On the “Direct Detection” of Gravitational Waves

Jamee Elder

March 28, 2023

1 Introduction

On September 14, 2015 the LIGO observatories in Hanford, WA and Livingston, LA detected gravitational waves for the first time. In the “discovery” paper the LIGO-Virgo Collaboration describe this event, “GW150914”, as the first “direct detection” of gravitational waves and the first “direct observation” of a binary black hole merger (Abbott et al. 2016a, 061102–1) *Prima facie*, these are somewhat puzzling claims. First, there is something counter-intuitive about describing such a sophisticated experiment as a “direct” detection, insofar as this suggests that the procedure was simple or straightforward. Even strong gravitational waves produce only a tiny change in the length of the $4km$ interferometer arms. The challenge in detecting gravitational waves is in showing that such a tiny change has really taken place. Doing so makes use of complex experimental apparatus, a library of templates of possible signals, and sophisticated data analysis techniques. To describe this whole complex process as “direct” is, on the surface at least, somewhat surprising. Second, black holes are not the kinds of things that we usually think of as “observable”—even if we happened to be sitting right outside the event horizon of one. So it is surprising to see the LIGO-Virgo Collaboration claim to have observed astrophysical black holes from their observatory here on Earth, let alone *directly* observed it.

In this paper, I aim to clarify debates about what counts as a direct (vs. indirect) detection in gravitational-wave astrophysics. In doing so, I highlight the epistemic considerations that lurk behind the verbal disputes. In particular, an analysis of these debates sheds light on some important features of model-based inference in measurement processes, pointing to an important epistemic difference between the empirical investigation of gravitational waves and of the compact binaries that produce them. By clarifying these debates, I clarify the nature of our empirical access to both gravitational waves and black holes. In particular, our empirical access to black holes depends on models of a separate target system in a way that our empirical access to gravitational waves does not. As I discuss elsewhere (Elder, forthcoming), this theory-ladenness, or model-dependence, turns out to be one of the key challenges of gravitational-wave astrophysics. A careful characterization of the difference between direct and indirect detection thus has payoffs for understanding the epistemic situation of gravitational-wave astrophysics as a whole.

The paper proceeds as follows. I begin, in Section 2, by describing a controversy that arose within the LIGO Scientific Collaboration about whether or not to describe their detection of gravitational waves as a "direct" detection (or observation). In Section 3 I provide philosophical analyses of three (related) empirical activities: "detection," "observation," and "measurement." In Section 4 I consider a recent attempt by Allan Franklin (2017) to clarify the direct/indirect distinction, but find it wanting. In section 4, I articulate my own positive account of the distinction between a direct and indirect detection in this context. Roughly, my account leverages the conceptual resources from recent work in the philosophy of measurement (especially Parker (2017)'s distinction between direct and derived measurements) to distinguish between these detections at the level of the modeling of the measurement processes. This distinction also has epistemic importance, because the choices scientists make about how to model measurement processes are related to the kinds of interventions they can perform to test the adequacy of their models. The direct/indirect distinction concerns the nature of the justification for confidence in the measurement outcome—in the direct case, this is based primarily on models of the measuring system, while in in the indirect case it also relies on models of a separate target system. Since astrophysical systems are not amenable to interventions, observation of the Hulse-Taylor system, and indeed the source of GW150914, cannot be "direct" in the same way that detections of gravitational waves are. Finally, in Section 4.3 I discuss the scope and significance of the distinction, thus understood. This includes applying the distinction to some other, related case studies, including previous (attempted) observations of gravitational waves and black holes.

2 Controversy Within the LIGO Scientific Collaboration

As Harry Collins discusses in his book, *Gravity's Kiss* (2017), there was significant controversy among LIGO scientists about how to describe their achievement. In particular, there was much discussion over whether to refer to it as a "direct" detection or observation of gravitational waves (and where in the paper to do so).

Calling the LIGO-Virgo detection a "direct detection" did have a clear aim: to distinguish it from other methods of (indirect) detection, including the previous detection of gravitational waves by Hulse, Taylor, and Weisberg (Hulse and Taylor 1975;

Taylor and Weisberg 1982). The decay in the orbital period of the Hulse-Taylor binary was determined from changes in the arrival times of the electromagnetic pulses from one of the components of the binary. The rate of decay was found to be consistent with the predicted decay due to gravitational radiation from such a binary. This provided compelling evidence for the existence of gravitational waves. Indeed, the LIGO-Virgo team credit Hulse, Taylor, and Weisberg with demonstrating the existence of gravitational waves:

> The discovery of the binary pulsar system PSR B1913+16 by Hulse and Taylor and subsequent observations of its energy loss by Taylor and Weisberg demonstrated the existence of gravitational waves. (Abbott et al. 2016a, 061102–1)

Describing the LIGO-Virgo detection as the first “direct detection” is intended to distinguish it from this previous demonstration. The use of the term “direct” is also supposed to distinguish the LIGO-Virgo method of detection from attempts at “indirect” detection of gravitational waves from signatures in the cosmic microwave background. For example, such methods were employed by the BICEP2 experiment.[1] This experiment was ultimately deemed to have failed to detect gravitational waves, but had it succeeded, it seems that it would have stood alongside the Hulse-Taylor-Weisberg detection as an indirect detection, in contrast to the “direct detection” by the LIGO-Virgo Collaboration (Collins 2017, 142). This paper will mainly focus on the distinction between the two successful detections by Hulse, Taylor, and Weisberg and the LIGO-Virgo Collaboration. However, I will briefly consider the case of BICEP2 in my discussion of the broader applicability of my account of direct detection in section 4.3 below.

Despite the seemingly clear goals of the “direct/indirect detection” terminology, the writing of the discovery paper generated discussion over whether or not to make such a distinction explicitly. This debate had both political and philosophical dimensions.

On the political side, there was concern that describing the detection as “direct” might be seen as diminishing the significance of the previous detections and offend members of the astrophysics community. For example, one unnamed scientist is quoted by Collins as advocating for a modest tone:

> I would also like to suggest that we avoid the phrases “direct observation” and “direct detection,” at least in the title, abstract, and introduction. While I know that we use those terms as a way of distinguishing what we do from other methods of GW detection . . . I have come to understand that there are

1. This experiment involved the detection of polarization patterns in the cosmic microwave background, which were initially attributed to gravitational waves. This claim was later retracted in response to evidence that this effect could be fully accounted for by other sources (such as dust).

> people in our broader community who think that this terminology is meant to diminish the importance of their work. In keeping with the modest tone I advocate, I see no benefit in using these particular terms. Indeed, without further definition, they do not convey any clear meaning. (Collins 2017, 147–8)

Others argued that LIGO was widely understood to be aiming for a direct detection, and expected to announce it as such when they had been successful. One member of the collaboration even surveyed popular science articles and concluded that the public would consider their achievement the first direct detection of gravitational waves (144–5). Another pointed out that the LIGO mission statement reads: "LIGO's mission is to open the field of gravitational-wave astrophysics through the direct detection of gravitational waves" (153). Under such circumstances, they argued, it would be strange and unreasonably modest *not* to claim the first direct detection. While this political dimension to the dispute is interesting, it doesn't tell us what, if anything, of conceptual or epistemic significance is at stake.

There was also a distinctly philosophical dimension to these arguments. Many thought it was clear that the LIGO-Virgo detection was direct, while the previous detection from the Hulse-Taylor binary was not. However, some argued that there was no clear distinction between the two on the basis of directness, or that neither could reasonably be called direct. For example, Taylor (perhaps unsurprisingly) questions whether the LIGO-Virgo detection was any closer to being a "direct" detection than that achieved with the Hulse-Taylor binary:

> In the binary pulsar experiment, and also in a LIGO-like experiment, one infers the presence of gravitational radiation based on effects it induces in a "detector." If a ruler could be used to measure the displacement of LIGO's test masses, I would grant that detection to be rather more "direct" than one based on timing measurements of an orbiting pulsar halfway across our Galaxy. However, LIGO can't use a ruler; instead they use servomechanisms, very sensitive electronics . . . and finally long sequences of calculations to infer that a gravitational wave has passed by. Such a detection, like the binary pulsar timing experiment, is arguably many stages removed from being what most people would call "direct." (148)

Some scientists within the LIGO-Virgo Collaboration denied that the direct/indirect distinction was a helpful one for explaining what is special about the LIGO detection. For example, one unnamed scientist says:

> Personally, I have never felt motivated by the direct detection phrase and thought it was a red herring. We saw some masses, which happened to be

> mirrors, moving under the influence of gravitational waves. Hulse, Taylor, and Taylor's later collaborators saw some masses, which happened to be neutron stars, moving under the influence of gravitational waves. What's the difference? (Collins 2017, 148)

However, the same scientist goes on to state what they take to be special about the LIGO detection, answering their own rhetorical question:

> The difference is that the Taylor crowd observed a GW transmitter and figured out how it worked! We figured out how to build a sufficiently sensitive GW receiver and *since we built it, we know exactly how it works*. If anybody misses the impact of those italicized words, then check out BICEP-2 and Planck's experience to date. Those words in italics represent a huge advance for GW physics and astronomy. (148)

It seems clear that the scientist did not take this difference to suggest a notion of directness. Indeed, they might be taken as suggesting that directness itself is a red herring, while the epistemically significant difference lies in what Collins calls the "unprecedented control over nature" (143). However, I will argue later (Section 4) that this control is in fact related to the relevant notion of "direct" in this context.

