Studies in History and Philosophy of Science xxx (xxxx) xxx

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: http://www.elsevier.com/locate/shpsa

Coincidence and reproducibility in the EHT black hole experiment

Galina Weinstein [a,b]

[a] The Department of Philosophy, University of Haifa, 199 Aba Hushi Ave., Mount Carmel, 3498838, Haifa, Israel
[b] The Interdisciplinary Center (IDC), Kanfei Nesharim, Herzliya, 46150, Israel

ARTICLE INFO

Keywords:
Black hole
General relativity
Event horizon telescope
Computer simulation
Philosophy of experimentation

ABSTRACT

This paper discusses some philosophical aspects related to the recent publication of the experimental results of the 2017 black hole experiment, namely the first image of the supermassive black hole at the center of galaxy M87. In this paper I present a philosophical analysis of the 2017 Event Horizon Telescope (EHT) black hole experiment. I first present Hacking's philosophy of experimentation. Hacking gives his taxonomy of elements of laboratory science and distinguishes a list of elements. I show that the EHT experiment conforms to major elements from Hacking's list. I then describe with the help of Galison's Philosophy of the Shadow how the EHT Collaboration created the famous black hole image. Galison outlines three stages for the reconstruction of the black hole image: Socio-Epistemology, Mechanical Objectivity, after which there is an additional Socio-Epistemology stage. I subsequently present my own interpretation of the reconstruction of the black hole image and I discuss model fitting to data. I suggest that the main method used by the EHT Collaboration to assure trust in the results of the EHT experiment is what philosophers call the Argument from Coincidence. I show that using this method for the above purpose is problematic. I present two versions of the Argument from Coincidence: Hacking's Coincidence and Cartwright's Reproducibility by which I analyse the EHT experiment. The same estimation of the mass of the black hole is reproduced in four different procedures. The EHT Collaboration concludes: the value we have converged upon is robust. I analyse the mass measurements of the black hole with the help of Cartwright's notion of robustness. I show that the EHT Collaboration construe Coincidence/Reproducibility as Technological Agnosticism and I contrast this interpretation with van Fraassen's scientific agnosticism.

1. Introduction

In April 2019, an international collaboration of scientists presented Event Horizon Telescope (EHT) observations collected in April 2017 of the radio source at the center of the galaxy Messier 87 (M87). The galaxy M87 has long been known from theoretical models and computer simulations to host a bright, radio source at its center with a supermassive black hole hypothesized to power this radio source. The M87 black hole is considered one of the two largest supermassive black holes in the sky along with the Sagittarius (Sgr) A black hole, a black hole at the center of the Milky Way, near the border of the constellations Sagittarius and Scorpius. Over 200 scientists across the world contributed to the EHT project. Observations were made with sufficient sensitivity to reconstruct images that show a bright orange-yellow asymmetric ring surrounding a dark shadow (a "crescent"). In 2019 the EHT Collaboration published its findings in six scholarly letters (EHT Collaboration 2019a, henceforth EHTC).

In this paper, I present a philosophical analysis of the EHT black hole experiment. I refer mainly to the philosophy of experimentation of Ian Hacking (Hacking 1983, 1988, 1989, 1992), Bas van Fraassen (van Fraassen 1980, 1989, 2001), Nancy Cartwright (Cartwright, 1991) and Peter Galison (Galison, 2019).

The rest of the paper is organized as follows. In Chapter 2, I first present Hacking's philosophy of experimentation. Hacking gives his taxonomy of elements of laboratory science and distinguishes a list of elements. I use examples from the EHT experiment to illustrate that the EHT experiment conforms to major elements from Hacking's list.

I then describe with the help of Galison's *Philosophy of the Shadow* how the EHT Collaboration created the famous black hole image. Galison, a member of the EHT Collaboration, outlines three stages for the reconstruction of the black hole image: *Socio-Epistemology*, *Mechanical Objectivity*, after which there is an additional *Socio-Epistemology* stage. The *Socio-Epistemology* and *Mechanical Objectivity* stages form a cyclic loop (Galison, 2019).

In Chapter 3, I present my own interpretation of the reconstruction of the black hole image and I discuss model fitting to data. I suggest that the

E-mail address: galiweinstein.mc2@gmail.com.

https://doi.org/10.1016/j.shpsa.2020.09.007
Received 18 September 2020; Accepted 20 September 2020

main method used by the EHT Collaboration to assure trust in the results of the EHT experiment is what philosophers call the *Argument from Coincidence*. I show that using this method for the above purpose is problematic. I present two versions of the *Argument from Coincidence*: Hacking's *Coincidence* (Hacking 1981, 1983) and Cartwright's *Reproducibility* (Cartwright, 1991) by which I analyse the EHT experiment. The same estimation of the mass of the black hole is reproduced in four different procedures. The EHT Collaboration concludes: the value we have converged upon is robust. I analyse the mass measurements of the black hole with the help of Cartwright's notion of robustness (Cartwright, 1991). I show that the EHT Collaboration construe *Coincidence/Reproducibility* as *Technological Agnosticism*. I contrast this interpretation with van Fraassen's *scientific agnosticism* (van Fraassen 1980, 2001).

2. The EHT laboratory

2.1. Hacking's philosophy of experimentation

What is Hacking's entity realism and laboratory science? In his book, *Representing and Intervening*, Hacking famously wrote that for his part he had never thought twice about scientific realism until a friend told him about a Robert A. Millikan oil drop-type experiment, the purpose of which was to try to detect and measure the charge of free quarks.[1] Hacking's friend used Millikan's nomenclature "blowing the spray" (Millikan, 1911, p. 13). From that very day, says Hacking, he became a scientific realist and invented the famous slogan: "If you can spray them, then they are real".[2] Experimenters spray electrons, "that in principle cannot be observed", manipulate them and use them as a tool to help them in hunting quarks. This, Hacking claims, commits them to believe in the reality of electrons. Experimenters have succeeded in doing so when they acquired several general ideas – e.g. debugging and properties about the behaviour of electrons that every experimenter is required to possess – which are not a matter of theoretical explaining. Using this background knowledge, experimenters describe and utilize electrons to investigate something else. Although we use incompatible theoretical models and different theories of electrons, we always refer to the same micro-entity (Hacking, 1983, pp. 22–24, p. 27).

Hacking ends his book with astrophysics and says that unlike micro-entities, astrophysical entities are never used to study something else and can only be measured. When we use micro-entities as tools or instruments of inquiry, we are entitled to regard them as real. Once we switch from astronomy (experiments on the moon and planets) to galactic and extragalactic astrophysics, we have entities, postulated by theories, with which we cannot interfere. Hacking confesses a certain skepticism about black holes, objects which he calls "dark matter" (i.e. "by definition matter that cannot be observed"). He suspects that we might equally represent the phenomena of the universe in another way, in which "dark matter" is precluded. Why on earth would Hacking want to do without black holes?[3] The answer lays in the manipulation of entities. We can know about black holes only through theory. According to Hacking, we do not know whether black holes are responsible for gravitational lensing. Since black holes are "dark matter", we are obliged to say that black holes are inferred, not observed (Hacking, 1983, pp. 271–275; Hacking, 1989, p. 559, pp. 561–562, p. 571).

2.2. Laboratory science

Hacking gives his taxonomy of elements of laboratory experiments and distinguishes a list of fifteen elements in laboratory science. Does the EHT experiment conform to major elements from the list of elements in Hacking's taxonomy of elements of laboratory science? Sheperd Doeleman, the founding leader of the EHT project, said: "We now have a laboratory to really study black holes in a way we never had before" (Doeleman, 2019b). Do the elements from Hacking's list fit into Doeleman's laboratory? Let us first examine the EHT apparatus and try to answer this question.

