-Virgo Collaboration. Observations of binary black hole systems are based on models of such systems, and confidence in the accuracy of these observations depends on the validity of the models being used. Thus using these observations to validate these models is problematically circular. (However, I briefly mentioned some ways that the LIGO-Virgo proceeds in validating their models in spite of this circularity.)

Overall, this paper shows how the methodology of the LIGO-Virgo experiments is intimately bound up with models—of binary black hole mergers and the gravitational waves that they produce. The success of these experiments rests on confidence in these models, which bridge the gap between theory and phenomena. The flip side of this is that much of the interesting, and challenging work in validating the LIGO-Virgo results lies in validating the models themselves with respect to both the equations of general relativity and the physical systems being observed.

[29] For further discussion of theory testing in gravitational-wave astrophysics, see Elder (forthcoming).

References

Abbott, B. P., et al. 2016a. "Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914." *Classical and Quantum Gravity* 33, no. 13 (June): 134001. https://doi.org/10.1088/0264-9381/33/13/134001.

———. 2016b. "Observing gravitational-wave transient GW150914 with minimal assumptions." *Physical Review D* 93 (12): 122004. https://doi.org/10.1103/PhysRevD.93.122004.

———. 2016c. "Properties of the Binary Black Hole Merger GW150914." *Physical Review Letters* 116 (24): 241102. https://doi.org/10.1103/PhysRevLett.116.241102.

———. 2016d. "Tests of General Relativity with GW150914." *Physical Review Letters* 116 (22): 221101. https://doi.org/10.1103/PhysRevLett.116.221101.

———. 2020. "A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals." *Classical and Quantum Gravity* 37, no. 5 (February): 055002. https://doi.org/10.1088/1361-6382/ab685e.

Abedi, Jahed, Hannah Dykaar, and Niayesh Afshordi. 2017. "Echoes from the Abyss: Tentative evidence for Planck-scale structure at black hole horizons." *Physical Review D* 96 (8): 082004.

Baker, John G, Manuela Campanelli, Frans Pretorius, and Yosef Zlochower. 2007. "Comparisons of binary black hole merger waveforms." *Classical and Quantum Gravity* 24, no. 12 (May): S25–S31. https://doi.org/10.1088/0264-9381/24/12/s03.

Baker, John G, Joan Centrella, Dae-Il Choi, Michael Koppitz, and James van Meter. 2006. "Gravitational-wave extraction from an inspiraling configuration of merging black holes." *Physical Review Letters* 96 (11). https://doi.org/10.1103/PhysRevLett.96.111102.

Blanchet, L., and T. Damour. 1986. "Radiative Gravitational Fields in General Relativity I. General Structure of the Field outside the Source." *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 320 (1555): 379–430.

Blanchet, Luc. 1987. "Radiative Gravitation Fields in General Relativity II. Asymptotic Behaviour at Future Null Infinity." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 409 (1837): 383–399.

———. 1998. "On the multipole expansion of the gravitational field." *Classical and Quantum Gravity* 15, no. 7 (July): 1971–1999. https://doi.org/10.1088/0264-9381/15/7/013.

Bokulich, Alisa. 2020. “Calibration, Coherence, and Consilience in Radiometric Measures of Geologic Time.” *Philosophy of science* (Chicago) 87 (3): 425–456.

Buonanno, Alessandra, and Thibault Damour. 1999. “Effective one-body approach to general relativistic two-body dynamics.” *Physical Review D* 59 (8): 084006. https://doi.org/10.1103/PhysRevD.59.084006.

———. 2000. “Transition from inspiral to plunge in binary black hole coalescences.” *Physical Review D* 62 (6): 064015. https://doi.org/10.1103/PhysRevD.62.064015.

Campanelli, M, C O Lousto, P Marronetti, and Y Zlochower. 2006. “Accurate evolutions of orbiting black-hole binaries without excision.” *Physical Review Letters* 96 (11). https://doi.org/https://doi.org/10.1103/PhysRevLett.96.111101.

Campbell, Donald, and Julian Staley. 1966. *Experimental and quasi-experimental designs for research.* Chicago: R. McNally.

Collins, Harry. 1985. *Changing Order: Replication and Induction in Scientific Practice.* University of Chicago Press.