In response to the controversy over the appropriate way to describe the LIGO-Virgo detections, those in charge of writing the discovery paper found a surprising way of resolving the issue: they polled the entire collaboration. Among other things, they asked collaboration members to vote on the use of terms like "direct", "detection" and "observation" in both the title and body of the paper. The results were as follows:

> Poll 1: No direct and no LIGO in the title. Preferences are Observation or Detection of GW from BBH merger. Poll 3: It is OK to use direct (detection/observation) in the body of the paper.[2] (156)

This result was reflected in the published paper.

This seems like an ideal place for philosophical intervention. A poll is hardly a philosophically satisfactory way to resolve a debate over the use of epistemically-loaded notions like direct detection and direct observation. The results give us no indication of *why* participants voted the way they did. Given that the poll was a guide to those writing the paper, we are left wondering how to understand the terminological choices that were made as well as their epistemic implications. Even if the poll results had reported the detailed reasoning of each scientist involved, this wouldn't necessarily settle the matter. The views of active scientists on the terminology they use are potentially

2. Collins omits mention of Poll 2 here because it concerns issues orthogonal to the present discussion.

important data, but these views need not be taken as definitive. Indeed, philosophy of science frequently involves conceptual clarification, revision or even outright disagreement with what scientists take themselves to be doing.

In what follows, I aim to give an account of the term "direct detection" that tracks the distinction being made by the LIGO-Virgo Collaboration in Abbott et al. (2016a), while also considering the epistemic significance of this distinction. Another attempt at this task can be found in Franklin (2017). Franklin's purpose in this paper is ostensibly to provide an account of what makes an observation direct or indirect in physics. This is carried out by detailed examination of a number of case studies, including both the LIGO-Virgo and Hulse-Taylor-Weisberg detections of gravitational waves. In Section 4.1 I discuss Franklin's approach in more detail and argue that he fails to provide either a systematic account of this distinction in general or a philosophically satisfying analysis of the gravitational wave cases. For now, it is worth setting out what he says about these two case studies:

> In terms of Shapere's discussion of "direct observation," we see that the LIGO-VIRGO [*sic*] collaboration did make a direct observation. The gravity waves interacted with the two interferometers. The fact that there was complex instrumentation and analysis should not change that conclusion. It seems fair to say that the binary pulsar observation was indirect. The existence of gravity waves was inferred from the decrease in period of the pulsar, which was transmitted to the detector by electromagnetic radiation. As discussed below, I do not believe that the direct-indirect distinction had any epistemological significance. (410)

This, I think, is an intuitive, common-sense way to distinguish between the two detections. However, I think it is in great need of philosophical clarification, which I endeavor to provide. Also, while I agree with Franklin that the relative directness of two observations need not determine which is epistemically superior, I nonetheless think that the distinction is of epistemic significance. This is because it concerns the nature of our justification for the claims we make on the basis of observations.[3]

3 Detection, Observation, and Measurement

The controversy among scientists over the use of terms like "direct detection" and "direct observation" raises some interesting questions. First, what do we mean by "detection" and "observation"? Are these two interchangeable terms (as suggested by the results of Poll 1, above) or do they have different meanings? Second, we can ask

3. I discuss this in Section 4.2.

what (if anything) the term "direct" adds in each case. Is there any epistemic significance to the distinction between direct and indirect? In this section, I address the first of these questions, starting with detection then moving on to observation and a third, related, notion: measurement. For the most part, I leave consideration of the second question—concerning the significance of directness—to Section 4.

3.1 Detection

The first clear characteristic of a detection is that it is empirical. Like other related empirical activities—experiments, measurements, observations, etc.—detections involve some kind of physical interaction with the world through which we learn about a system of interest—the target system.

A second characteristic is that it aims to establish the existence of some entity (or phenomenon): a successful detection provides evidence that an entity of that kind exists (or existed) within the target system. Indeed, it must give us sufficient reason to accept the existence of that entity, relative to the background knowledge and standards of acceptance of the relevant scientific community. For example, in particle physics, the current standards of acceptance for detection of a new particle dictate that the statistical significance of the signal must be at least 5σ. I also take it that an account of successful detection should leave room for scientists to get it wrong. A detection that is seen as successful by the scientific community at one time can be shown to have been unsuccessful in light of new information, or by the revised standards of acceptance of a scientific community at a later time.[4]

Detection may take very different forms, depending on the field of science and the entity we are trying to detect. In the simplest case, detection may involve an observation in the ordinary sense, using our eyes or other sense organs. For example, the detection of kākāpō[5] on Hauturu may be a simple matter of seeing a kākāpō with one's own eyes. In more complicated cases, detection involves some kind of measuring process, such as the extremely complex measuring processes involved in the detection of the Higgs boson. There are also intermediate cases, involving simple, standard tests—such as the use of blue litmus paper to detect an acid.

Additionally, it seems that detection need not involve measuring or observing the

4. Discussion of the success of a detection also raises the question of realism and antirealism. On a realist interpretation, the success of a detection will partly rest on whether the entity in question is really present in the system where it is purportedly detected. Antirealists will have to spell out the success of a detection differently, without the realist commitment to the literal existence of the detected entity. I set aside such metaphysical questions for the purposes of this paper.

5. Kākāpō are a species of flightless parrot, native to New Zealand. Since they are critically endangered, surviving kākāpō are kept on predator-free islands.

entity itself. For example, it might be sufficient for detection of a kea[6] to note certain characteristic signs of their having been there—olive-green feathers, bird droppings, destruction of the rubber parts of one's car, etc. As long as these provide satisfactory evidence that the entity in question (a kea) has been there, this might be considered a detection of a kea. It seems to be this feature of detection that leads us to want to distinguish between direct and indirect detection. At first blush, this is the difference between detecting the entity itself and detecting some consequence of that entity's existence. However, the complexity of modern measuring processes makes this distinction rather less clear-cut in practice.

The purpose of detection can also vary, in ways that are already implicit in the above examples. First, detection can sometimes be for discovery, establishing the existence of a new, previously-undetected entity (e.g., the discovery of the Higgs boson). Second, a detection can corroborate existing evidence for such a purported discovery. Third, detection can play the somewhat different role of locating a familiar kind of entity within some spatio-temporal region. The kākāpō and the kea examples are detections of this kind. Another example of this kind of detection is a carbon monoxide detector: such a device alerts us to when a dangerous gas is present in our home. Finally, detection often acts as part of some broader empirical investigation in which we seek to learn more about an entity, its properties, and the target systems in which it is found. In such cases, we both locate an entity and learn something about what it is we are detecting. These roles of detection are neither exhaustive nor mutually exclusive. The LIGO-Virgo detections, for example, play all of the above roles except the first (given that the Hulse-Taylor-Weisberg detection sufficed for "discovery").

So far, I have said that detection can be either an observation or an (instrumental) measurement. Of course, depending on the notion of observation we choose, many (or all) measurements may also qualify as observations. So far, I have not presupposed anything about the overlap of these different empirical activities. But what are observations and what are measurements? I address these questions in sections 3.2 and 3.3 respectively.

3.2 Observation

Of the three empirical activities discussed here—detection, observation, and measurement—the notion of "observation" has received the most scrutiny in the philosophy literature. Views about what is "observable" range from extremely strict to extremely permissive.

Toward the strict end of the spectrum is Bas van Fraassen, who argues that something is literally observed only if it is perceived through the unaided senses (van

6. The kea is another New Zealand parrot, found in alpine regions. They are known for their intelligence and mischievous behaviour.

Fraassen 1980, 2008). Similarly Bogen and Woodward (1988) take observation to be tied to human perception, or to "processes which can be usefully viewed as extensions of perception."[7] According to these strict views we should reject the LIGO-Virgo Collaboration's claims to have directly observed a binary black hole merger, since this merger was imperceptible to our human-typical sense organs. Of course, members of the LIGO-Virgo Collaboration are perfectly aware that gravitational waves are imperceptible. They are simply choosing to use the term "observation" in a way that is at odds with the strict sense advocated by some philosophers.

On the other end of the spectrum, Shapere (1982) offers a permissive view of (direct) observation according to which we can use solar neutrinos to observe the inside of the sun. On Shapere's analysis, a direct observation is any transfer of information from source to detector via a physical messenger (photons, neutrinos, etc.) that is sufficiently uncorrupted by interference, *according to our best science and background knowledge*. He puts it as follows:

> x is directly observed (observable) if:
>
> 1. information is received (can be received) by an appropriate receptor; and
> 2. that information is (can be) transmitted directly, i.e., without interference, to the receptor from the entity x (which is the source of the information). (492)

Here, "current physical knowledge" tells us what counts as an "appropriate receptor," or as "information." This background knowledge is broken up into the "theory of the source," the "theory of the transmission," and the "theory of the receptor." Judgments about whether a candidate observation counts as such depend on each of these "theories" (i.e., how the source, transmission, and receptor are modeled).

Overall, Shapere argues that current usage by physicists generalizes the philosopher's restrictive notion of observation by allowing a broader range of receptors (i.e., including measuring instruments) and kinds of information (e.g., neutrinos, gravitational waves) to play roles in the observation process. Based on this analysis, Shapere says that we can use neutrinos to "directly observe" the interior of the sun, since neutrinos from this region are transmitted to our detectors with virtually no interference. However, we cannot similarly observe the interior of the sun using photons, since the sun is opaque to

7. It is worth noting that Bogen and Woodward certainly do not claim that scientists need to adhere to their strict notions of observation. Indeed, Bogen and Woodward (1988) explicitly say that they take no issue with scientists using the term "observe" in a different way. Their targets are not scientists but philosophers, especially those caught up in discussions of the theory-ladenness of observation and those (like van Fraassen) who take a strict view of observation while also claiming that theories explain what is observable (which they deny).

electromagnetic light, and consequently photons from this region cannot be transmitted to our detectors.

