On the “Direct Detection” of Gravitational Waves

Jamee Elder

September 6, 2024

1 Introduction

In 2015 the Laser Interferometer Gravitational Wave Observatory (‘LIGO’), comprising 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. 2016, 061102–1).[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. Detecting this change requires sophisticated instrumentation, modeling, and data analysis techniques. Describing this complex detection process as “direct” is, on the surface, surprising. Second, a black hole’s event horizon is often thought to render it invisible. So it is surprising to see the LIGO-Virgo Collaboration claim to have observed astrophysical black holes, let alone *directly* observed them.

In this paper, I provide an account of direct (vs. indirect) detection in gravitational-wave astrophysics (the science of detecting gravitational waves and using them to observe the universe). In doing so, I highlight the epistemic considerations that lurk behind existing debates over the application of the term “direct”. My analysis elucidates 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 [reference omitted for peer review]this theory-ladenness, or model-dependence, turns out to be one of the key challenges of gravitational-wave astrophysics. A careful

---

[1] Virgo is a third observatory located in Italy. Together, the LIGO and Virgo observatories form a network of gravitational-wave interferometers used by the LIGO-Virgo Collaboration to detect gravitational waves.

characterization of the difference between direct and indirect detection thus has payoffs for understanding our empirical grasp on both gravitational waves and the astrophysical systems that produce them.

The paper proceeds as follows. In Section 2, I describe a controversy that arose within the LIGO Scientific Collaboration about whether or not to describe their detection of gravitational waves as "direct". In Section 3 I provide philosophical analyses of three (related) empirical activities: "detection," "observation," and "measurement." In Section 4 I first argue against a recent account of the direct/indirect distinction given by Allan Franklin (2017), then articulate my own positive account. I also discuss the scope and significance of the distinction, thus understood (in Section 4.3).

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.

Overall, this paper solves a puzzle about what counts as a "direct" detection of gravitational waves in a way that is true to scientific useage, and (more importantly) both philosophically precise and epistemically perspicuous. Having done so, this paper provides a foundation for a broader project of analyzing the epistemic situation of (gravitational-wave) astrophysics.

2 Controversy Within the LIGO Scientific Collaboration

As Harry Collins discusses in his book, *Gravity's Kiss* (2017), significant debate ensued among LIGO scientists about whether to describe GW150914 as a "direct" detection/observation of gravitational waves.

Calling the LIGO-Virgo detection a "direct detection" had one clear aim: distinguishing it from other methods of (indirect) detection, especially the previous gravitational-waves detection by Hulse, Taylor, and Weisberg (Hulse and Taylor 1975; Taylor and Weisberg 1982). For this detection, the decay rate of the orbital period of the Hulse-Taylor binary was found to be consistent with the predicted decay due to gravitational radiation. This rate was empirically determined from changes in the arrival times of the electromagnetic pulses from one of the binary's component stars. This was widely considered to be compelling evidence for the existence of gravitational waves, including by the LIGO-Virgo team:

> 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. 2016, 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 other attempts at "indirect" detection, such as the BICEP experiment.[2] This paper will 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 BICEP 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 detection, offending members of the astrophysics community. While this political dimension to the dispute is interesting, it doesn't tell us what of conceptual or epistemic significance is at stake, so my discussion will focus on the philosophical dimension.[3]

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. For example, Taylor (unsurprisingly) questions whether the LIGO-Virgo detection was any closer to being a "direct" detection than the Hulse-Taylor-Weisberg detection:

> 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

---

[2] See Section 4.3 for discussion of BICEP.

[3] Skulberg and Elder (In Preparation) discusses the social and political dimensions.

> 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." (Collins 2017, 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? (148)

However, the same scientist goes on to state what they take to be special about the LIGO detection:

> 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)

The quoted scientist did not take this difference to suggest a notion of directness. Indeed, they suggest 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 (Section 4) that this control *is* 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 to resolve the issue: a poll. 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.[4] (156)

---

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

This result was reflected in the published paper.