———. 2017. *Gravity’s Kiss: The Detection of Gravitational Waves.* Cambridge MA: MIT Press.

Cook, Thomas D. 1979. *Quasi-experimentation: design & analysis issues for field settings.* Boston: Houghton Mifflin.

Cronbach, Lee, and Paul Meehl. 1955. “Construct Validity in Psychological Tests.” *Psychological Bulletin* (Washington, etc.) 52.

Dethier, Corey. 2020. *Multiple Models, Robustness, and the Epistemology of Climate Science.* Notre Dame, Indiana.

Duez, Matthew D, and Yosef Zlochower. 2018. “Numerical relativity of compact binaries in the 21st century.” *Reports on Progress in Physics* 82, no. 1 (November): 016902. https://doi.org/10.1088/1361-6633/aadb16.

Elder, Jamee. 2020. “The Epistemology of Gravitational-wave Astrophysics.” PhD diss., University of Notre Dame.

———. 2021a. “Independent Evidence in Multi-messenger Astrophysics.”

———. 2021b. “On the ‘Direct Detection’ of Gravitational Waves.”

———. Forthcoming. “Theory Testing in Gravitational-wave Astrophysics.” In *Philosophy of Astrophysics: Stars, Simulations, and the Struggle to Determine What is Out There,* edited by Nora Mills Boyd, Siska De Baerdemaeker, Vera Matarese, and Kevin Heng. Synthese Library.

Fillion, Nicolas, and Sorin Bangu. 2015. "Numerical Methods, Complexity, and Epistemic Hierarchies." *Philosophy of science* 82 (5): 941–955.

Franklin, Allan. 1994. "How to Avoid the Experimenters' Regress." *Studies in History and Philosophy of Science Part A* 25 (3): 463–491. https://doi.org/10.1016/0039-3681(94)90062-0.

———. 2015. "The Theory-Ladenness of Experiment." *Journal for General Philosophy of Science* (Dordrecht) 46 (1): 155–166. https://doi.org/10.1007/s10838-015-9285-9.

Gelfert, Axel. 2012. "Scientific Models, Simulation, and the Experimenter's Regress." In *Models, Simulations, and Representations,* edited by Cyrille Imbert and Paul Humphreys, 145–167. Routledge studies in the philosophy of science 9. New York: Routledge.

Glymour, Clark. 1975. "Relevant Evidence." *The Journal of Philosophy* 72 (14): 403–426.

Hannam, Mark, Sascha Husa, John G. Baker, Michael Boyle, Bernd Brügmann, Tony Chu, Nils Dorband, et al. 2009. "Samurai project: Verifying the consistency of black-hole-binary waveforms for gravitational-wave detection." *Physical Review D* 79 (8): 084025. https://doi.org/10.1103/PhysRevD.79.084025.

Havas, Peter. 1989. "The Early History of the Problem of Motion in General Relativity." In *Einstein and the history of general relativity: based on the proceedings of the 1986 Osgood Hill Conference, North Andover, Massachusetts, 8-11 May 1986,* edited by Don Howard and John Stachel. Einstein studies v. 1. Boston: Birkhäuser.

———. 1993. "The Two-Body Problem and the Einstein-Silberstein Controversy." In *The Attraction of gravitation: new studies in the history of general relativity,* edited by John Earman, Michel Janssen, and John Norton. Einstein studies v. 5. Boston: Birkhäuser.

Holst, Michael, Olivier Sarbach, Manuel Tiglio, and Michele Vallisneri. 2016. "The emergence of gravitational wave science: 100 years of development of mathematical theory, detectors, numerical algorithms, and data analysis tools." *Bulletin of the American Mathematical Society* 53:513–554. https://doi.org/https://doi-org.proxy.library.nd.edu/10.1090/bull/1544.

Kennefick, Daniel. 2007. *Travelling at the Speed of Thought.* Princeton University Press.

Le Tiec, Alexandre. 2014. "The overlap of numerical relativity, perturbation theory and post-Newtonian theory in the binary black hole problem." *International Journal of Modern Physics D* 23 (10): 1430022. https://doi.org/https://doi.org/10.1142/S0218271814300225.