The apparatus. According to Hacking, *an apparatus* is used in the laboratory "in isolation to interfere with the course of that aspect of nature that is under study". Hacking presents "the materiel of the experiment" consisting of three parts, each associated with a set of instruments: first, a target is prepared by certain devices after which an apparatus is used to interfere with it in some way. Finally, there is a detector that measures the result of the interference with the target (Hacking, 1988, p. 509; Hacking, 1992, pp. 46–47). Keep in mind that Hacking admits that if his accounts deserve to be arranged according to any familiar philosophical "-ism", then it is materialism. "That is most notably true of my account of the laboratory style" (Hacking, 2002, p. 194).

What kind of apparatus is the EHT? It is an *interferometer*. A Very Long Baseline Interferometry (VLBI) array of telescopes has been assembled by researchers from several countries in the world. The global array of widely separated telescopes, located thousands of kilometers apart from one another, operates in tandem, as a single Earth-sized telescope capable of achieving an ultra-high-resolution. Earth-rotation synthesis is also employed. As the Earth is rotating, the telescopes sweep different parts of the sky and scan the target from different angles; the baseline (separation between two telescopes in the array) changes and new measurements are obtained.

Notice that the core of M87 is about 55 million light-years from Earth and its apparent size is extremely small in astronomical terms when viewed from Earth. To resolve something that small from such a large distance we need a giant telescope with a diameter equal to 13,000 km. Since it is impossible to create such a telescope, an array of telescopes is built instead. The angular resolution is set by the spacing between dishes where there is no measurement. To achieve an ultrahigh angular resolution of the VLBI array, one must either increase the maximum separation between two telescopes in the array, the baseline L_{max}, or decrease the observing wavelength λ. Angular resolution – called *beam width* or *beam size* in radio astronomy and *diffraction limit* in optical astronomy – is given by the equation (the angular resolution of the EHT array):

$$\theta \sim \frac{\lambda}{L_{max}} \sim \frac{1.3mm}{10,000km} \sim 20\mu as. \quad (1)$$

It is impossible to indefinitely increase the distance between two telescopes because Earth has a fixed diameter. Expanding the EHT to include radio telescopes on satellites in space would increase the distance between two telescopes and would dramatically increase resolution. Indeed, space-based interferometry is being seriously considered (Bouman et al., 2016, p. 914).

If we speak in terms of the diffraction limit in optical astronomy, then equation (1) is *Abbe's diffraction limit equation*. Hacking points out that Ernst "Abbe was interested in resolution. […] G. B. Airy, the English Astronomer Royal, had seen the point already when considering the properties of a telescope needed to distinguish twin stars. It is a matter of diffraction" (Hacking, 1981, p. 311). In his book, *Representing and Intervening*, Hacking provides a quote saying that the images of minute objects are not depicted microscopically, as required by the laws of refraction, but rather depend entirely on the laws of diffraction. He explains that what this means is that "we do not see, in any ordinary sense of the word, with the microscope" (Hacking, 1983, p. 187). Images of astrophysical objects, as Hacking notes, similarly depend on the

[1] Hacking wrote his book during the time at which the quark experiments caused a great stir.

[2] Hacking's slogan has been criticized by philosophers on the ground that it begs the question. Hillary Putnam has stated that we cannot claim to spray an electron unless we are already justified in claiming that it is real. Hacking's slogan merely begs the question because if they are real, then you can spray them and if they are not real, then you cannot spray them (Putnam, 1984, p. 5).

[3] Based on Hacking's question concerning numbers in mathematics (Hacking, 2014, p. 241).

diffraction limit equation. Likewise, we do not see in any ordinary sense of the word with radio telescopes.

Here is where the interferometer comes into the picture. The radio telescopes across the stations of the VLBI array create an Earth-sized *radio interferometer*: the signals from each telescope are combined to form an interference pattern, interference fringes. One of the first astronomical interferometers was Michelson's stellar interferometer. In 1890 Albert Abraham Michelson invented the optical stellar interferometer, which works on the principle of *optical interferometry*. The stellar interferometer operates in the same way as Thomas Young's two-slit experiment. Light from a source of light (starlight) falls on two widely separated mirrors (like two separated slits) and is reflected into the Hooker telescope at Mount Wilson Observatory. The two beams are combined in the telescope's objective to make interference fringes. In 1920 Michelson and Francis Gladheim Pease measured with Michelson's stellar interferometer the diameter of a star called Alpha Orionis (Michelson & Pease, 1921, p. 259).

Hacking describes the differential interference contrast (DIC) microscope, which works on the principle of *optical interferometry* and is an optical microscope-interferometer. The microscope interferometer operates by splitting a light source into two parts: half the light (specimen light beam) goes to the specimen while the other (reference light beam) is recombined later before observation. According to Hacking, the light beam that traverses the specimen manipulates it (Hacking, 1983, p. 198). Unlike the DIC, Michelson's optical stellar interferometer and radio interferometers are all detectors. A radio telescope collects radio waves and radiation. It does not measure the result of the modification of the target by an apparatus. Indeed, Hacking argues that astrophysics is not laboratory science. Given that the latest technology is used with the most advanced telescopes and modern astrophysicists study astronomical objects by images that are reconstituted electronically, says Hacking, then telescopes not only improve data obtained by astronomers but also render astronomy completely independent of direct observations and manipulation (Hacking, 1992, pp. 33–35).

Ideas: An important element of Hacking's laboratory science is *background bodies of knowledge*: established or working theories, background knowledge and assumptions about the subject matter (Hacking, 1988, p. 509). A disagreement on the terms used by Hacking is worthy of mentioning. Galison refers to Hacking's background knowledge as "the establishment of knowledge prior to experimentation" (Galison, 1988, p. 525). But Hacking disagrees with Galison writing that it is tempting to follow Galison and take his "ideas" as the "establishment of knowledge prior to experimentation" but "that suggests something put in place before the experiment and enduring throughout it". Hacking's picture of experimentation is one of potential modification of any of the elements, including the prior knowledge (Hacking, 1992, p. 44, p. 50).

Since I focus on Hacking's taxonomy, I will also use his language. What are the elements of Hacking's *ideas*?

1) There is a *question* or *questions* about some subject matter. When a question is about a theory, Hacking speaks of the theory in question.
2) A *systematic theory* "of a general and typically high-level sort about the subject matter, and which, by itself, may have no experimental consequences".
3) *Topical hypotheses* (phenomenology). We formulate phenomenological hypotheses, *topical hypotheses* whose purpose is to fit the high-level theory with experiments.
4) *Background knowledge* which is not systematized and plays little part in writing up an experiment because it is taken for granted. But an experiment without background knowledge makes no sense
5) *Modeling of the apparatus.* According to Hacking, we first explain how the apparatus works and behaves. That is, we create "Intellectual models of apparatus" (Hacking, 1988, p. 509; Hacking, 1992, pp. 45–46). Nowadays, however, computer simulations model the behaviour of instruments. In the words of Margaret Morrison, simulations "guide the setup of new experiments" (Morrison, 2015, pp. 223).

Questions: I start with the first element. The members of the EHT collaboration asked two main questions: Is the core of M87 a 3.5×10^9 solar masses black hole or a 6.5×10^9 solar masses black hole? Is it a black hole at all?