We have now seen two apparently conflicting viewpoints about how to use the term "observation." On the strict view, observation is about perception. This emphasizes a distinction between observing a *representation* of an entity produced by a measuring process—a page of data, a digital display on a screen, an image of a paramecium—and observing the entity itself. On the permissive view, there is no such privileging of human sense organs. Other instruments are taken to be on par with our eyes, and thus what we might call "instrumental observation" is merged with ordinary perception to form a broader category of observation. This view emphasizes that much of scientific practice has moved away from reliance on human sense perception as a means of gathering information about the world.

Without claiming to have resolved this dispute (indeed, I see no strong motivation for thinking we need to choose between the two uses), I will nonetheless assume a permissive view of observation in what follows, since this is the better fit with scientific practice (in particular, the practice of astrophysics as an "observational" science). The LIGO-Virgo Collaboration didn't observe a binary black hole merger in the strict sense, but it seems that they *did* observe it in some more permissive sense. Thus only the permissive sense provides a candidate notion for understanding the claims being made by the LIGO-Virgo Collaboration.

3.3 Measurement

So far, I have suggested that both detections and observations are (at least sometimes) measurements. In astrophysics, where ordinary human perception alone is of very limited use, most if not all detection and observation takes the form of some kind of measurement. This suggests a fruitful way forward for analyzing LIGO's "direct detection" and "direct observation" claims: connecting these claims to recent work in the philosophy of metrology (the study of measurement). In what follows, I will briefly summarize some of this literature and apply it to the two key cases of interest: LIGO-Virgo and the Hulse-Taylor binaries.

One important feature of measurement is that it is representational: the outcome of a measurement is a symbolic (usually mathematical) representation of some feature of a physical state. A traditional view, the "representational theory of measurement," emphasizes this feature. In particular, proponents of the representational theory focus on the construction of measurement scales that relate *empirical* relational structures (e.g., rigid rods, along with their ordering and concatenation) and *numerical* relational structures (e.g., a set of numbers and specific mathematical relations between them)(Tal 2013, 1164).

In the last decade or so, a contemporary view of measurement has emerged in the work of authors such as van Fraassen (2008), Tal (2012, 2013, 2016, 2017), and Parker (2017), that emphasizes some of the more "applied" aspects of measurement neglected by the representational theory (e.g., error, uncertainty, and calibration). On this view:

> Measurement is a kind of empirical information-gathering activity, involving physical interaction with the entity measured, which locates the entity in a logical space. Especially in contemporary measurement, this locating activity often involves a form of model-based inference—an inference from the state(s) of one or more physical processes to the value(s) of one or more parameters thought to characterize the entity under study, where this inference is guided by a model of the measuring process. [. . .] Measurement outcomes, when complete, include not only a best-estimate value for a parameter but also a well-motivated uncertainty estimate (279)

According to this view, measurement is a procedure (partly physical interaction, partly model-based inference) through which we generate a selective representation of the state of a physical system (including an indication of our uncertainty).

Introducing some terminology is helpful here. Drawing on Parker (2017), Tal (2017), and Bokulich (2020), I take an "instrument indication" to be the physical state of an apparatus used in measuring, such as a pointer position or a digital display showing a numerical readout; a "raw instrument reading" (or "quantified indication") to be the preliminary assignment of some value to some parameter about which the apparatus is supposed to be informative (under favorable circumstances); and a "measurement outcome" (or "measurement result") to be a selective representation of the system under measurement, inferred from one or more instrument indications.

Parker (2017) goes on to classify measurements according to the layers of inference involved in going from *instrument indication* to *measurement outcome*—how "direct" this chain of inferences is. Since I have said that detections (often) take the form of measurements, I take Parker's taxonomy of measurements to provide a candidate for classifying detections (as direct or indirect) as well. I therefore devote some space here to considering her taxonomy. Parker describes three main kinds of measurement: direct, derived, and complex.

A direct measurement is one where the instrument indication is produced without explicit symbolic calculation and the "raw instrument reading" gives a preliminary value to the parameter being measured. This characterization does not rule out the use of some explicit calculations to correct for interfering factors or to estimate uncertainties, but it does rule out measurement outcomes that represent a different parameter than the one assigned a preliminary value by the raw instrument reading.

A derived measurement is one where there is at least one additional layer of inference. In a derived measurement, measurement outcomes are calculated or derived from a directly measured value for another parameter, based on some kind of reliable scientific principles or definitions. A full model of the measuring process thus includes everything as for a direct measurement—assumptions about physical interactions and data processing—plus assumptions about how the directly measured parameters relate to the derived parameter that is ultimately of interest.

Finally, a complex measurement is one where multiple direct and/or derived measurements are used together to generate a measurement outcome that is more informative than the direct and/or derived measurements used to generate it. This involves a further layer of inference, this time about how to combine information obtained through multiple measurements. Complex measurements take a wide range of forms, including (but not limited to): multiple measurements of the same parameter with different instruments; a set of measurements that serve as a sample from which to estimate an aggregate or population-level parameter; and measurements where structure is added to initial data to arrive at the data model.

Now we can consider how the LIGO-Virgo detections of gravitational waves fit into this framework. To do so, I will first provide a description of how these detections proceed.

The LIGO and Virgo interferometers undergo a physical interaction with gravitational waves. The passing gravitational wave effectively changes the length of the two perpendicular interferometer arms ($4km$ long for LIGO) by a very tiny amount (on the order of $10^{-18}m$). This length change is given by:

$$\Delta L(t) = \delta L_x - \delta L_y = h(t)L$$

where L_x and L_y are the lengths of each of the perpendicular arms (designated as the x and y axis respectively), and h is the gravitational-wave strain as projected onto the detector. The strain is measured by using a photodetector to register the relative change in phase of laser light sent along the two perpendicular arms. This works because the phase shift depends on the path difference of the rays and is thus probing $\delta L_x - \delta L_y$.[8]

Each interferometer produces gravitational-wave strain data as a time series, sampled more than 16,000 times per second. A sample of this data is provided in figure 1. This can be regarded as the raw instrument reading, although there is some data processing built into its production. For example, converting the light detected at the photo detector into strain data depends on calibration, so for some events there is a second version of the data based on updated calibration protocols (Veitch et al. 2015).[9] This

8. For a philosophical introduction to this process and the related physics, see Elder (2020, ch.1).

9. Note that since calibration cannot be done using real gravitational waves, the calibration process

```
# time (seconds)             strain * 1.e21
2.500000000000000000e-01 2.454791884395226415e-02
2.500610351562500000e-01 1.529268268197186628e-02
2.501220703125000000e-01 6.372337209458739903e-03
2.501831054687500000e-01 -2.075009545624776318e-03
2.502441406250000000e-01 -9.889093246500136117e-03
2.503051757812500000e-01 -1.690355466802163920e-02
2.503662109375000000e-01 -2.296401895143313290e-02
2.504272460937500000e-01 -2.794594009509657889e-02
2.504882812500000000e-01 -3.176983420095493005e-02
2.505493164062500000e-01 -3.441166524283004874e-02
2.506103515625000000e-01 -3.590663183802679514e-02
2.506713867187500000e-01 -3.634551559571452123e-02
2.507324218750000000e-01 -3.586396766973119071e-02
2.507934570312500000e-01 -3.462629599896537014e-02
2.508544921875000000e-01 -3.280620402422139426e-02
2.509155273437500000e-01 -3.056745208078989751e-02
2.509765625000000000e-01 -2.804750301187722283e-02
2.510375976562500000e-01 -2.534673932921384723e-02
2.510986328125000000e-01 -2.252482663905932408e-02
2.511596679687500000e-01 -1.960454591214734824e-02
2.512207031250000000e-01 -1.658223206938517699e-02
2.512817382812500000e-01 -1.344294408579713336e-02
2.513427734375000000e-01 -1.017772839163448920e-02
```

Figure 1: Selection of LIGO strain data provided by the Gravitational Wave Open Science Center. (Vallisneri et al. 2015).

data is then subjected to sophisticated data analysis procedures. Notably, the most effective search technique involves matched-filtering, a signal-processing technique in which correlations between the (noisy) data and template waveforms are sought. Comparison of the data to a library of approximately 250,000 templates leads to a data model and reconstructed waveform.[10] A confirmed detection requires coincident detection of (matching) signals across at least two detectors. Once the signal has been identified, a Bayesian inference package called LALInference is also used to infer information about the source parameters and underlying physics (Veitch et al. 2015). Numerical relativity simulations are also produced using parameters consistent with the estimated source parameters.

The data analysis procedures produce a number of representations of both the gravitational waves and the astrophysical objects that generated them. In this case, the most famous such representation is the one depicted in figure 2, depicting the strain amplitude over time for GW150914.

Importantly, this image comes along with some values for the "false alarm rate" and the statistical significance—indications of how seriously we ought to take the given image as a representation of a gravitational wave. These are calculated based on analyses of data recorded by LIGO-Virgo during a window of time surrounding the event. In the case of GW150914, this window was approximately 16 days. Using time-slides of 0.1s—that is, repeatedly shifting the Livingston data by 0.1s relative to the Hanford data and searching the offset data for coincident events—this 16 days of data is used to generate approximately 608,000 years worth of background data from which to calculate the false alarm rate for GW150914. As stated in Abbott et al. (2016a, 061102–7), this technique gave a false alarm rate of 1 event per 203 000 years, equivalent to a significance greater than 5.1σ. If we are thinking of the measurement of GW150914 as a detection of gravitational waves then we should take the measurement outcome to be one of the representations of the inferred waveform of the gravitational wave, along with the estimated statistical significance of the signal.