This presents an opportunity for philosophical intervention. A poll is a philosophically unsatisfactory resolution to a debate over the use of epistemically-loaded notions like direct detection and direct observation. We are left wondering how to understand the terminological choices that were made as well as their epistemic implications.

In what follows, I provide an account of the term “direct detection” that tracks the distinction being made by the LIGO-Virgo Collaboration in Abbott et al. (2016), while also considering the epistemic significance of this distinction. Another attempt at this task can be found in Franklin (2017). Franklin aims 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. With respect to the two gravitational-wave detections he says:

> [T]he LIGO-VIRGO 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. (410)

This is an intuitive, common-sense way to distinguish between the two detections. However it is in great need of philosophical clarification, which I will now provide.

3 Detection, Observation, and Measurement

The controversy over the use of the terms “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 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. I leave consideration of the second question—concerning the significance of directness—to Section 4.

3.1 Detection

When considering what a “direct detection” is, the first question to ask is: “what is a *detection*?” None of the participants in the debates I describe below explicitly state what they mean by detection, but Dudley Shapere (1982, 488) comes closest, saying that “detection” is (often) used interchangeably with “observation”. However, since “observation” is also a disputed term, this is somewhat unhelpful for present purposes. To the best of my knowledge, this relative neglect of the term “detection” is characteristic of existing philosophical literature, so I devote some space here to spelling out how I think the term should be understood.

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 establishes the existence of some entity (or phenomenon) within the target system. A successful detection gives 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.

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(a species of flightless parrot, native to Aotearoa New Zealand) on Hauturu (Little Barrier Island) 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 entity itself. For example, it might be sufficient for detection of a kea (another parrot from Aoteroa New Zealand, known for their intelligence and mischievous behaviour) to note certain characteristic signs of their having been there—olive-green feathers, bird droppings, a shredded tent, 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 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 locate 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 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."[5] 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 aware that gravitational waves are imperceptible. They are simply using the term "observation" in a way that is at odds with the strict sense advocated by some philosophers.

---

[5] It is worth noting that Bogen and Woodward (1988) explicitly say that they take no issue with scientists using the term "observe" in a different, more permissive way.

Toward the permissive 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).

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 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. I return to Shapere's view of (direct) observation in Section 4.1, in the context of Franklin (2017)'s analysis of direct observation.

3.3 Measurement

So far, I have suggested that both detections and observations are (at least sometimes) measurements. In astrophysics, where human perception alone is of very limited use, most 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., among rigid rods) and *numerical* relational structures (e.g., among sets of numbers)(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), Parker (2017), and Tal (2012, 2013, 2016, 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 (Parker 2017, 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 e.g.,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.[6]

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 $\delta L_x - \delta L_y$.[7]

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).[8] These data are 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.[9] 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