Systematic theory: Hacking defines the systematic theory as the high-level theory, which in our case is the classical general theory of relativity (GR) and general relativistic magnetohydrodynamics (GRMHD). In GR, we consider two cosmological cases (exact solutions of Einstein's field equations):

1) The uncharged non-rotating *Schwarzschild black hole* discovered in 1915–1916 by Karl Schwarzschild (Schwarzschild, 1916).
2) The uncharged rotating *Kerr black hole* discovered in 1963 by Roy Kerr (Kerr, 1963).

The horizon of both types of black holes is a one-way area: particles can enter the horizon from outside but cannot leave it from inside. Both solutions are unique: any non-spinning/spinning uncharged black hole is described by the Schwarzschild/Kerr metric. The Kerr black hole (metric) is described by GR. Magnetohydrodynamics (MHD) is the framework that governs the dynamics of the accretion flow and jets around the Kerr black hole. The plasma is treated as a fluid and astrophysicists search for numerical methods to integrate the GRMHD equations.

Theorems and topical hypotheses: In the 1960s and 1970s four theorems and hypotheses were formulated:

1) The *Singularity theorem.* Roger Penrose wrote: "the presence of a trapped surface always does imply the presence of some form of space-time singularity". That is, timelike or spacelike geodesics will be incomplete and a singularity must arise when a stellar-mass black hole is formed. Penrose further showed that "Ultimately the field settles down to becoming a Kerr solution" (Penrose, 1969/2002, p. 1152, p. 1157).
2) The *Cosmic Censorship hypothesis.* There are solutions to the field equations of general relativity describing naked singularities not hidden by an event horizon. Penrose wrote: "In short, the singularity is visible, in all its nakedness, to the outside world!" He then asked: "does there exist a 'cosmic censor' who forbids the appearance of naked singularities, clothing each one in an absolute event horizon?" Penrose answered: it is not known whether naked singularities will ever arise in a collapse which starts off from a nonsingular initial state. He then made "a few highly speculative remarks" (Penrose, 1969/2002, p. 1160, p. 1162). The *Cosmic Censorship hypothesis* is the idea that physics censors naked singularities by always enshrouding them with a horizon. Demonstrating that the core of M87 has an event horizon would not conclusively disprove the existence of naked singularities elsewhere. Moreover, discrediting the cosmic censorship hypothesis "would not be a death blow to general relativity" because its field equations allow for the naked singularities solutions (Psaltis & Doeleman, 2015, p. 77).
3) The *Blandford-Znajek mechanism.* In 1969 Penrose wrote: "Let me suggest another method which actually tries to do something a little different, namely extract the 'rotational energy' of a 'rotating black hole (Kerr solution)'" (Penrose, 1969/2002, p. 1160). Penrose referred to stellar-mass black holes. Less than ten years later in 1977, Roger Blandford and Roman Znajek extended Penrose's "mechanical extraction of energy" process to a Kerr supermassive black hole in a MHD environment: "The simplest application of these ideas is to a model of an active galactic nucleus containing a massive black hole surrounded by an accretion disc", i.e. a supermassive black hole (Blandford & Znajek, 1977, p. 434, p. 451). Most galaxies are

thought to host a supermassive black hole in their center. On the other hand, stellar-mass black holes are formed by the gravitational collapse of stars. These black holes are around a few to ten times the mass of the Sun and are scattered throughout galaxies.

4) The *no-hair theorem.* In John Archibald Wheeler's formulation of the *no-hair theorem*, "a black hole has no hair". The only properties a black hole conserves out of the matter that falls into it are the mass, the angular momentum (spin) and the electric charge. Put differently, all the physical properties of the infalling matter into a black hole are eliminated and we cannot distinguish between two black holes with the same mass, angular momentum and electrical charge (Wheeler, 1981, pp. 32–33). "If the no-hair theorem is false, general relativity will, at minimum, have to be modified" (Psaltis & Doeleman, 2015, p. 77, p. 79).

In 1973 James Maxwell Bardeen conjectured that there was an orbit of photons near the event horizon of a Kerr black hole. The Kerr solution describes the spacetime geometry around astrophysical rotating black holes. Bardeen imagined a source of illumination behind the black hole whose angular size is large compared with the angular size of the black hole. "As seen by a distant observer the black hole will appear as a 'black hole' in the middle of the larger bright source […] The rim of the 'black hole' corresponds to photon trajectories which are marginally trapped by the black hole; they spiral around many times before they reach the observer" and are deflected by the extreme gravitational fields around the black hole which cause strong gravitational lensing (Bardeen, 1973, pp. 230–231). Photons that can still escape the black hole to the observer experience strong gravitational redshift. In 1979, assuming the observer is practically at infinity and the source of illumination is the plasma in the accretion disk, Jean-Pierre Luminet numerically simulated a Schwarzschild black hole. He reconstructed a black-and-white image of a photon ring around the black hole (Luminet, 1979, p. 228, p. 235). In 2000, Heino Falcke et al. wrote that "a marked deficit of the observed intensity inside the apparent boundary" is produced "which we refer to as the 'shadow' of the black hole" (Falcke et al., 2000, p. L14). Falcke et al. considered a supermassive black hole, which is billions of times the mass of the Sun and lies at the center of the galaxy.

Experimental astrophysicists subsequently interpreted the no-hair theorem as follows: for a black hole of known mass, the size and shape of the shadow of the black hole remain nearly unchanged. Alternatively, this could be formulated as follows: shadows of black holes always appear nearly circular (Psaltis & Doeleman, 2015, pp. 77–79; Psaltis et al., 2015, p. 1). This is a *topical* hypothesis. Astrophysicists subsequently took up the task to demonstrate that the shadows of black holes appear nearly circular by reconstructing an image of a lensed ring around a black shadow from interferometric data.

As Hacking realized, Pierre Duhem said that if an experiment was inconsistent with theory, one could revise astronomy or revise the theory of the telescope. If the latter was done, the telescope would probably be rebuilt, creating a different instrument. One could, according to Hacking, modify theory in two ways: either revise the systematic theory, or revise the auxiliary hypotheses – in which Hacking includes both topical hypotheses and modeling of the apparatus (Hacking, 1988, p. 511). In 1992 Hacking concocted what he called, the *self-vindicating* character of laboratory science. Theories of the laboratory sciences, says Hacking, are not directly checked by comparison with the world. Instead, "what happens to a laboratory science as it matures and stabilizes" is that "there evolves a curious tailor-made fit between our ideas, our apparatus, and our observations" (Hacking, 1992, p. 33, p. 58). The use of the words "tailor-made fit" implies a demon arranging things in such a way that there is the above "tailor-made fit". Hacking though is against this kind of demon (Hacking, 1981, pp. 314–315).

Background knowledge: What is the required *background knowledge*? Images are reconstructed from data; models are fitted to data and hypotheses are tested by applying Bayesian statistics. Methods of Bayesian statistical inference play little part in writing up the EHT experiment because they are taken for granted. This background knowledge is nevertheless crucial for the understanding of the EHT experiment. While having an unprecedented resolution, the EHT also collects sparse and noisy data. In order to deal with this problem, it is necessary to develop new algorithms for calibration, imaging and model fitting.[4] These methods utilize machine learning techniques, image processing, an evolutionary paradigm and Bayesian inference. This knowledge is also taken for granted and plays little part in writing up an experiment. Even so, the EHT Collaboration developed seven new tailor-made pipelines – three calibration and three imaging pipelines and one pipeline called THEMIS for model fitting – for the EHT experiment:

Three Calibration algorithms: Three independent pipelines were tailored specifically for phase calibration and processing of 2017 EHT data: 1) EHT-HOPS [Haystack Observatory Postprocessing System (HOPS)] 2) rPICARD (Radboud Pipeline for the Calibration of high Angular Resolution Data) is based on the Common Astronomy Software Applications (CASA) software package. 3) AIPS (Astronomical Image Processing System). HOPS, CASA and AIPS have been in use for many years but groups of members of the EHT Collaboration created three Python scripts with modules from the EHT analysis toolkit called eat library (EHTC 2019c, p. 2, pp. 5–8, p. 23; github.com/sao-eht/eat).