The LIGO-Virgo detection of GW150914 should be classified as a complex measurement within Parker's scheme. Recall that a complex measurement is one where multiple direct and/or derived measurements are used together to generate a measurement outcome. In the case of LIGO-Virgo, no measurement outcome is possible without (at least) *two* interferometers, since the search for gravitational waves is a search

involves modeling the interferometer's response to gravitational waves using lasers to move the interferometer's test masses (mirrors in each arm). Errors in the calibration model must later be modeled as part of the parameter estimation process—the data analysis step in which inferences are made about the astrophysical source of the gravitational waves.

10. I discuss the use of matched filtering by the LIGO-Virgo Collaboration in greater detail in Elder (2020, 82–7). See also Patton (2020).

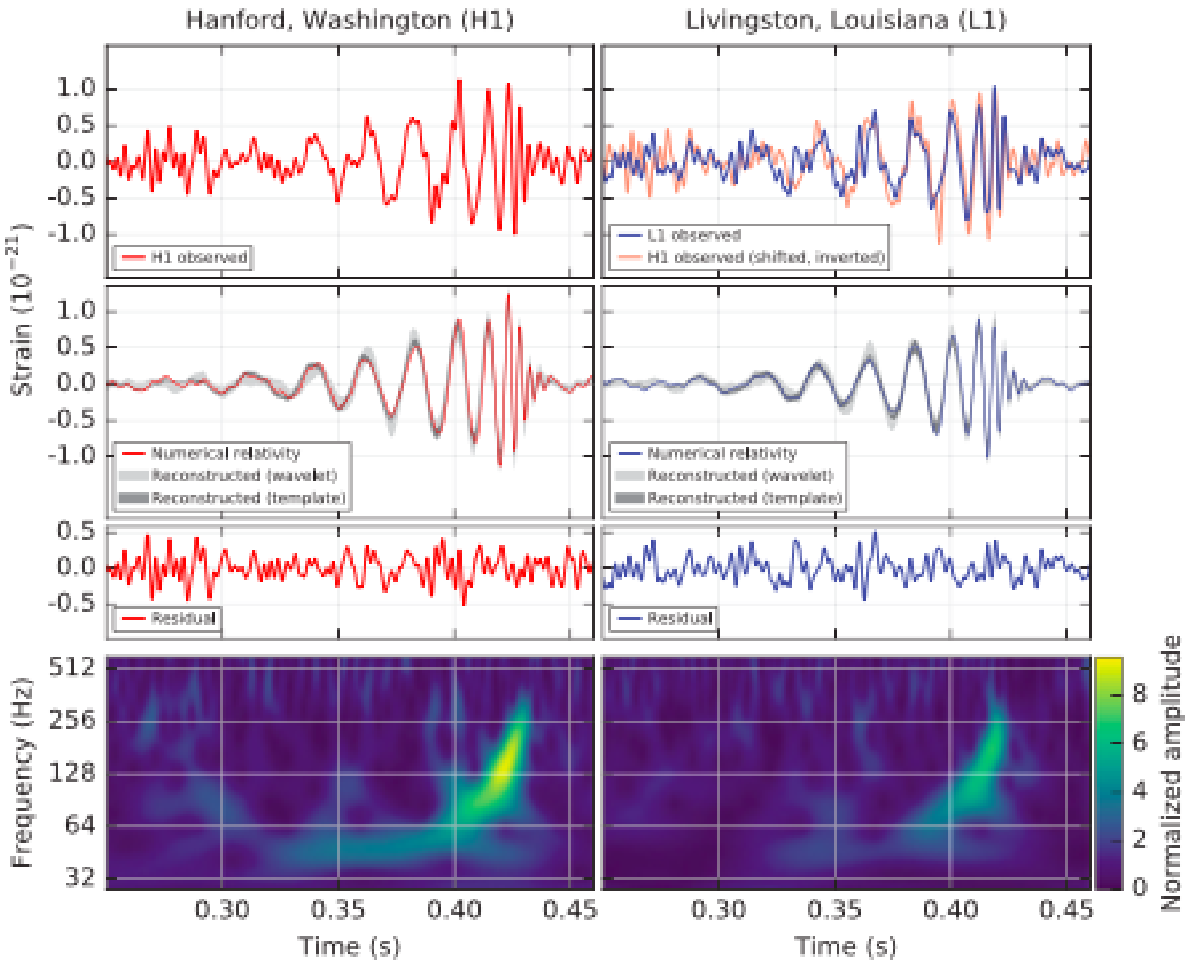

Figure 2: Image reproduced from Abbott et al. (2016a, 061102–2). Original caption: "The gravitational-wave event GW150914 observed by the LIGO Hanford (H1, left column panels) and Livingston (L1, right column panels) detectors. [...] Top row, left: H1 strain. Top row, right: L1 strain. [...] Second row: Gravitational-wave strain projected onto each detector in the 35–350 Hz band. Solid lines show a numerical relativity waveform for a system with parameters consistent with those recovered from GW150914 confirmed to 99.9% by an independent calculation based on [Campanelli et al. (2006)]. Shaded areas show 90% credible regions for two independent waveform reconstructions. One (dark gray) models the signal using binary black hole template waveforms. The other (light gray) does not use an astrophysical model, but instead calculates the strain signal as a linear combination of sine-Gaussian wavelets. These reconstructions have a 94% overlap, as shown in [Abbott et al. (2016b)]. Third row: Residuals after subtracting the filtered numerical relativity waveform from the filtered detector time series. Bottom row: A time-frequency representation of the strain data, showing the signal frequency increasing over time."

for *coincident signals* in both detectors. In fact, if we take the parameter of interest to be the strain, then it begins to look like we need not just one measurement of the parameter of interest but a series of measurements over the time during which the gravitational waves were passing through. Each detector produces its own strain data for this time period—the raw instrument readings—and joint analysis of both datasets is what is used to produce a measurement outcome. Indeed, 16 days worth of measurements were required in order to generate the measurement outcome, given the method used to generate background data and hence a "well-motivated uncertainty estimate." The detection of GW150914 is certainly a complex measurement because it took two detectors and 16 days worth of measurements in each in order to produce the measurement outcome: the reconstructed waveform of the gravitational wave together with the estimated uncertainty associated with the event.

What about the Hulse-Taylor-Weisberg detection? This also qualifies as a complex measurement according to Parker's classification, since the detection of orbital decay, and hence gravitational waves, was based on roughly 1,500 measurements of pulse arrival times. Thus it seems that the binary pulsar detection and the LIGO-Virgo detection are classified the same way (and for similar reasons) in Parker's scheme. On the surface, then, it might seem that this classification of measurements sheds little light on the difference between these detections, and on the supposed distinction between direct and indirect detections.

Although I do not think that Parker's classification system solves the puzzle, I do think that it contains an important insight that resonates with what the scientists themselves have to say about the distinction. Thus, in the next section, I propose a way of distinguishing between direct and indirect detection that draws on the distinction Parker makes between direct and derived measurement.

4 The Direct Detection of Gravitational Waves

In the remainder of this paper, I will analyze the significance of adding the terms "direct" and "indirect" to detection/observation claims. This task is complicated by the fact that that "detection" and "observation" are often used interchangeably within a field, and inconsistently across fields. The implications of adding the term "direct," similarly seem to vary by field (or sub-field).

In particle physics, for example, the term "observation" is sometimes used simply to to indicate a level of confidence in the results. Thus, according to Staley (2004), "observation of the top quark" indicates a confident existence claim, while "evidence for the top quark" is a more cautious claim that such a hypothesis is supported by the data. On this use of the term, "observation" indicates nothing about the methods of detection used. It is a marker of the quality (but not the nature) of the evidence. Accounts of astrophysical observation, such as Shapere's, *might* be applicable in such contexts (perhaps with modifications) but such an application is unlikely to track the use of the term reported by Staley. The point here is that different epistemic communities use terms in different ways and follow different epistemic norms.

My focus will be less on trying to accurately reconstruct a single account of how terms such as "direct observation" are used across contexts (indeed, I suspect that this is an impossible task) but rather to assess the potential epistemic significance of adding terms like "direct" and "indirect" in different situations—especially the cases at hand: the detection of gravitational waves and the observation of binary black hole mergers. These correspond to two different roles that the LIGO-Virgo interferometers play: on the one hand, a complex experimental system for the detection of a new phenomenon (gravitational waves) and, on the other hand, a new kind of "receptor" for observing distant astrophysical events (compact binary mergers). The interferometers can thus be viewed through complementary lenses as analogous to either a particle accelerator or a telescope, depending on which goal we have in mind.

4.1 Franklin's Account of Direct Observation

The intuitive distinction between direct and indirect implicit in the contrast between the LIGO-Virgo and the Hulse-Taylor-Weisberg detections concerns the role that gravitational waves play in the measurement process. In the former case, but not the latter case, gravitational waves are directly interacting with the experimental system. Thus in the LIGO-Virgo case, it seems, we are detecting the gravitational waves themselves, while in the Hulse-Taylor-Weisberg case we are detecting a downstream effect: electromagnetic waves from which the existence of gravitational waves is inferred. I think that this intuitive picture is on the right track. However, it is in need of philosophical refinement. Before giving my own positive view of the needed refinement, I will first take a short detour through Franklin's attempt at clarifying the situation, in part to illustrate why further clarification is necessary.

In his mammoth paper "Is Seeing Believing?: Observation in Physics" (2017), Franklin is ostensibly aiming to give an account of what makes an observation direct or indirect in physics. This is carried out by detailed examination of a number of case studies, from Galileo's observation of the Jovian moons to the CMS detection of the Higgs boson and the LIGO-Virgo detections of gravitational waves. These rich case studies offer great insight into the variety of "observations" deemed direct or indirect by different scientific communities. However, while Franklin claims to handle them all using Shapere's account of direct observation, it is difficult to discern any single consistent standard that is being applied across all of the case studies.