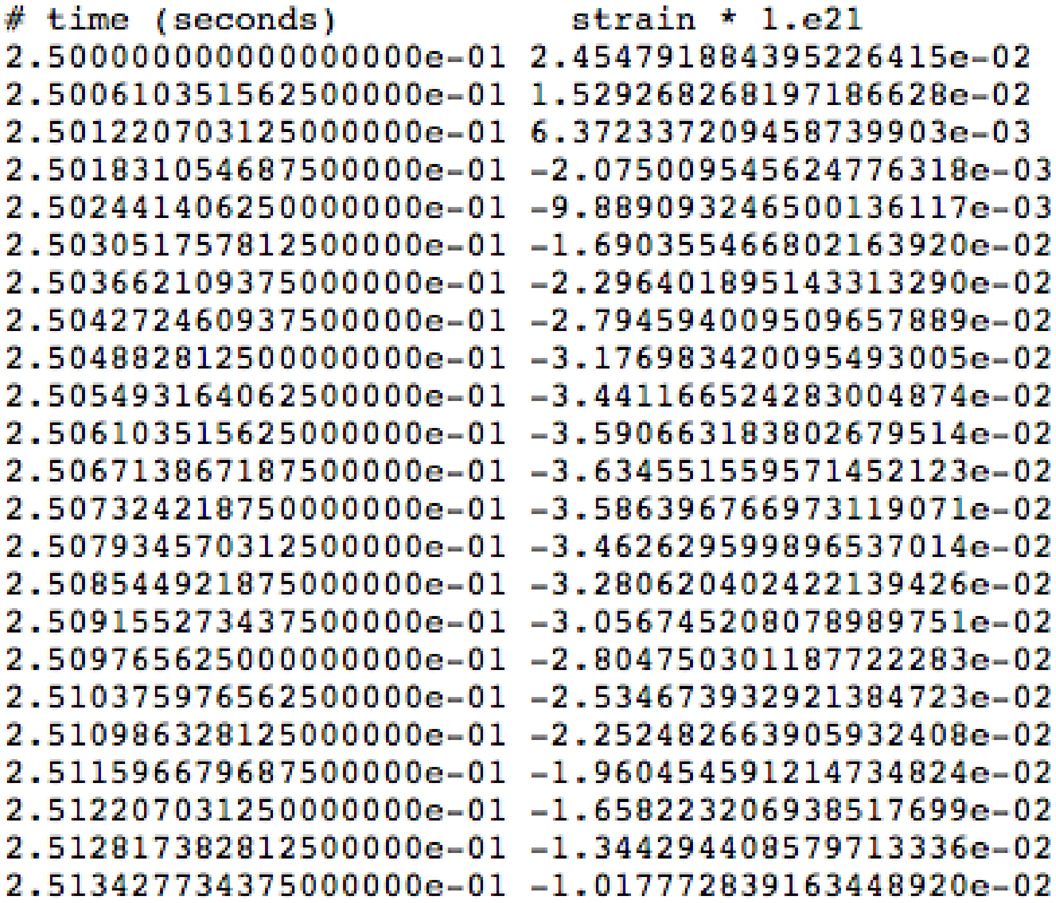
---

[6] Specifically, how they proceeded at the time of the first operating run, O1.

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

[8] Since calibration cannot be done using real gravitational waves, the calibration process involves modeling the interferometer's response to gravitational waves using lasers to move the interferometer's test masses (mirrors in each arm).

[9] I discuss matched filtering in greater detail in Elder (2020, 82–7). See also Patton (2020).



```
# time (seconds)          strain * 1.e21
2.50000000000000000e-01 2.45479188439522641 5e-02
```

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

most famous such representation is the one depicted in figure 2, depicting the strain amplitude over time for GW150914.