To avoid self-tuning (an automatic controlled process) on statistical (meaningless) noise, the EHT-HOPS pipeline employs the *round-robin, leave-one-out cross-validation* (LOOCV) approach for atmospheric phase corrections. According to Christopher Hitchcock and Elliot Sober, the LOOCV strategy protects against overfitting, i.e. fitting the training set (data) too well but not being able to work on a new data set (Hitchcock & Sober, 2004, pp. 13–15). The data to be corrected is split into equally sized segments (the training set). One segment is left out (validation set). The estimated phase from the average of the training set is used to correct the remaining data in the validation set. The algorithm performs iterations of training and cross-validation. In each iteration, one segment of the validation set is selected, left out, for validation with the remaining segments of data used for training. The process cycles through the data to cover the full data segments and phase corrections are never estimated from the same data, which would otherwise be done by self-tuning.

Three imaging algorithms: First, the imaging techniques used in generating the 2017 EHT black hole images are based on two strategies: the traditional deconvolution CLEAN algorithm and the new RML (Regularized Maximum Likelihood) method. Imaging algorithms are broadly categorized into two methodologies: *inverse* modeling and *forward* modeling. The former includes CLEAN, while the latter includes RML (EHTC 2019d, p. 4).

CLEAN: The original CLEAN algorithm was created in 1974 by Jan Arvid Högbom. The sky is only sparsely sampled by the VLBI array. Thus, the existence of certain points where no data are available (lack of measurements by telescopes) means that in principle an infinite number of possible highest intensity points (images) could be consistent with our visibility data. CLEAN begins with:

(1) an inverse Fourier transform of the sampled visibilities (the *dirty image*).
(2) The algorithm then proceeds to find the points of highest intensity in the *dirty image.*

[4] The algorithms and software packages used by the EHT Collaboration are found on https://github.com/eventhorizontelescope and the 2017 EHT data is found on https://eventhorizontelescope.org/for-astronomers/data. The laboratories that created the algorithms and techniques that generated the 2017 EHT image are: the Harvard and Smithsonian Center for Astrophysics (CfA), the MIT Computer Science and Artificial Intelligence Laboratory and the MIT Haystack Observatory.

(3) The process is iterative and iterations are continued until a desired noise level (a stopping criterion) is achieved in the *dirty image*.
(4) Finally, CLEAN takes the accumulated point source model and convolves the image with a restoring beam, a *clean beam*, Gaussian beam of full width at half maximum (FWHM) $\frac{\lambda}{L_{max}}$ [see equation (1)], instead of the dirty beam.
(5) The algorithm adds the residuals of the *dirty image* to this image to form the CLEAN image.

RML: In RML, the image with the highest likelihood is selected from an infinite number of images and is fit to empirical data. RML is based on maximum likelihood estimation, MLE: finding an image that minimizes a function that is a sum of a chi-squared χ^2 function corresponding to the data and another function called *regularizer*. Regularization may include smoothness (i.e. requiring that the image be smooth) or sparsity (Bouman, 2020a; EHTC 2019a, p. 4; EHTC 2019d, p. 4).

With the above methods, three imaging pipelines, DIFMAP, SMILI (Sparse Modeling Imaging Library) and eht-imaging, were designed using minimal imaging commands and trained on synthetic data. The training of the pipelines was an important stage in reconstructing the 2017 EHT images (discussed in Section 2.1.4). A CLEAN Python script had been developed for the established DIFMAP (Difference Mapping) software package originally written by Martin Shepherd in the 1990s. After loops of cleaning and self-calibration, the DIFMAP script generates final cleaned images.

The SMILI and eht-imaging pipelines were specifically developed by members of the EHT Collaboration to handle the challenges of the EHT experiment. The SMILI Python-interfaced library is based on sparse sampling, i.e. reconstruction of images from sparse data. NumPy and SciPy power the main tasks of the pipelines and the main imaging methods are powered by scipy.optimize.minimize. The two above imaging scripts are based on RML and employ an iterative imaging loop, alternately imaging and then self-calibrating the data. The two pipelines also manipulate images and blur them (with a Gaussian). The three imaging pipelines are open-source software libraries (Bouman, 2019, p. 41; Bouman, 2019; EHTC 2019d, pp. 11–14, p. 31; Chael, 2020; achael.github.io/eht-imaging/; github.com/astrosmili/smili).

Three pipelines for model fitting: Model fitting to data was performed by three different methods: 1) Nested sampling (NS) algorithms. 2) Genetic algorithms (GA). 3) THEMIS, a Bayesian framework, which uses a Markov chain Monte Carlo (MCMC) sampler. Only THEMIS was specifically developed for the EHT experiment (Broderick et al., 2020, p. 4; pp. 15–16; EHTC 2019f, p. 6, p. 15; https://github.com/joshspeagle/dynesty).

First, a GRMHD Simulation Library was generated from several different codes: 1) The Black Hole Accretion Code (BHAC) performs magnetohydrodynamical simulations of an accretion flow onto a black hole. The ions (protons) and the electrons in the accretion disk plasma travel quite a long distance along magnetic field lines before being scattered. This and other factors complicate the calculations and require more computational power. 2) A system called H-AMR (Hierarchical Adaptive Mesh Refinement) is used, which accelerates GRMHD calculations by implementing the AMR strategy: a method that reduces the amount of computations. 3) Two additional codes, a 3D version of High Accuracy Relativistic Magnetohydrodynamics (iharm3D) and KORAL [Kod radiacyjny L (in Polish)], solve the GRMHD conservation laws (GRMHD equations written in conservation form) by a shock-capturing method (Porth et al., 2019, pp. 6–8, p. 31; Sądowski et al., 2013, pp. 3535–3537; github.com/AFD-Illinois/iharm3d).

The EHT Collaboration created a Python module ehtplot with a color submodule, a perceptually uniform colourmap that was used in the simulations and reconstructions. The colours of the images represent the brightness temperature which does not necessarily correspond to any physical temperature of the radio-emitting plasma in the jets and accretion disk. Andrew Chael, a member of the team that interpreted the 2017 EHT image, adds that it takes multiple weeks on a supercomputer to run the GRMHD simulations, to see how the gas is orbiting around a black hole (Black Hole Initiative, 2019; Wielgus et al., 2020, p. 8; http://github.com/liamedeiros/ehtplot).

Now let us go back to the three pipelines:

1) NS algorithms: NS sampling is an Approximate Bayesian Computation method. In Bayesian statistics, we start with Bayes' theorem and update the prior probability distribution $\pi(\Theta)$ of the model M parameters Θ upon receiving new data D to obtain the posterior probability distribution $P(\Theta)$ of Θ:

$$P(\Theta|D,M)=\frac{P(D|\Theta,M)P(\Theta|M)}{P(D|M)}\equiv P(\Theta)=\frac{\mathscr{L}(\Theta)\pi(\Theta)}{Z}. \tag{2}$$

$\mathscr{L}(\Theta)$ is the likelihood that the data would fit the given Θ. We integrate over $\mathscr{L}(\Theta)$ and get the probability that the data D fits the model M. That is, we obtain the marginal likelihood $P(D|M)$:

$$P(D|M)=\int P(D|\Theta,M)P(\Theta|M)d\Theta\equiv\mathscr{Z}=\int\mathscr{L}(\Theta)\pi(\Theta)d\Theta, \tag{3}$$

with the integral taken over all possible Θ of M. $\mathscr{Z}$ stands for the Bayesian evidence.