At the outset, Franklin outlines Shapere's account, and emphasizes the provision that the transmission must involve minimal interference between its emission at the source and eventual arrival at the detector. This, after all, is the difference that Shapere cites between solar neutrinos and solar photons. The contrast that Shapere envisages here is not between direct and indirect observation, but rather between (direct) observation and inferences based on observation:

> Although claims about the center of the sun based on photons, like those based on neutrinos, are "based on observation", the sense in which they are so based is not that of being "directly", but only "inferentially" so based. But since the operative contrast here is between "observational" and "inferential", the term 'direct' in 'direct observation' has the function only of emphasizing that conclusions about the source are arrived at by observation, and not by inference based on observation. (That is, the idea of "indirect observation" plays no role at all.) (Shapere 1982, 512)

Later in the paper, Franklin points to there being "degrees of directness, depending on the length of the inference chain linking the experimental results to the conclusion" (Franklin 2017, 380). While this is in some ways similar to the picture Shapere describes, I think that basing directness on the number of inferences required to justify the claim is a mistake. First of all, it is not clear that there is a determinate number of inferences required to justify any scientific claim. The number of inferences involved will always depend on what premises one allows oneself to take as brute facts, and the ways that one is willing to model the systems involved (the level of abstraction, idealization, or simplification that are considered appropriate for the purposes of a particular inference). As such, I am not convinced that the standard can be objectively applied to substantially different cases, relying on different kinds of premises and inferences. In other words, providing a measure on the length of an inference chain, and hence on 'directness' so-conceived, looks to be problematically subjective, except perhaps in cases where one inference chain is a proper subset of the other.[11]

11. My objection to Franklin here might also be taken as an objection to Parker's taxonomy, given that both accounts concern the number of inferences implicated in the measurement or observation. However, I do not think that the objection works against Parker. The reason is that Parker is concerned with whether or not there are *any* additional inferences required going from the raw instrument reading to the measurement outcome. This doesn't involve trying to count inferences in the way that I take to be problematic in Franklin. However, I do think that a version of this objection applies to Parker's account (and also to my own account of direct/indirect detection). This involves re-modeling the measurement process so that different inferences are/are not incorporated into the model of the measuring instrument, thus changing the status of the raw instrument reading. I address this objection below (in Section 4.2 with respect to my own account of direct/indirect detection.

Furthermore, it seems that this criterion risks getting the wrong answer in the LIGO-Virgo case. There is a sense in which inferring the existence of gravitational waves from the Hulse-Taylor binary is more straightforward than the inference using the LIGO-Virgo interferometers. The LIGO-Virgo detection relies on new, incredibly complex equipment and a detailed understanding of how its components will interact with passing gravitational waves. This even includes the recording of thousands of auxiliary channels to monitor various sources of noise. It seems likely that the layers of inference necessary to claim a gravitational wave detection are at least as numerous as in the Hulse-Taylor-Weisberg case, given the number of assumptions about the instrument itself that must be included in the inference chain.

It is possible that what Franklin has in mind is, inspired by Shapere, counting the inferences that go beyond the observation itself. Shapere's comments on the distinction between the "observation" and the "inferential" suggest that only further inferences about the source get counted as "inferential", while the rest of the chain of inferences are neatly packaged together as the "observation". In this case, it is hard to know what to say about the LIGO-Virgo case, since Shapere's account doesn't have a natural interpretation in the absence of a separation between the observed entity (gravitational waves) and the receptor (the interferometer). For example, information about the gravitational waves might be seen as trivially transmitted without interference to the interferometer (thus counting as direct observation). Nonetheless, the signal is subject to significant interference within the interferometer. This points to a general problem in applying Shapere's account of observation too widely. Shapere himself thought that his view would require modifications when applied to particle physics, where the distinction between the "theory of the source", "theory of the transmission" and "theory of the receptor" no longer stands up (Shapere 1982, 512). Without further clarification from Franklin on what modifications he is assuming, it is difficult to properly evaluate this version of the "chain of inference" approach to directness.

Franklin applies a third and remarkably different distinction between direct and indirect observation in the context of particle physics. Here, he insists that directness has to do with the dispensability of statistical arguments in making a detection claim. For example, the CMS detection of the standard model Higgs particle is deemed to be indirect for the following reason:

> It seems clear that the discovery of the Higgs boson was not a direct observation. What was observed was a statistically significant signal above background at a definite mass. [. . .] The recorded event was definitely not "golden." Statistical arguments were both needed and provided. (Franklin 2017, 394)

It seems that, according to Franklin, an observation cannot be "direct" if it relies (too heavily) on statistical arguments. It is not enough, according to Franklin, to observe a "statistically significant signal above background", given that this signal *could* have been produced by background processes.

This statistics-based way of spelling out the distinction between direct and indirect observation has nothing to do with the other ways Franklin discusses. Furthermore, applying this standard to the LIGO-Virgo case risks committing him to calling it an indirect observation. This is because there are strong parallels between the statistical arguments made for the Higgs detection and the arguments in favor of GW150914 being a genuine detection. To understand why, it is sufficient to consider the use of time slides in determining the statistical significance of the gravitational wave signal.

Time slides are used to establish background data against which the purported signal can be judged to be significant. As previously mentioned, this process involves repeatedly shifting the Livingston data by 0.1s relative to the Hanford data and searching the offset data for coincident events. For GW150914, this was done for around 16 days worth of data during which the interferometers were judged to be operating in approximately the same state as they were during the detections.[12] Repeating this procedure over and over produced around 608,000 years worth of data. Searching for coincident events in all of this time-shifted data is supposed to provide a background of events that would be generated by chance fluctuations in the two interferometers. The upshot is that the statistical reasoning required to make the LIGO-Virgo Collaboration's discovery claim is strikingly similar to that which Franklin points to in the CMS case as the reason it is *not* a direct detection. In the LIGO-Virgo case as in the CMS case, such statistical arguments were used and indeed *required*.

Finally, when distinguishing between the LIGO-Virgo and Hulse-Taylor-Weisberg cases, Franklin falls back on something closer to the the intuitive, common-sense difference between them:

> In terms of Shapere's discussion of "direct observation," we see that the LIGO-VIRGO [*sic*] collaboration did make a direct observation. The gravity waves interacted with the two interferometers. The fact that there was complex instrumentation and analysis should not change that conclusion. It seems fair to say that the binary pulsar observation was indirect. The existence of gravity waves was inferred from the decrease in period of the pulsar, which was transmitted to the detector by electromagnetic radiation. As discussed below, I do not believe that the direct-indirect distinction had any epistemological significance. (Franklin 2017, 410)

12. This is perhaps a little misleading, because the states of the interferometer systems are constantly changing and require regular adjustments. For some discussion of what is really meant by "the same state" here, see Collins (2017).

Here, we return to the claim that the LIGO-Virgo detection was direct simply because gravitational waves were interacting with the interferometer. However, applying this standard back to the Higgs detection would seem to imply that this detection was *direct*, since this particle did interact with the experimental system. If, as instructed here with respect to gravitational waves, we *ignore* the role of complex instrumentation and analysis, it is hard to see why we should consider the Higgs detection indirect.

Franklin's case studies demonstrate the messiness of the direct/indirect distinction, and illustrate how the content of this distinction changes both over time and between communities and contexts. I do not think it is unreasonable to apply different standards of "directness" to case studies drawn from different eras and fields—in this sense I do not think that there is significant disagreement between my views and Franklin's. However, I do think that it is misleading to suggest that these different standards are based on Shapere (1982)'s account of "direct observation".

Although Franklin fails to provide a common, unifying analysis of what this distinction is supposed to be across his case studies, I am inclined to say that there was no success to be had in such a task. Indeed, I suspect that part of the confusion within the LIGO-Virgo Collaboration over whether to call their detection "direct" stems from the fact that it wasn't clear which scientific community the norms should be drawn from. When it comes to observing compact binaries, the goals of the LIGO-Virgo Collaboration are clearly astrophysical. However, the detection of gravitational waves itself is not distinctly astrophysical, in the sense that the target phenomena—the detected gravitational waves—are, at least temporarily, actually present in the terrestrial detectors. Here, the analogy with particle physics may be more apt (e.g., in terms of the carefully constructed device designed to detect a very tiny effect within it). If so, it may be fruitful (as Franklin suggests) to consider the LIGO-Virgo detection by the lights of what Peter Galison (1997) calls the "visual" or "image" tradition and the "electronic" or "logic" tradition (and later, the hybrid of these). However, I leave further exploration of these analogies between gravitational wave detection and particle physics as a subject for future work.

4.2 Refining the Intuitive Picture

In the previous section I said that the intuitive picture of the difference between the LIGO-Virgo and Hulse-Taylor-Weisberg detections was to do with the fact that gravitational waves interact with the measuring device in the former case but not the latter. However, I think that this picture is in need of clarification and refinement. I take up that task now and offer my own positive view of the distinction between direct and indirect detection in the context of detections of gravitational waves.

Unlike Franklin's account, my refinement is based on an analysis of measurement rather than observation. One reason for this is that I take Shapere's notion of observation (which Franklin appeals to) to be most readily applicable to astronomical observations. It is most naturally applied to LIGO-Virgo in the context of the observation of compact binary mergers, rather than the detection of gravitational waves. Additionally, I agree with Boyd (2018a, 2018b) that focusing on what is observable (or directly observable) through philosophical accounts of observation probably isn't the most perspicuous way of getting at the epistemically interesting features of empirical science. In contrast, recent work in the philosophy measurement (and philosophy of data, e.g., Boyd (2018a), Leonelli (2016), and Bokulich and Parker (2021)) seems to be better-suited to the task.