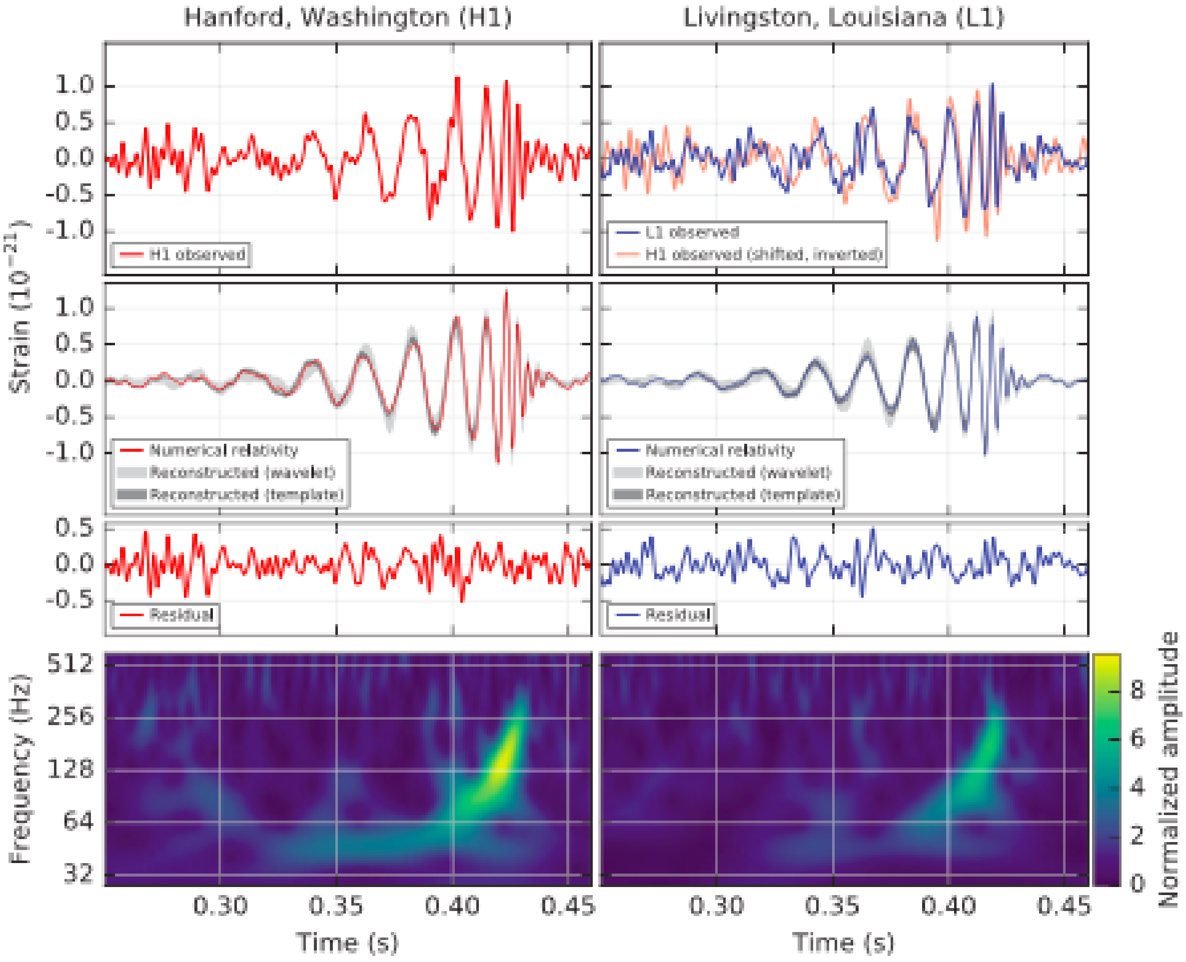


Figure 2: Image reproduced from Abbott et al. (2016, 061102–2).

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—the original 16 days of data produces a "new" 16-day dataset for each time slide. This provides a total of approximately 608,000 years worth of background data from which to calculate the false alarm rate for GW150914. The idea here is that the time slides preclude any genuine gravitational wave signal from being temporally correlated across the two data sets; if the waves were synchronized according to the genuine time stamps, they will no longer be so according to the shifted time stamps. Thus any event detected when searching the time-shifted data is assumed to be a coincidence—a fake event generated by various noise sources. Essentially, this procedure is supposed to simulate the rates at which background processes would produce events like GW150914 if we were able to perform the (impossible) task of keeping the interferometers running in the same state, within the same environment, for 608,000 years. As stated in Abbott et al. (2016, 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 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 —the raw instrument readings—and joint analyses of both datasets over a 16 day period were used to produce a measurement outcome.

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.

My aim is not 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"—especially the case at hand: the detection of gravitational waves.

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.

This intuitive picture—call it the "intuitive view"—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 detour through Franklin's attempt at clarifying the situation, in part to illustrate why further clarification is necessary.

Franklin (2017) sets out to use Shapere's account of direct observation to classify a range of case studies—including LIGO-Virgo, Hulse-Taylor-Weisberg, and the CMS detection of the Higgs boson—as direct or indirect observations. However, as I will show, Franklin fails to provide a common, unifying analysis of the direct/indirect distinction that applies across all of his case studies. Instead, his paper seems to make the distinction in at least four inequivalent ways, none of which successfully capture the distinction between LIGO-Virgo and Hulse-Taylor-Weisberg.