Now we can compare models in the light of the 2017 EHT data using $P(D|M)$. Model selection is performed by calculating the Bayes factor B of two competing models. We wish to compare two competing models, M_1 and M_2. $P(D|M_1)$ and $P(D|M_2)$ are the marginal likelihoods and $P(M_1)$ and $P(M_2)$ are the prior probabilities of the two models. The posterior odds ratio of two models given the same data is:

$$\frac{P(M_1|D)}{P(M_2|D)}=\frac{P(D|M_1)}{P(D|M_2)}\frac{P(M_1)}{P(M_2)}=B=\frac{\mathscr{Z}_1}{\mathscr{Z}_2}\frac{\pi_1}{\pi_2}. \tag{4}$$

NS is designed to evaluate the evidence $\mathscr{Z}$ but as a by-product it can further sample the posterior probability distribution. The evidence $\mathscr{Z}$ in equation (3) is calculated using *live points* that randomly sample from the prior and are subject to the following condition: the live points having the lowest likelihood are repeatedly replaced with new live points having higher likelihoods. *Dynamic NS* has been developed to increase the accuracy of nested sampling, i.e. to sort likelihoods more efficiently and speed the process. The EHT Collaboration used both NS and the dynesty Python Bayesian *Dynamic NS* sampling code (EHTC 2019f, p. 4, p. 6).

2) GA: In the 1960s people began to talk about problems of adaptation such as how would one improve performance in successive plays of a complex game when the solution of the game was unknown and probably too complex to implement even if it were known? Or how does evolution produce increasingly fit organisms under environmental conditions which necessarily involve a great deal of uncertainty in relation with individual organisms? John Holland proposed to solve these problems by genetic optimization algorithms (Holland, 1973, pp. 88–89, pp. 98–99).

GenA, a GA written in Python, was employed for comparing GRMHD simulations to data. GA are optimization algorithms based on Darwin's principle of survival of the fittest. GA fit models to data by finding the model parameters which minimize a χ^2 statistic. GenA implements the *differential evolution* (DE) algorithm. DE is an optimization algorithm inspired by natural evolution theory: a randomly initialized population of parameters (chromosomes, individuals) is generated. The population of parameters evolves over several generations, improved (*selection*) by iterations in which *mutation* and *crossover* (mixing and mating) operators are used. DE is a kind of natural selection of individuals. These are the major steps of GA (DE):

1) GA creates a random population of model parameters (parents).
2) The next step is: fitness of parameters with the data. Fitness is inversely proportional to χ^2. The fitness of each parameter (F_ν, θ_g and PA, see Section 3.2.1 for definition) with the data is computed such that maximizing fitness of parameters to data minimizes the value of χ^2 [See equations (5) and (6) in the next section].
3) Based on their fitness, best-fit parameters (parents) that move to the next generation are *selected*. They pass their genes to their offsprings, new parameters.
4) New parameters are created by combining the parents (*crossover*) and then added to the population of parameters.
5) The *mutation* operator is used to probabilistically enforce diversity into the population. It is applied to the offsprings of the *crossover* process at a certain predefined rate, which guarantees that good offsprings are not extinct. During the *mutation*, one or more genes within a chromosome are randomly selected and replaced (values of the bits of the gene are reversed).
6) Steps 2–5 are repeated and carried over several generations and GA stops when the value of the fitness reaches a given limit.

GA have been developed whose *selection* operator is not dependent on *crossover* and *mutation* (steps 4 and 5 above):

1. A GA based on a predator-prey model. A predator, which is randomly placed, catches the weakest (worst) prey in the population and eliminates it.
2. An elitist selection model according to which only a limited number of the best-fit individuals are stored and passed to the next generation.
3. Nondominated Sorting GA (NSGA) implements an elitist selection process for multi-objective optimization. A population of parents is generated randomly. Offsprings are compared to parents. If the offsprings do not dominate their parents, they are sorted, moved to the next generation and the parents are replaced. If the offsprings dominate their parents, they are not elite offsprings and are therefore rejected. The process of selecting the non-dominated offsprings continues until the initial population of parents is replaced. GenA utilizes the NSGA-II to explore the parameter space (EHTC 2019f, pp. 5–6).

3) THEMIS: Recall that in Bayesian statistical inference, models and data are compared using likelihoods, which express the probability that the data were obtained from the model. The dependence of the likelihood on the model parameters (incorporating prior probability distributions on the parameter values) is assessed via MCMC samplers. An MCMC sampler is a random walk method through parameter space for performing Bayesian inference. It randomly samples the posterior probability distribution of parameters and generates a sequence (a chain) of random samples of parameters that fit the data. The parameter is dependent upon the previous one in the chain. Although MCMC may be very efficient with huge amounts of data as well as many parameters, it is unable to handle large amounts of samples and complex data and it is time-consuming. Bayesian methods quantify epistemic uncertainty but a gross computational penalty is incurred by using time-consuming samplers.

To solve the above problems, optimization methods have been integrated with MCMC (this does not solve all problems). Two types of hybrid MCMC samplers are used in THEMIS: an affine invariant MCMC and a (differential evolution) DE-MCMC. In the DE-MCMC many chains are run in parallel. DE-MCMC is especially efficient in sampling from models with highly correlated parameters. THEMIS, a Bayesian parameter estimation and model comparison framework, implements a parallel-tempering algorithm for both samplers. Parallel-tempering is a method based on an analogy with statistical physics, called thermodynamic integration, to calculate the Bayesian evidence. Many MCMC chains are run in parallel on tempered versions (copies) of the original likelihood function at different temperatures (Broderick et al., 2020, pp. 14–15).

Algorithms were rifling through thousands of GRMHD model images. An additional method was required to sort model images. THEMIS-*Average Image Scoring* (THEMIS-AIS) was used to give scores to images and decide whether to accept a model even if the reduced chi-squared statistic $\chi^2_\nu \gg 1$, where $\chi^2_\nu \equiv \frac{\chi^2}{\nu} \equiv \frac{\chi^2}{N-M}$. ν represents the degrees of freedom, N is the number of data values and M is the number of free parameters. If $\chi^2_\nu \approx 1$, then the corresponding model is likely to be a description of the data and this model is accepted because $\frac{\chi^2}{\nu} \approx 1 \rightarrow \chi^2 = \nu$. THEMIS-AIS is a method by which the average snapshot frame is obtained from GRMHD simulations after which the χ^2_ν distance between the average frame and the data is measured: the similarity between the average images and the data is calculated.

Modeling of the apparatus: Before performing the actual 2017 observations, the experiment and observations with the EHT at 1.3. mm had both been modeled (simulated) using MAPS (MIT Array Performance Simulator). Synthetic data corresponding to the coverage of the stations of telescopes in the EHT array were generated from simulations of observations of the core of M87. The synthetic data were corrupted with realistic levels of thermal noise and station-based systematic errors. At first, simulated observations for the array consisting of six telescopes at different sites on Earth were performed. The simulation assumed similar conditions to the ones in the actual experiment (angular resolution, exposure time of observation and other features which were close to the observational conditions). A large range of physical conditions had been taken into consideration after which reconstructed images from simulated data were generated. The simulated experiment demonstrated that the EHT would achieve the effective sufficient resolution to produce images of a shadow of the core of M87 (Bouman et al. 2016, 2020a).