To begin refining the intuitive picture, it should be noted that it simply is not true (we assume) that gravitational waves are only interacting with the experimental system in the LIGO-Virgo case. Gravitational waves, like electromagnetic waves, travel at the speed of light. This means that the gravitational waves emitted by the Hulse-Taylor binaries interact with the radio telescopes in essentially the same way that they interact with the LIGO interferometers. The difference between the two cases cannot be the proximity of the waves to the detector, or even an interaction between the two. Instead, what matters is the nature of the detector itself and the explanation for why the collected data is informative about gravitational waves.

In the LIGO-Virgo case, the interferometer detects gravitational waves. These waves are interacting with it such that the recorded data accurately reflects the properties of the gravitational waves. In the Hulse-Taylor-Weisberg case, the radio telescopes detect electromagnetic radiation from the binary pulsar and from this we infer something about the distant system and the gravitational waves it emits. Although there are, presumably, gravitational waves interacting with this detector too, this interaction is not reflected in the data, which are data about the timing of electromagnetic pulses. Indeed, note that gravitational waves passing through the detector play no role in our explanation of how the detector is ultimately used to provide evidence for the existence of gravitational waves. A version of this point was made by one of the scientists within the LIGO-Virgo Collaboration. Paraphrasing the point made by a scientist, Collins writes:

> If something had got in the way of the gravitational waves on their way from the orbiting pulsar to Earth, and diverted them off somewhere else so they did not impact on the Earth, it would have made no difference to that inference, so long as the radio waves indicating the way the stars were moving still got through. But in the case of LIGO-Virgo, what is seen is the impact of the gravitational waves on an instrument built with the specific purpose of reacting to the waves and converting them into electrical signals so that they can be measured. And, as the emailer said, if something got in the way of those gravitational waves and diverted them off somewhere else, LIGO would not see them. (Collins 2017, 149)

The view expressed here can be understood as a counterfactual account of what constitutes *relevant* interaction between the detector and gravitational waves. That is, gravitational waves interact with the detector in a way that is relevant to the detection only if their absence would imply that no detection occurs. The interaction between the detector and the gravitational waves passing through it makes a difference to whether or not a detection occurs in the LIGO-Virgo case, but not in the Hulse-Taylor case. Thus the presence of gravitational waves in the detector system is relevant to the detection only in the former case. This counterfactual judgment is grounded in an understanding of the LIGO interferometers, which were "built with the specific purpose of reacting to the waves and converting them into electrical signals so that they can be measured". Thus we see that the core difference between the two cases doesn't concern the presence or absence of gravitational waves *per se*, but rather in the design of the detectors and, more abstractly, our model of how the detector works.

That is, we can understand the difference as being that the LIGO interferometers are understood to be *detectors of gravitational waves* while the radio telescopes used by Hulse and Taylor are understood to be *detectors of electromagnetic waves*, which are then used to make inferences about gravitational waves in a distant target system.

The core of my proposal is to employ Parker's distinction between direct and derived measurements in distinguishing between the modeling of these two detections. Recall that, for Parker, the key difference between a direct and a derived measurement is whether or not the raw instrument reading gives a preliminary value to the parameter being measured. Despite the complexity of the subsequent inferences mapping raw instrument readings to measurement outcomes and detection claims (in both cases), I think that there is a clear difference between the LIGO-Virgo and Hulse-Taylor-Weisberg detections at the level of the raw instrument reading. That is, the individual measurements made by the LIGO interferometers are understood to be direct measurements of the strain induced by the gravitational wave, while the individual measurements in the Hulse-Taylor-Weisberg case are not. Instead, these data indicate the arrival times of electromagnetic pulses from a binary pulsar system. Thus with Parker's distinctions in hand, we can spell out the difference between a *detector of gravitational waves* and an instrument that measures radio waves in order to infer the existence of gravitational waves. Thus the difference between the two detections lies in the modeling of the respective measurement processes.

However, there is an obvious response to this analysis: Why not model the situation differently? We are presumably free to model complex measuring situations in a range of different ways. This freedom leads to the worry that re-description of the situation will lead to different conclusions about whether a detection is direct or indirect. In this case, we might be concerned that simply saying that the LIGO (and Virgo) interferometers are gravitational wave detectors while the radio telescopes are electromagnetic wave detectors both begs the question (in simply *stipulating* that only the former detects gravitational waves *themselves*) and obscures the fundamental similarity between the two cases.

For example, what if we simply count the binary system as part of the detector in the Hulse-Taylor-Weisberg case? Then, we can describe the two detections in very similar ways:

> **LIGO-Virgo:** A binary system emits gravitational waves as its orbit decays. Through physical interaction with the detector system (the interferometer), information about these waves is encoded in electromagnetic pulses that are recorded by a photodetector and used to make inferences about gravitational waves in their causal past.
>
> **Hulse-Taylor-Weisberg:** A binary system emits gravitational waves as its orbit decays. Information about this effect is encoded in electomagnetic pulses. These are transmitted from one part of the detector system to another (from the binary to the telescope), recorded by the telescope and used to make inferences about gravitational waves in their causal past.

Looked at from this perspective, it does look like the two detections are fundamentally similar. Both involve detecting gravitational waves through their interaction with test masses, and in both cases this information is finally recorded via electromagnetic radiation.

It is tempting to dismiss my re-description of the two events as forced. One reason is relative scale; given the vast distance between the telescope and the binary pulsar, we might be tempted to say that the Hulse-Taylor binary cannot reasonably be modeled as the same system. However, I do not think we should be satisfied with this response. After all, the individual LIGO interferometers are themselves far apart, and the parts of the planned LISA interferometer will be further still. These distances are tiny compared to the distance from Earth to the Hulse-Taylor binary, but it is not clear what principled reason we could have for distinguishing based on this scale difference alone. I am not willing to dismiss the re-description on this basis. Indeed, I take this re-description seriously as pointing to some subtle challenges in distinguishing between the two detections in a philosophically rigorous manner. Furthermore, I think that the objection can be applied quite generally, simply by changing which inferences are built into the system that we group together as the measuring instrument, and which are left as explicit additional inferences. I take this objection to apply to both my account of direct/indirect detection and to Parker's taxonomy as described above. Given the strong connections between our accounts, defending my account can also be taken as at least a limited defense of the distinction between direct and derived measurements more broadly from objections of this kind.

The key difference between these two measurements is in how we model the detector and how we justify our confidence in its outputs as representative of the parameter of interest in the target system. These systems can (as I have just shown) be described in various ways. However, the choice to classify the radio telescope as a detector, while denying that the Hulse-Taylor binary is even a part of a detector, is not arbitrary. To see why, the key is to recognize how we come to consider an instrument as an appropriate measuring device of a particular kind. We usually do so on the basis of a range of interventions and tests (e.g., coherence tests, (Bokulich 2020)) to ensure that we understand the behavior of the instrument under a range of conditions. For example, in the case of LIGO-Virgo, the introduction of a "blind injection" (a fake signal) using lasers offered reason to think that the detection procedures worked. This followed extensive calibration and testing of the interferometer and its subsystems. Similarly, the radio telescopes used in the Hulse-Taylor-Weisberg detections were built for the purpose of detecting radio waves and scientists have a good understanding of how they work. The successful use of this kind of telescope over many years allows us to be confident in the reliability of the data it provides us with. However, we cannot say the same about the Hulse-Taylor binary pulsar. Hulse and Taylor did not build this binary system and we cannot control it in any way. Thus, the way in which we go about justifying our confidence in models of this system is different than in cases where it is possible to perform controlled interventions.[13] To put this another way, the way we justify our confidence in the "theory of the source" and the "theory of the receptor" are generally very different.

Overall, the difference between the LIGO-Virgo and Hulse-Taylor-Weisberg detections can be summarized as follows. The LIGO-Virgo detection GW150914 is a *direct* detection, because the LIGO interferometers are *gravitational wave detectors*. The "raw instrument reading" is strain data, representing a feature of the entity being measured—the strain associated with a passing gravitational wave. In modeling this system, the interferometer can essentially be black boxed as a mapping from the target system to a selective representation of that system. In contrast, the Hulse-Taylor-Weisberg detection of gravitational waves is *indirect* because the radio telescopes used are *radio wave detectors* (i.e., detectors of a certain band of electromagnetic radiation). In modeling this system, the telescope can be black boxed, modeled abstractly in terms of its inputs and outputs, which pertain to electromagnetic radiation rather than gravitational radiation. While there may be an element of

13. Hacking (1982) goes further and suggests that we only really believe in something when we can successfully perform such interventions. I think that this goes too far and under-values the empirical investigations that have been undertaken in astronomy. However, Hacking's emphasis on the epistemic significance of intervention is clearly relevant to the views I express here concerning direct and indirect detection.

convention in the choice to treat the telescope, but not the Hulse-Taylor binary, as a detector, this choice is not arbitrary. Rather, it is based on the interventions (e.g., calibration and coherence testing) that we can perform to justify our confidence in treating its output as representative of some parameter of interest in the target system.