First, Franklin claims to be using Shapere's account (discussed in Section 3.2) as the basis for his analysis. In doing so, he begins by emphasizing Shapere's provision that the transmission must involve minimal interference between its emission at the source and eventual arrival at the detector. The contrast that Shapere envisages here is not between direct and indirect observation, but rather between (direct) observation and inferences based on observation:

> [S]ince 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)

However, Franklin uses Shapere's criterion as the basis for a distinction between direct and indirect observation—call this the "no interference" view.

The main problem with this no interference view as applied to LIGO-Virgo is that Shapere's account lacks a natural interpretation in the absence of a separation between the observed entity (gravitational waves) and the receptor (the interferometer). 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. It simply doesn't seem right that the direct/indirect distinction in the LIGO-Virgo case should be spelled out in terms of lack of interference. 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 (512).

A second, related, view—call it the "degrees of directness" view—arises elsewhere in the paper, when Franklin argues there are "degrees of directness, depending on the length of the inference chain linking the experimental results to the conclusion" (Franklin 2017, 380). This has some similarity to the picture Shapere describes, but recasts it from a binary distinction to a matter of degree. This view, equating the degree of directness to the number of inferences required to justify the claim, faces problems. The first problem is that 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). 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.[10]

The second problem with the degrees of directness view is that it risks getting the wrong answer in the LIGO-Virgo case. In some ways, 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 number of inferences 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.

A third view—call it the "dispensibility of statistics" view—emerges in Franklin's particle physics case study. 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. [...] Statistical arguments were both needed and provided. (Franklin 2017, 394)

Here, Franklin implies that 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 "dispensability of statistics" view is clearly inequivalent to any (binary or spectrum) version of Shapere's account. As such, it is impossible to interpret Franklin's paper as providing a universal account of the direct/indirect distinction across case studies.

---

[10] This objection might also be taken as an objection to Parker's taxonomy. Parker initially looks to dodge this objection by proposing a binary distinction between direct/derived based on whether there are *any* additional inferences required going from the raw instrument reading to the measurement outcome. However, I think that a version of this objection applies to Parker's account—and to my account of direct/indirect detection. This version involves re-modeling the measurement process so that different inferences are/are not incorporated into the model of the measuring instrument, thus changing what counts as the raw instrument reading. I address this objection in Section 4.2.

The "dispensability of statistics" view also isn't a good candidate for analyzing the LIGO-Virgo case, since doing so would commit Franklin to classifying the LIGO-Virgo detection as indirect. 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 LIGO-Livingston data by 0.1s relative to the LIGO-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.[11] Repeating this procedure over and over produced around 608,000 years worth of data. Searching for coincident events in all of these 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*.

The three preceding views that can be discerned in Franklin's discussion ("no interference", "degrees of directness", and "dispensability of statistics") are thus neither equivalent to one another, nor good candidates for distinguishing between the LIGO-Virgo and Hulse-Taylor-Weisberg cases. Franklin's discussion of gravitational wave detection instead reflects a fourth view, which falls back on something closer to the the "intuitive view":

> 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)

Here, we return to the claim that the LIGO-Virgo detection was direct simply because gravitational waves were interacting with the interferometer. In Section 4.2, I provide a philosophical refinement of this intuitive view.

---

[11] For discussion of what is really meant by "the same state" here, see Collins (2017).

Contra Franklin (2017), I think that his case studies illustrate how the content of the direct/indirect distinction *changes*, both over time and between communities and contexts. Nonetheless, it is also apparent that this distinction has been consistently used to convey different levels of confidence and to validate different kinds of measurement procedures across these many contexts. In other words, the nature of the distinction changes, but it is generally taken to have epistemic weight. For another interesting historical example, see Mitchell (2012)'s discussion of direct/indirect methods for measuring sound waves in the 19th Century. In this case, a direct/indirect distinction analogous to both the "intuitive view" and Parker's view was deployed to argue for the superiority of direct over indirect methods (456–8).

4.2 Refining the Intuitive Picture

The "intuitive view" of the difference between the LIGO-Virgo and Hulse-Taylor-Weisberg detections is based on the idea that gravitational waves interact with the measuring device in the former case but not the latter. However, 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.

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.