2.3. Empirical data

According to Hacking's definition of laboratory science, the stages in handling data include the following: *Data generators*: The machine, tools and scientists are data generators and they produce raw data. By data Hacking means uninterpreted graphs recording variation over time, photographs, and tables. These are raw data. *Data assessment*: This is the first stage of data processing, which may include a calculation of the probable error, i.e. debugging. Such procedures are supposed to be what Hacking calls, "theory neutral". But they are applied only by people who thoroughly understand the experiment. *Data reduction*: In the second stage of data processing vast amounts of unintelligible numerical data are transformed by supposedly theory-neutral computational techniques into manageable quantities or displays. *Data analysis*: In the third stage of data processing, the events studied in an experiment are selected, analysed, and presented by computer but the programs for analysing the data are not exactly theory-neutral methods. *Interpretation of the data* is dependent on theory.

Hacking gives the example of pulsars: once scientists discovered a theory to explain pulsars, it was possible to go back over the data of radio astronomers and find ample evidence of pulsars that could not have been interpreted as such until there was such a theory (Hacking, 1992, pp. 48–49). Although the above stages are typical of laboratory science, they also apply to experiments in astrophysics, as Hacking himself attests when he gives the above example of pulsars, which illustrates the interpretation of data. I shall therefore examine the 2017 EHT data and its processing according to Hacking's five stages:

The raw data: The EHT collected radio waves from M87's core. Every antenna in a VLBI array records the incoming electromagnetic field as a function of time, frequency, and polarization.

First stage: *Data assessment*. This stage includes *Debugging*. Hacking points out that if we manipulate an object, we can eliminate possible sources of error (or as he calls it, "bugs"). He, therefore, pauses to reflect on debugging. Many of these bugs are never understood but they are

eliminated by trial and error. Debugging is not a matter of theoretically explaining or predicting what is going wrong because errors mean all the events that are not understood by any systematic theory. Hacking defines three different kinds of bugs (Hacking, 1982, p. 77, p. 81; Hacking, 1983, p. 269; Hacking, 1992, p. 49):

(1) *The essential technical limitations* that in the end are taken into consideration in the analysis of error. That is, there is always an irremediable error.
(2) *Simple mechanical defects* scientists never think of until these are forced on them.
(3) *Hunches that scientists have about what might go wrong* in the experiment: "Good experimenters guard against the absurd".

Errors always creep into experiments and it's not a matter of whether we can manipulate an astrophysical entity or not. The same holds for the EHT experiment.

First kind of bugs: Among the *essential limitations* that impeded the EHT observations were errors caused by weather conditions. Let us first consider in more detail the EHT apparatus. In April 2017, the Core of M87 was observed with eight telescopes at six geographic sites: Atacama Desert in Chile, Maunakea in Hawaii, Mt. Graham in Arizona, Pico Veleta in Spain, Sierra Negra in Mexico and the South Pole station at Antarctica. It was the first time that the small array called Atacama Large Millimeter/submillimeter Array (ALMA) was included in the VLBI array. Before Alma could participate in the EHT observations, it had to be modified. The ALMA array is composed of 66 dishes which probe the sky at millimeter and submillimeter wavelengths. The upgrade consisted of adding a new capability that allows all 66 antennas to act together as a single telescope. The antennas are configured in different ways, spacing them at distances from 150 m to 16 km (almaobservatory.org).

The VLBI telescopes were borrowed by the EHT Collaboration specifically for observing M87. Not only were those eight telescopes borrowed by the EHT Collaboration, but during 2014–2017, the telescopes and equipment had been modified and upgraded by engineers of the EHT project. Since the VLBI array was an array of different telescopes and different equipment, it was necessary to develop a purpose-built array with unique capability by modifying telescopes and equipment and introducing new systems and significant changes to VLBI equipment. Those changes allowed for the use of existing telescopes and facilities (most of which were not originally conceived to operate as EHT elements) at 230 GHz. Although operating as a single instrument spanning the globe, the upgraded array, namely the EHT, nonetheless remains a heterogenous array, a mixture of new and well-exercised stations, with varying designs of telescopes and operations and with telescopes of various sensitivities (EHTC 2019b, pp. 18–22; EHTC 2019c, pp. 2–3; Doeleman, 2019a; Bouman, 2020b).

It is very expensive to operate the entire VLBI array. Moreover, the EHT Collaboration is provided with a ten-day window every year in which all telescopes in the array are available for use. But every year that window will be in a different month or worse, within this window only five observing nights are triggered due to weather constraints. In June 2017, Doeleman wrote in a blog post that during observations members of the team at all sites went through the checklists twice each observing day to ensure things were ready. An extremely precise, specially-developed Hydrogen maser atomic clock is installed in every telescope in the EHT array. The atomic clocks are so accurate that on a period of hundred million years they would be off by less than a second. The atomic clocks measure small differences in the arrival time of the radio waves coming from the target object to each telescope of the array. Members of the team checked that the following worked: the hydrogen maser atomic clocks were stable, the high-speed data recorders were ready, signal processing instruments were tuned up, and synchronization to GPS was complete (corrections for the ionospheric and tropospheric delays of the GPS signal were made). But the one thing beyond their control was the weather.

At a central command room at the Smithsonian Astrophysical Observatory in Cambridge, Ma, weather data were collected each observing day from around the array. A few hours before the start of observing each night, all sites were checked to see if they were technically ready and with good weather. Then an array-wide decision was made whether or not to observe the target that night: "Will the heavy clouds surrounding a mountain top telescope dissipate, or will they settle in for the night? Is the weather risky at many sites, or maybe just one? And even if the sky above clears up, might ground conditions early in the evening leave a dish iced up and unusable?" The decision was based on predictions, the experience of people at the stations, weather and climate models and forecast data provided by the National Centers for Environmental Prediction (NCEP).

Notwithstanding the above difficulties, "in a true Goldilocks coincidence", explains Doeleman, the 2017 observing campaigns were scheduled for five nights during the April 5–14, 2017 Coordinated Universal Time (UTC) ten-night window when M87 was a night-time source and the weather tended to be best on average on all sites. Observations were made on April 5, 6, 10, and 11, and not on April 7, 8 and 9, due to the chance of strong winds and snow forecast at one station. Weather was good to excellent for all other stations and the EHT array of dishes collected radio signals from the core of M87 (Doeleman, 2017; 2020a; EHTC 2019b, p. 9).

Second stage: *Data reduction*. First, what do we mean by data reduction? To answer this question, it is essential to understand that detecting radio waves from the core of M87 requires high resolution and very high frequencies (230 GHz). In these frequencies, the accretion gas around the black hole and its forward jet become translucent and the radio waves can penetrate interstellar gas and dust and Earth's atmosphere. For this reason, dishes that collect radio signals from the sky have a wide bandwidth recording system. But collecting electromagnetic signals at multiple frequencies produces data with a wide range of signal-to-noise ratio (SNR).[5] One of the consequences of that is the huge amount of data recorded by the EHT, which must be reduced.

Now let us describe how data is reduced. Through the receiver and the electronics at each telescope of the EHT array the weak radio signals collected by the dish were amplified and then converted into electronic signals, digitally processed to produce a digital version of the signals called, digital back end (DBE). The DBE was recorded by four Mark 6 recorders at up to a total rate of 16 Gbps (Gigabits per second) directly to four × sixteen hard drives (whose capacity was almost one million gigabytes = 1 petabyte) together with the time signals provided by the atomic clocks installed in each station. To get a sense of the amount of EHT data involved in the experiment, the hard drives collectively accumulated over 10,000 h of recording at up to 5100 m of altitude without a disk failure.