This way of looking at it seems to resonate with what at least some scientists see as the epistemic difference between LIGO and Hulse-Taylor-Weisberg. Recall the scientist who though that directness was a "red herring" but thought that there was an important difference between the two detections, saying: "The difference is that [...] we figured out how to build a sufficiently sensitive GW receiver and *since we built it, we know exactly how it works*" (Collins 2017, 148, emphasis in original). However, unlike the quoted scientist, I take this epistemic difference to track the difference between direct and indirect detection. It is precisely because we build and manipulate our measuring devices (even complex ones like LIGO-Virgo) that we are able to treat them simply as a detector of a certain kind. Treating the detector as a black box, we then consider the layers of inference required to go from the output data—the raw instrument reading—to the final measurement outcome. Even in complex cases involving multiple detectors and measurements, we call a detection "direct" when the data is data about the entity we are trying to detect—in the LIGO-Virgo case, strain data. In these cases, confidence in the detector combined with confidence in the data analysis methods, explains the confidence that scientists have in the detection. In contrast, we call a detection "indirect" when it is also necessary to rely on a model of a separate target system in order to justify our detection claim. In such cases, we still black box the detector, but subsequent inferences must rely on models of the target system in order to go from raw instrument reading to measurement outcome—we cannot rely on our confidence in our detector alone.

With this picture in mind, we can make sense of some of the conflicting views about the similarities and differences between the LIGO-Virgo and Hulse-Taylor-Weisberg detections. It is true that there are strong similarities between the two in terms of the physical interactions involved—at least on a certain level of abstraction. In both cases, we must learn about gravitational waves through electromagnetic radiation, which is what is ultimately recorded to obtain data. However, it would be a mistake to take this to imply that the kinds of inferences we are making about the entire causal chain are the same, or that we must justify them in the same way. In the LIGO-Virgo case, our basis for taking the electromagnetic data to be informative about the strain of the passing gravitational wave is that the detector has been carefully set up such that this is so. In the Hulse-Taylor-Weisberg case, the basis for taking the electromagnetic data to be informative about gravitational waves is instead based on models of a distant target system with which we cannot interact. In this case we take ourselves to have directly detected electromagnetic signals from a binary system, from which we infer the existence of (or indirectly detect) gravitational waves.

So what is the epistemic significance of a direct detection, as compared to an indirect detection? Certainly direct detections are not inherently better than indirect detections.[14] Either way, we are making model-based inferences, and the security of these inferences depends on how well-justified our models are, within the domain that we are using them. In cases where we are confident about our models of the target system, an indirect detection of some entity in that system might be much better justified than a direct detection where we are less confident about our model of the measuring apparatus.

Nonetheless, I have shown that there is an important epistemic dimension to the direct/indirect distinction I have described. A direct detection is a model-based inference based primarily on the model of the measuring instrument—a model that is generally validated through interventions to build, calibrate, and test the system it represents. In contrast, indirect detections rely on a model of a separate target system, so inferences from the data—specifically, the raw instrument reading—to the target phenomena rely on models of both the measuring system and a separate target system. In the case of detecting a phenomenon like gravitational waves, which sits near the limits of our empirical grasp, a direct detection may be particularly important, since indirect detections depend (to a greater extent) on theoretical descriptions of gravitational waves for which there is a lack of independent evidence—at least prior to the LIGO-Virgo direct detection.

4.3 Scope and Significance

So far, my analysis has focused on making sense of the distinction between the 'direct' detection performed by LIGO-Virgo and the earlier 'indirect' detection using the Hulse-Taylor binary. I believe that my account of the direct/indirect distinction in this paper takes a common intuition about this distinction—that gravitational waves are physically present and interacting with the detector in the former, but not the latter case—and makes it philosophically precise. In order to do so, my account locates the difference between the two at a higher level of abstraction: in the modeling of the measurement process, and the related explanations of how the interferometer data comes to be a partial representation of gravitational waves. One further question that remains is the following: what is the scope of the direct/indirect distinction I have described?

I have already said that my analysis cannot apply to all uses of the terms "direct detection" or "direct observation" since these terms have demonstrably been used in different and sometimes conflicting ways (Franklin 2017; Staley 2004; Shapere 1982). Indeed, in the same paper (even the same sentence) that the LIGO-Virgo Collaboration call GW150914 the first direct detection of gravitational waves, they describe it as the first direct observation of a binary black hole merger:

14. In this, I am in agreement with Franklin, who says that "it depends" (Franklin 2017, 410).

> A century after the fundamental predictions of Einstein and Schwarzschild, we report the first direct detection of gravitational waves and the first direct observation of a binary black hole system merging to form a single black hole. (Abbott et al. 2016a, 061102–1)

Whatever "direct" might mean in this context, it is clear that the LIGO-Virgo "direct observation" of a binary black hole merger is *not* a "direct detection" according to the analysis I have just provided. Just as the Hulse-Taylor-Weisberg detection was "indirect" because the inference relied on models of a separate target system, the use of the LIGO and Virgo interferometers to observe *the source of* GW150914 must count as an "indirect detection" of a binary black hole merger. Thus the epistemic benefits of being a direct detection as stated above do not apply; when we focus on the role of the LIGO-Virgo interferometers in observing compact binary mergers all of the relevant inferences must be justified using models of a distant target system. This need not render them epistemically suspect. However, the model-dependent observation of remote target systems provides some key epistemic challenges that the methods of astrophysics must overcome (Anderl 2016; Elder 2020).

Despite this, I expect that my analysis of direct detection in this paper extends beyond its original context. To motivate this expectation, I will briefly consider how the analysis can be applied to some related cases: two unsuccessful attempts to measure gravitational waves—by Joseph Weber, and by the BICEP experiment; and first observation of a black hole "shadow" by the Event Horizon Telescope Collaboration.

Joseph Weber (1968, 1969, 1970) claimed to have detected gravitational waves using aluminium cylinders ("Weber bars") suspended on steel wires. The idea was that a passing gravitational wave would cause the Weber bars to vibrate at their resonant frequency of approximately $1660Hz$, which would then be converted to an electrical signal. Weber claimed to see coincident signals in bars separated by a distance of $1,000km$ (Weber 1969). Weber's results were widely discredited by the scientific community; attempts to replicate the experiment failed, and Weber's calibration and data analysis methods were considered to be inadequate (Franklin 1994).[15] According to my analysis, Weber's measurements (if successful) would have counted as *direct* detections of gravitational waves. Weber's bars play a similar role in the detection that gravitational wave interferometers do in the LIGO-Virgo detection, in that they interact with gravitational waves and convert the gravitational wave signal to an electrical signal. The data (while noisy) is taken to be a partial representation of these passing gravitational waves. Weber's bars are (unsuccessful) gravitational wave detectors, and Weber's failure to detect gravitational waves was ultimately attributed to this instrument not being sensitive enough to detect realistic gravitational wave. His

15. For a different perspective on this case (to which Franklin (1994) is responding), see Collins (1985).

detection claim was discredited on the basis of his inadequate treatment of his own instrument and data, rather than on inadequate modeling of a separate target system.

The BICEP (Background Imaging of Cosmic Extragalactic Polarization) experiment aims to detect gravitational waves through observations of the cosmic microwave background (CMB)—highly isotropic radiation produced by the very early universe. One iteration of this program used the BICEP2 telescope at the south pole, which observed at 150GHz from 2010 to 2012. The goal of these observations was to measure the polarization pattern of the CMB, in order to detect signatures of primordial gravitational waves. Although this experiment was initially thought to have successfully detected gravitational waves, these signatures were later found to be attributable to dust.[16] According to my analysis, this experiment (if successful) would have provided another *indirect* detection of gravitational waves. This is easily seen by the fact that the detector—the BICEP2 telescope—is designed to detect a specific frequency band within the electromagnetic spectrum. Detection of gravitational waves relied on modeling of a distant target—the very early universe, along with other intermediate systems such as sources of gravitational lensing and dust. Given that the eventual retraction of the detection claim was due to the realization that the dust had been inadequately modeled, this example also points to the epistemic significance of the distinction I have laid out: in cases involving a remote target system, it can be fiendishly difficult to establish confidence in the model of the target.

The Event Horizon Telescope (EHT) is an experiment that uses Very Long Basline Interferometry (VLBI)—a technique that combines radio telescopes separated by large distances (long baselines) to perform observations as if they were components of a single, much larger, telescope. The resolution of a telescope increases with its size. Using an array of telescopes where the longest baselines approximate the diameter of the Earth, the EHT is able to achieve a resolution equivalent to a telescope the size of the Earth. More precisely, pairs of telescopes in the array sample the "visibility" (or "fringe visibility function"), which is the Fourier transform of the intensity distribution of the source. Inferring the appearance of the source (producing an image) involves sampling the visibilities and then using sophisticated imaging algorithms to determine the most likely source model whose Fourier transform matches observations. The EHT's sampling of the visibilities is very sparse, which means that these algorithms must do more work to make up for the lost information; in principle, a range of sources will be compatible with the sampled visibilities (The Event Horizon Telescope Collaboration et al. 2019b). Nonetheless, in 2019 the EHT Collaboration announced their successful imaging of the region surrounding a black hole—specifically, the supermassive black hole at the centre of the M87 galaxy—complete with a "shadow" and a bright crescent produced by

16. Keating (2018) provides a detailed account of this saga. See also Ade et al. (2014) for scientific details.

accreted matter (The Event Horizon Telescope Collaboration et al. 2019a). The EHT imaging presents a more complicated case for interpreting according to my analysis. On the one hand, the measured data is about visibilities. Turning these into a representation of the source takes sophisticated additional analysis. In this sense, it looks to be an indirect detection of the source. On the other hand, these imaging algorithms are sometimes described by scientists as "part of the telescope", and can be understood as part of the synthesis of an Earth-sized telescope from a small number of distributed components. While these algorithms build in some assumptions about the source being observed, they do not assume a theoretical model of the source object in the way that LIGO-Virgo parameter estimation does. It is only in the interpretation of the image that such modeling assumptions are needed. In this sense, imaging by the EHT may have more in common with the direct detection of gravitational waves by LIGO-Virgo than it initially seems. And indeed, we might be inclined to say that the visibilities themselves count as a representation of the source (though the eventual measurement outcome takes a very different form). For now, I leave the interpretation of this case to future work.[17]

Overall, I think that these cases show that my analysis provides a fruitful framework for thinking about the nature of different detections that emphasizes the justification for confidence (or lack thereof) in the detection. In doing so, I have also shown how recent work in philosophy of measurement can provide insight into debates and areas of confusion that are not explicitly framed in terms of measurement.