To begin refining the intuitive picture, note that it is technically *false* 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 arrive at the radio telescopes simultaneously with the detected radio pulses. They then interact with the radio telescopes in essentially the same way that they interact with the LIGO interferometers (periodic squeezing and stretching of distances according to the frequency of the waves). 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 are 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 reflect 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. (Collins 2017, 149)

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 (modeled as) *detectors of gravitational waves* while the radio telescopes used by Hulse and Taylor are (modeled as) *detectors of electromagnetic waves*, which are then used to make inferences about gravitational waves in a distant target system. In other words, the raw instrument reading gives a preliminary value for a parameter describing gravitational waves (strain) in the former case, but not the latter.

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 gravitational-wave strain, 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.

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 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.

I take this re-description seriously as pointing to some subtle challenges in distinguishing between the two detections in a philosophically rigorous manner. Furthermore, 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 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.[12] 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 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

---

[12] Hacking (1982) goes further and suggests that we *only* believe in the reality of an entity when we can perform such interventions. (I disagree).

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 are 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.

So what is the epistemic significance of a direct detection, as compared to an indirect detection? Direct detections are not inherently better than indirect detections.In cases where we are confident about our models of the target system, an indirect detection of some entity in that system may be better justified than a direct detection where we are less confident about our model of the detector.

Nonetheless, there is an important epistemic dimension to the direct/indirect distinction I have described, because the distinction points to the nature of the justification of the measurement outcome. In particular, labeling a detection indirect signals that modeling of a separate target system is implicated.

In the case of detecting novel phenomena like gravitational waves, which sit 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, I have given a novel account of the distinction between the ‘direct’ detection performed by LIGO-Virgo and the earlier ‘indirect’ detection using the Hulse-Taylor binary. My account 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. 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 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:

> 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. 2016, 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 2023, 2024).

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 the 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. These 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).[13] 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 to gravitational wave interferometers; they interact with gravitational waves and convert the gravitational wave signal to an electrical signal. Weber's failure to detect gravitational waves was ultimately attributed to this instrument not being sensitive enough to detect realistic gravitational waves. His 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 signatures of primordial 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. Initially, BICEP claimed to have detected gravitational waves but the apparent signal was later found to be attributable to galactic dust—microscopic interstellar matter that radiates microwaves.[14] If successful, BICEP would have provided another *indirect* detection of gravitational waves because the detector—the BICEP2 telescope—detects electromagnetic radiation in a specific frequency band. Detection of gravitational waves relied on modeling of a distant target—the very early universe, along with other intermediate systems such as dust and sources of gravitational lensing.

The Event Horizon Telescope (EHT) is an experiment that uses Very Long Baseline 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 large telescope. More precisely, pairs of telescopes in the array sample the "visibility" 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. In 2019 the EHT Collaboration announced their successful imaging of the region surrounding the supermassive black hole at the center of M87—featuring a bright ring (produced by accreting matter) enclosing a "shadow" region (The Event Horizon Telescope Collaboration et al. 2019). The EHT imaging presents a more complicated case for interpreting according to my analysis. On the one hand, the measured data are about visibilities. Turning these into an image 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

---

[13] Cf. Collins (1985).

[14] See Keating (2018) and Ade et al. (2014) for an account of the saga, and for scientific details (respectively).

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 (Doboszewski and Elder 2024). 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. For now, I leave the interpretation of this case to future work.[15]

Overall, these cases show that my account provides a fruitful framework for analyzing the nature of different detections, emphasizing the justification for confidence (or lack thereof) in the detection. This account has some strong connections to existing work on related issues, which are worth drawing attention to.

First, my account has connections 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.

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 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

---

[15] (Skulberg and Elder, 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.

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-Virgo directly detected gravitational waves, but it doesn't tell us what a direct detection is. In this paper, I have provided an analysis of "direct detection" that is both philosophically adequate and true to scientists' own usage in this context. In addition to the specifics of my account, this paper exemplifies how recent work in the philosophy of measurement can help clarify confusion over terms being used by scientists and, in doing so, put us in a better position to understand the distinct epistemic challenges that such terms are used to signal.