Second kind of bugs: At this stage, the EHT Collaboration encountered *simpler mechanical and technical defects*, Hacking's second type of errors. The telescopes are placed at a high altitude. But the hardware specifications of hard drives onto which data were recorded state a lower maximum operating altitude. At first, a disk failure was experienced during the observations at high-altitudes. The low ambient air density at high altitudes required sealed helium-filled hard drives. Consequently, a decision was made to hermetically seal hard drives which were used for operations at high altitudes, avoiding the need to build pressure chambers around the recorders.

Thus, totally almost five petabytes were collected from all telescopes. Such a massive amount of data could not be sent over the internet. The hard drives were therefore shipped via FedEx from their respective telescopes to the Massachusetts Institute of Technology (MIT) Haystack Observatory in the US and the Max Planck Institute (MPI) for Radio Astronomy in Bonn, Germany for correlation. There the data were read

[5] The quality of the signal is expressed in terms of SNR, which is the ratio of signal amplitudes to the standard deviation of the noise.

from each hard drive to a supercomputer, a correlator, at up to 4096 MB per second. The next step was processing the data by the correlator. The data were processed using the DiFX (distributed F-X) software correlator (F stands for Fourier transform and X represents cross-multiply). The five petabytes were translated to about a tera byte-worth of correlations. Lindy Blackburn, who led the data processing team, has said that the final 2017 EHT black hole image produced by the EHT collaboration fits in about one kilobyte-worth of information. This image can be tweeted, while the initial data weighs about half a ton. The data were adjusted and synchronized to a common time reference. Since each telescope was at a different position on Earth, each had a slightly different view of the core of M87. Using time-stamps created by the atomic clocks at each site, the correlator's software matched up and compared the data streams from every possible pairing of EHT's eight telescopes. The signals interfered with each other and interference fringes were formed (EHTC 2019b, p. 11; Hearing, 2019, pp. 4–5; Doeleman, 2019a; Black Hole Initiative, 2019).

The EHT Collaboration once again encountered *simpler mechanical and technical defects*, Hacking's second type of errors. A failure of a hard drive in one of the removable modules was discovered in the Mark 6 recording system connected to the James Clerk Maxwell Telescope in Hawaii. This caused a loss of a very small fraction of data and therefore required no special handling because it was automatically corrected by DiFX that adjusted data based on the amount of data obtained from each station of the EHT array. Further, unanticipated issues with ALMA were discovered: a small glitch in the ALMA correlator, the tuning of one of the ALMA generators was specified to insufficient precision, and the ALMA delay system automatically removed the bulk atmospheric delay from above the ALMA array. The correlator software DiFX furthermore removed the bulk atmospheric delay from above each station of the EHT array, resulting in a double correction for ALMA. This problem was solved by an adjustment that was made for ALMA: the normal atmospheric correction at all stations other than ALMA was merged with a no-atmospheric correction at ALMA (EHTC 2019b, pp. 8–11; EHTC 2019c, pp. 27–28).

You should bear in mind that after correlation of the (DBE) baseband data, the complex *visibility* data V_{ij} is obtained. V_{ij} is the output of the correlator and the fundamental data product which gives information of both the *amplitude* and the *phase* of the fringes of the signal on a baseline, between two stations i and j.

Third kind of bugs: But at this point, we arrive at Hacking's third type of bugs, *hunches about what might go wrong*. Blackburn has explained that there is a lot of noise in the system: "The atmosphere is radiating and causing ten thousand times more noise than signal in our data so, we have to average billions of samples to get something out of that and that's why we take so much data so, we can average down and extract those weak signals" (Black Hole Initiative, 2019). The received radio signals from most astronomical sources are extremely weak in comparison to the noise in the instruments of the VLBI array. An astronomical signal is not detectable above the noise until phase corrections are applied. Since the signal is masked by noise, noise in the signal might be mistaken for genuine astronomical signal, and vice versa.

Regarding the visibility, the V_{ij} data were calibrated by the three calibration pipelines EHT-HOPS, rPICARD and AIPS – phase errors (fringe-fitting) and atmospheric (turbulence and variations) errors were corrected and fixed – and the amount of data was consequently reduced so that noisy data were removed. But Hacking argues that "there are also estimates of systematic error". *Systematic errors* are estimated by theories of the detector, apparatus, and target, and by deductions from topical hypotheses. The estimation of systematic errors requires knowledge of the theory of the apparatus (Hacking, 1988, p. 510).

Although calibration was made by the three processing pipelines, residual station-based amplitude and phase calibration errors nevertheless persisted and overwhelmingly dominated the remaining thermal noise ε. The reason for that is that at high frequencies, visibility amplitude calibration becomes extremely difficult, not to speak of visibility phase calibration. Accordingly, *amplitude gain errors* (g_i and g_j) and *phase errors* (ϕ_i and ϕ_j) corrupt the measured visibility, V_{ij}: $V_{ij} = g_i g_j \mathscr{V}_{ij} e^{i(\phi_i - \phi_j)} + \varepsilon$, where $\mathscr{V}_{ij}$ stands for the unknown model visibility, i.e. the ideal measured visibility with no atmospheric corruption.

Recall from Section 2.1.2 that the distance between two stations is $\approx 10,000\ km$. There is a difference of arrival times (time delay) of the incoming radio waves at each station. Radio waves from M87 reach station i before they reach station j. Moreover, turbulence, refraction and water vapor in the atmosphere cause attenuation (amplitude errors) and phase delay of waves at each station. Residual gain errors are systematic errors introduced at the instruments (such as poor pointing or focus of the antenna) during the propagation of the radio waves through the atmosphere and the antenna (Bouman, 2019).

In order to deal with the above problem, the imaging pipelines perform *self-calibration*. The eht-imaging pipeline performs self-calibration by minimizing:

$$\chi^2 = \sum_{i<j} \frac{\left(|V|_{ij} - g_i g_j |\mathscr{V}|_{ij}\right)^2}{\sigma_{ij}^2}, \tag{5}$$

for each baseline in the EHT array, where $|V|_{ij}$ is the measured visibility amplitude, $|\mathscr{V}|_{ij}$ stands for the model visibility amplitude and σ_{ij} represents the uncertainty on V_{ij}.

While in progress, the imaging + self-calibrating pipelines gradually remove g_i and g_j by going through a loop of multiple iterations. V_{ij} is compared each time to the model image, i.e., V_{ij} is calibrating itself by comparing to $\mathscr{V}_{ij}$.

Gain amplitudes g_i represent approximately between 40 and 143 additional superfluous nuisance parameters per data set. According to the Bayesian Information Criterion, however, there is a penalty term that prevents from complex models described by a large number of parameters from being chosen. The nuisance gain parameters are subsumed into the likelihood and incorporated as model parameters by a method called analytic marginalization. Assuming Gaussian priors, the log-likelihood:

$$\mathscr{L} = -\sum_{i<j} \frac{\left(|V|_{ij} - g_i g_j |\mathscr{V}|_{ij}\right)^2}{2\sigma_{ij}^2}, \tag{6}$$

is maximized and a marginalization over all station gain amplitudes is performed. Within THEMIS, at each MCMC step, the station gains are addressed by marginalizing over the nuisance parameters. THEMIS self-calibrates the station gains to the model at each likelihood evaluation (Broderick et al., 2020, pp. 14–15, p. 35; EHTC 2019f, p. 25; Wielgus et al., 2020, p. 9).