This paper provides a novel account of the distinction between direct and indirect detection. However, this account has strong connections to existing work on related issues.

First, my account bears a strong family resemblance to Parker's taxonomy of measurements. Both accounts make use of a distinction between direct and derived measurements based on the layers of inference needed to go from the raw instrument reading to the measurement outcome. Although my account is not strictly speaking an application or elaboration of Parker's (recall that the measurements discussed here are "complex" on her account) it shares a common core and is partly inspired by her account. This connection also suggests fruitful comparisons with other studies of derived measurements. For example, there are strong parallels between Ohnesorge (2022)'s excellent discussion of problems of coordination in geodesy and a circularity problem that arises in gravitational-wave astrophysics, due to LIGO-Virgo observations of black holes being derived measurements of inaccessible target systems (among other things) (Elder, forthcoming).

Second, the distinction that I have made has connections to work on the distinction between experiment and observation. While the nature, and even the existence of such a

17. In particular, (Skulberg, Elder, and Field, In Preparation) takes up the the question of what is meant by "direct" in the context of the EHT, from both historical and philosophical perspectives.

distinction is disputed (e.g., Malik (2017)) the difference has often been taken (following Hacking (1989)) to concern manipulation and control; experimenters are taken to create new phenomena via controlled interventions, while mere observation involves no such controlled creation of new phenomena. What I have said in this paper has connections to this in terms of the emphasis on the role of controlled interventions. However, I think that the case of the LIGO-Virgo gravitational wave detection is a problematic case for the experiment/observation distinction. This is because it has many of the hallmarks of an experiment—controlled interventions into the world that create new phenomena in the detector (e.g., the movement of test masses). Indeed, this detection stands in contrast to astronomical observations in epistemicially-significant ways, such that it has much in common with other large scale experiments such as those in high energy physics (e.g., the ATLAS experiment at the Large Hadron Collider). An account of experiment/observation such as Hacking's that classes LIGO-Virgo detections as observations and high energy physics detections as experiments obscures epistemically-salient similarities between these two large-scale epistemic activities.

Third, my distinction has some connections to work on the distinction between experiments and simulations. In particular, my account takes some inspiration from Winsberg (2009)'s account of this distinction, according to which simulations and experiments differ in terms of the arguments given for their external validity. For simulations, this is based on confidence in the available principles for dynamical modeling of the target system, while for experiments this is based on scientists taking the "object" (the experimental system) to be of the same kind of system to the target in relevant ways, such that it is expected to exhibit relevantly similar phenomena. My account of direct/indirect detection similarly locates the distinction to be at the more abstract level of the kind of argument that must be given for the validity of the results in each case.

5 Conclusion

The discovery paper for GW150914 claims that LIGO directly detected gravitational waves, but it doesn't tell us what a direct detection is. Drawing on both the controversy among LIGO scientists over these descriptions and on existing philosophical literature, I have provided an analysis of these terms that is both philosophically adequate and true to the use of these terms by the scientists themselves. This distinction, thus interpreted, also has some epistemic significance because it relates to the ways in which scientists justify their models of different parts of the measurement process.

Recent philosophical work on measurement has the potential to shed light on several related debates in philosophy of observation, experiment, and simulation. I take this paper to be one example of how such analysis can both clarify confusion over the terms being used by scientists and put us in a better position to understand the epistemic challenges associated with observations—such as BICEP2, LIGO-Virgo observations of black holes, etc.—that are "indirect" in the sense I have described.

References

Abbott, B. P., et al. 2016a. "Observation of Gravitational Waves from a Binary Black Hole Merger." *Physical Review Letters* 116 (6): 061102. https://doi.org/10.1103/PhysRevLett.116.061102.

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

Ade, P. A. R., et al. 2014. "Detection of B-Mode Polarization at Degree Angular Scales by BICEP2." *Physical Review Letters* 112 (24). https://doi.org/10.1103/physrevlett.112.241101.

Anderl, Sibylle. 2016. "Astronomy and Astrophysics." In *Oxford Handbook of Philosophy of Science,* edited by Paul Humphreys. Oxford University Press, April. https://doi.org/10.1093/oxfordhb/9780199368815.013.45.

Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *Philosophical Review* 97 (3): 303–52.

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

Bokulich, Alisa, and Wendy Parker. 2021. "Data models, representation and adequacy-for-purpose." *European journal for philosophy of science* (Netherlands) 11 (1): 31–31.

Boyd, Nora Mills. 2018a. "Evidence Enriched." *Philosophy of science* 85 (3): 403–421.

———. 2018b. "Scientific Progress at the Boundaries of Experience." PhD diss., University of Pittsburgh.

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.

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.

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

Elder, Jamee. Forthcoming. "Black Hole Coalescence: Observation and Model Validation." In *Working toward Solutions in Fluid Dynamics and Astrophysics: What the Equations Don't Say,* edited by Lydia Patton and Erik Curiel. Springer Briefs.

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/https://doi.org/10.1016/0039-3681(94)90062-0.

———. 2017. "Is Seeing Believing?: Observation in Physics." *Physics in Perspective* 19. https://doi.org/https://doi-org.proxy.library.nd.edu/10.1007/s00016-017-0210-y.

Galison, Peter Louis. 1997. *Image and logic: a material culture of microphysics.* Chicago: University of Chicago Press.

Hacking, Ian. 1982. "Experimentation and Scientific Realism." *Philosophical Topics* 13 (1): 71–87. https://doi.org/https://doi.org/10.5840/philtopics19821314.

———. 1989. "Extragalactic Reality: The Case of Gravitational Lensing." *Philosophy of Science* 56 (4): 555–581.

Hulse, R. A., and J. H. Taylor. 1975. "Discovery of a pulsar in a binary system." *Astrophysical Journal* 195:L51–L53. https://doi.org/10.1086/181708.

Keating, Brian. 2018. *Losing the Nobel Prize : a story of cosmology, ambition, and the perils of science's highest honor.* New York: W.W. Norton & Company.

Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study.* Chicago: University of Chicago Press.

Malik, Saira. 2017. "Observation Versus Experiment: An Adequate Framework for Analysing Scientific Experimentation?" [In eng]. *Journal for general philosophy of science* (Dordrecht) 48 (1): 71–95.

Ohnesorge, Miguel. 2022. "Pluralizing measurement: Physical geodesy's measurement problem and its resolution." *Studies in History and Philosophy of Science* 96:51–67. ISSN: 0039-3681. https://doi.org/https://doi.org/10.1016/j.shpsa.2022.08.011. https://www.sciencedirect.com/science/article/pii/S003936812200125X.

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.

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.

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.

Skulberg, Emilie, Jamee Elder, and Grace Field. In Preparation. "What is a 'direct' image of a shadow?: A history and epistemology of 'directness' in black hole imaging."

Staley, Kent W. 2004. *The evidence for the top quark : objectivity and bias in collaborative experimentation.* Cambridge, UK ; New York: Cambridge 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.

———. 2016. "Making Time: A Study in the Epistemology of Measurement." 67 (1): 297–335.

———. 2017. "Calibration: Modelling the measurement process." *Studies in history and philosophy of science. Part A* (England) 65-66:33–45.

Taylor, J. H., and J. M. Weisberg. 1982. "A new test of general relativity - Gravitational radiation and the binary pulsar PSR 1913+16." *Astrophysical Journal* 253:908–920. https://doi.org/10.1086/159690.

The Event Horizon Telescope Collaboration et al. 2019a. "First M87 event horizon telescope results. I. The shadow of the supermassive black hole." *Astrophys. J. Lett* 875 (1): L1.

———. 2019b. "First M87 event horizon telescope results. IV. Imaging the central supermassive black hole." *The Astrophysical Journal Letters* 875 (1): L4.

Vallisneri, Michele, Jonah Kanner, Roy Williams, Alan Weinstein, and Branson Stephens. 2015. "The LIGO Open Science Center." *J. Phys. Conf. Ser.* 610 (1): 012021. https://doi.org/10.1088/1742-6596/610/1/012021. arXiv: 1410.4839 `[gr-qc]`.

van Fraassen, Bas C. 1980. *The Scientific Image.* Oxford University Press.

van Fraassen, Bas C. 2008. *Scientific Representation : Paradoxes of Perspective.* Oxford; New York: Clarendon Press ; Oxford University Press.

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.

Weber, J. 1968. “Gravitational-Wave-Detector Events.” *Phys. Rev. Lett.* 20 (23): 1307–1308. https://doi.org/10.1103/PhysRevLett.20.1307. https://link.aps.org/doi/10.1103/PhysRevLett.20.1307.

———. 1969. “Evidence for Discovery of Gravitational Radiation.” *Phys. Rev. Lett.* 22 (24): 1320–1324. https://doi.org/10.1103/PhysRevLett.22.1320. https://link.aps.org/doi/10.1103/PhysRevLett.22.1320.

———. 1970. “Anisotropy and Polarization in the Gravitational-Radiation Experiments.” *Phys. Rev. Lett.* 25 (3): 180–184. https://doi.org/10.1103/PhysRevLett.25.180. https://link.aps.org/doi/10.1103/PhysRevLett.25.180.

Winsberg, Eric. 2009. “A Tale of Two Methods.” *Synthese (Dordrecht)* (Dordrecht) 169 (3): 575–592. ISSN: 0039-7857.