My novel account of the direct/indirect distinction takes a common (but technically false) intuition—that gravitational waves interact with the detector in the direct, but not the indirect case—and makes it philosophically precise. To do so, my account locates the difference between the two detections 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. In both cases, the detection takes the form of a model-based inference. The difference is that the indirect detection relies heavily on the model of a distant target system in order to justify the inference, while the direct detection does not.

The direct/indirect distinction, thus interpreted, has epistemic significance (even though direct detections are not inherently better). This is due to differences in the strategies available for validating models of a measuring instrument vs. models of a separate target system. While validation of the former can rely on interventions that establish the reliability of the instrument under controlled circumstances, such strategies are not available when the target system is physically distant and cannot be intervened upon—as is generally the case in astrophysics. For this reason, direct detection (in the sense that I have articulated in this paper) may be both particularly challenging and particularly valuable when detecting novel phenomena in astrophysics.

Acknowledgments

[Omitted for peer review]

References

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

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. https://doi.org/10.1093/oxfordhb/9780199368815.013.45.

Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *Philosophical Review* 97 (3): 303–52. https://doi.org/10.2307/2185445.

Bokulich, Alisa. 2020. "Calibration, Coherence, and Consilience in Radiometric Measures of Geologic Time." *Philosophy of Science* 87 (3): 425–456. https://doi.org/10.1086/708690.

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.

Doboszewski, Juliusz, and Jamee Elder. 2024. "Robustness and the Event Horizon Telescope: the case of the first image of M87*." *Philosophy of Physics.*

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

———. 2023. "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. https://doi.org/10.1007/978-3-031-25686-8.

———. 2024. "Independent evidence in multi-messenger astrophysics." *Studies in History and Philosophy of Science* 104:119–129. https://doi.org/10.1016/j.shpsa.2024.02.006.

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

———. 2017. “Is Seeing Believing?: Observation in Physics.” *Physics in Perspective* 19. https://doi.org/10.1007/s00016-017-0210-y.

Hacking, Ian. 1982. “Experimentation and Scientific Realism.” *Philosophical Topics* 13 (1): 71–87. 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.

Malik, Saira. 2017. “Observation Versus Experiment: An Adequate Framework for Analysing Scientific Experimentation?” *Journal for General Philosophy of Science* 48 (1): 71–95.

Mitchell, Daniel Jon. 2012. “Measurement in French Experimental Physics from Regnault to Lippmann. Rhetoric and Theoretical Practice.” *Annals of Science* 69 (4): 453–482. https://doi.org/10.1080/00033790.2012.730437.

Parker, Wendy S. 2017. “Computer Simulation, Measurement, and Data Assimilation.” *British Journal for the Philosophy of Science* 68 (1): 273–304. 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/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/10.1086/289075.

Skulberg, Emilie, and Jamee Elder. 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.” PhD diss., University of Toronto. http://search.proquest.com/docview/1346194511/.

———. 2013. “Old and New Problems in Philosophy of Measurement.” *Philosophy Compass* 8 (12): 1159–1173. 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* 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. 2019. “First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole.” *The Astrophysical Journal Letters* 875 (1): L1.

Vallisneri, Michele, Jonah Kanner, Roy Williams, Alan Weinstein, and Branson Stephens. 2015. “The LIGO Open Science Center.” *Journal of Physics: Conference Series* 610 (1): 012021. https://doi.org/10.1088/1742-6596/610/1/012021.

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

———. 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.” *Physical Review Letters* 20 (23): 1307–1308. https://doi.org/10.1103/PhysRevLett.20.1307.

———. 1969. “Evidence for Discovery of Gravitational Radiation.” *Physical Review Letters* 22 (24): 1320–1324. https://doi.org/10.1103/PhysRevLett.22.1320.

———. 1970. “Anisotropy and Polarization in the Gravitational-Radiation Experiments.” *Physical Review Letters* 25 (3): 180–184. https://doi.org/10.1103/PhysRevLett.25.180.

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