Lovelace, Geoffrey, Carlos O Lousto, James Healy, Mark A Scheel, Alyssa Garcia, Richard O'Shaughnessy, Michael Boyle, et al. 2016. "Modeling the source of GW150914 with targeted numerical-relativity simulations." *Classical and Quantum Gravity* 33, no. 24 (November): 244002. https://doi.org/10.1088/0264-9381/33/24/244002.

Maggiore, Michele. 2008. *Gravitational waves. Volume 1, Theory and experiments.* Oxford: Oxford University Press.

Meskhidze, Helen. 2017. "Simulationist's Regress in Laboratory Astrophysics."

Morgan, Mary, and Margaret Morrison. 1999. "Models as Mediating Instruments." In *Models as Mediators: Perspectives on Natural and Social Science,* edited by Mary Morgan and Margaret Morrison, 10–37. Cambridge University Press.

Parker, Wendy S. 2017. "Computer Simulation, Measurement, and Data Assimilation." *British Journal for the Philosophy of Science* 68 (1): 273–304. https://doi.org/https://doi.org/10.1093/bjps/axv037.

Pati, Michael E., and Clifford M. Will. 2000. "Post-Newtonian gravitational radiation and equations of motion via direct integration of the relaxed Einstein equations: Foundations." *Physical Review D* 62 (12): 124015. https://doi.org/10.1103/PhysRevD.62.124015.

Patton, Lydia. 2020. "Expanding theory testing in general relativity: LIGO and parametrized theories." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 69:142–153. https://doi.org/https://doi.org/10.1016/j.shpsb.2020.01.001.

Peters, P. C. 1964. "Gravitational Radiation and the Motion of Two Point Masses." *Physical Review (U.S.) Superseded in part by Physical Review A, Physical Review B: Solid State, Physical Review C, and Physical Review D* 136 (4B). https://doi.org/10.1103/PhysRev.136.B1224.

Pretorius, Frans. 2005. "Evolution of binary black-hole spacetimes." *Physical Review Letters* 95 (12). https://doi.org/10.1103/PhysRevLett.95.121101.

Shapere, Dudley. 1982. "The Concept of Observation in Science and Philosophy." *Philosophy of Science* 49 (4). https://doi.org/https://doi.org/10.1086/289075.

Sperhake, Ulrich. 2015. "The numerical relativity breakthrough for binary black holes." *Classical and Quantum Gravity* 32 (12): 124011. https://doi.org/10.1088/0264-9381/32/12/124011.

Staley, Kent W. 2004. "Robust Evidence and Secure Evidence Claims." *Philosophy of Science* 71 (4): 467–488. https://doi.org/10.1086/423748.

Suppes, Patrick. 1962. "Models of Data." In *Logic, Methodology and Philosophy of Science Proceedings of the 1960 International Congress,* edited by Ernest Nagel, Patrick Suppes, and Alfred Tarski, 252–61. Stanford University Press.

Tal, Eran. 2012. *The Epistemology of Measurement: A Model-Based Account.* http://search.proquest.com/docview/1346194511/.

———. 2013. "Old and New Problems in Philosophy of Measurement." *Philosophy Compass* 8 (12): 1159–1173. https://doi.org/https://doi.org/10.1111/phc3.12089.

Veitch, J., et al. 2015. "Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library." *Physical Review D* 91 (4): 042003. https://doi.org/10.1103/PhysRevD.91.042003.

Will, Clifford M., and Alan G. Wiseman. 1996. "Gravitational radiation from compact binary systems: Gravitational waveforms and energy loss to second post-Newtonian order." *Physical Review D* 54 (8): 4813–4848. https://doi.org/10.1103/PhysRevD.54.4813.

Winsberg, Eric. 2019. "Computer Simulations in Science." In *The Stanford Encyclopedia of Philosophy,* Spring 2019, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University.

Yunes, Nicolás, and Frans Pretorius. 2009. "Fundamental theoretical bias in gravitational wave astrophysics and the parametrized post-Einsteinian framework." *Phys. Rev. D* 80 (12): 122003. https://doi.org/10.1103/PhysRevD.80.122003. https://link.aps.org/doi/10.1103/PhysRevD.80.122003.