As seen from equations (5) and (6), minimizing the χ^2 and maximizing the log-likelihood are two sides of the same optimization coin.

To recover information about the *phase* of the core of M87, the visibility phases are reformulated in terms of closure phases. The closure phase is the product of three visibilities (from three stations). It is formed from baseline visibilities on a closed triangle *ijk*.[6] Closure quantities across days remain stable. Closure amplitudes are unaffected by g_i, g_j and closure phases are unaffected by ϕ_i, ϕ_j and depend only on the apparent structure of the target source and thermal noise (random noise, which, as stated above, is approximated as Gaussian noise).

[6] Consider three telescopes *ijk*. The visibility we measure on baseline *ij* is: $(u, v)_{ij}$ + phase error at the station $(i - j)$. The measured visibility on baseline *jk* is: $(u, v)_{jk}$ + phase error $(j - k)$ and the one measured on baseline *ki* is: $(u, v)_{ki}$ + phase error $(k - i)$. Now the phase errors $(i - j) + (j - k) + (k - i)$ are all canceled when the phase visibilities are added up, leaving a quantity: $(u, v)_{ij} + (u, v)_{jk} + (u, v)_{ki}$ such that the baselines "close". The logarithmic closure amplitude is constructed from combinations of visibilities measured on four stations and is insensitive to variations in the amplitude gains (EHTC 2019d, p. 4).

Having reconstructed images by the three imaging pipelines, validation tests were performed to assess their reliability. One of the tests checked self-calibration. Observations of the core of the galaxy 3C279 were performed with the EHT in April 2017. Images were reconstructed and self-calibrated by the three imaging pipelines and then compared to the 2017 EHT M87 images. The EHT Collaboration showed that station gains in both cases were broadly consistent and then concluded: consistency between the two sources M87 and 3C279 and among the three different pipelines provides confidence that the gain corrections are not imaging artefacts or missing structures in the 2017 EHT images (EHTC 2019d, pp. 22–24). In other words, the EHT Collaboration "guard against the absurd" to use Hacking's words (third type of errors) (EHTC 2019a, p. 3; EHTC 2019b, p. 2, pp. 7–8, p. 14; EHTC 2019c, p. 7, pp, 13–18; EHTC 2019d, p. 2, p. 12, EHTC 2019f, pp. 4–5, p. 15, p. 25).

Third stage: *Data analysis*: After calibration the remaining data were organized and then represented in the form of a graph showing the final observed visibility amplitudes received by a pair of telescopes as a function of distance between the telescopes (baseline). The graph shows peaks in the observed visibility amplitudes that appear as small "hills" that come up and down again (the data look like a Bessel function) and appear across all four observed days: the peaks have two minima (nulls) on either side and at the location of the minima the visibility amplitudes of the data are very low. The first of the nulls occurs at $\sim 3.4G\lambda$ (giga-lambda) and the second is observed at $\sim 8.3G\lambda$. The high peak between these two nulls is at $\sim 6G\lambda$. With the new visibility amplitudes, a slight anisotropy was discovered: the peaks were neither exactly symmetric nor were they spaced equally apart on the graph. Doeleman and most of the members of the collaboration agreed: "We knew we were onto something" (EHTC 2019c, p. 24; Bouman, 2019; Doeleman, 2019a).

2.4. Reconstructing images

The fourth stage in Hacking's data-treatment process is *Interpretation of the data*. 2017 EHT images were reconstructed using the engineering release of data. From the epistemological standpoint, images were generated in three stages. Galison calls the first stage, *Socio-Epistemology*, the second stage, *Mechanical Objectivity* and the third is again *Socio-Epistemology*. The *Socio-Epistemology* and *Mechanical Objectivity* stages form a cyclic loop:

Socio-Epistemology. First, images are blindly reconstructed, evaluated and interpreted (Galison, 2019). The 2017 EHT April 11 data were selected. The reason the EHT Collaboration used this data set was that on the April 11 observing day, the EHT array of radio antennas covered the largest area for M87 and the amplitude calibration among stations was the most stable (EHTC 2019d, p. 9). In June 2018 the EHT Collaboration split up into four teams in different regions of the world and each team chose a different imaging method and worked in isolation for seven weeks on the data. The teams neither talked nor crossed photos and that enabled them to avoid getting into group-think, which would otherwise happen if they "were all in the same room, examining the same data, all using the same algorithms" (Doeleman, 2019a). Each one of the four teams relied upon the judgment of its members to select a different imaging method to convert the data into images. Teams 1 and 2 used eht-imaging and SMILI, while Teams 3 and 4 used a version of the CLEAN algorithm.

Several weeks later on July 24, 2018, they all came together and revealed for the first time the four images they produced. Although not precisely identical, when comparing images, all four teams obtained an image with a central dark area surrounded by an asymmetric ring.

The image-generation process in the first stage often depended on the idiosyncrasies of the observer and required a large amount of manual intervention, i.e. a "knowledgeable user". After the first stage, there was the problem of demonstrating "that we hadn't all just injected a human bias favoring a ring into our software" (Bouman, 2019; Galison, 2019).

Mechanical Objectivity. The essential point is that each one of the three scripted pipelines, DIFMAP, SMILI and eht-imaging has fixed settings (e.g. pixels' size) and also parameters taken as arguments (for RML methods these are regularizers) such as the image total flux density. Those imaging parameter combinations for the three different imaging pipelines are derived by training the algorithms and conducting parameter space surveys on synthetic data (Bouman, 2020a; EHTC 2019d, p. 10):

1) *Four geometric models*: four simple geometric models of objects were selected – a uniform ring, a crescent (an asymmetric ring around a shadow), a uniform disk, and two circular objects or blobs – all approximately reproduce the features of the 2017 EHT data. That is, for all models the first null is consistent with what is seen in the measured visibility amplitudes (EHTC 2019f, pp. 3–4). Unlike the black hole, the EHT Collaboration made the four models (Galison, 2019). The three imaging pipelines were trained on the four morphologies. First, the pipelines were trained on a ring and a crescent. Subsequently, they were also trained on shapes that did not look like a crescent: a filled disk and double blobs.
2) *Finding fiducial parameters*: synthetic data sets were subsequently produced from the four model images by the eht-imaging library with the arrangement of the EHT on all four days of the 2017 observations. Station-based residual gain errors were further added. The *Top Set* of imaging parameter combinations for each pipeline was searched for. Thirty thousand parameters were tested, during which many similar-looking images were reconstructed from synthetic data and blurred to the resolution of the EHT. Those images were compared to images that were generated from the 2017 EHT data. The best-performing hyper-parameters called *fiducial imaging hyper-parameters* were determined from the Top Set parameters for each pipeline. The hyper-parameter selection procedure was tested to see whether or not it would lead to reconstructed images that were similar to the training set. The disk model was excluded from the training set. All three imaging pipelines nonetheless produced (from the synthetic data) an image with a disk shape that does not resemble any of the other models. The fiducial parameters that best produced images of the disk shape were those that described the crescent, the ring, and the double objects models. By repeating the same process with each of the four geometric models, the set of fiducial imaging hyper-parameters was cross-validated. The chosen fiducial imaging hyper-parameters were able to distinguish between the four geometric shapes and perform well on average in reconstructing images of the four shapes (of what was there).
3) *Reconstructing fiducial images*: Katie Bouman describes the above procedure: synthetic data were generated "as if the Event Horizon Telescope were actually seeing a disk on the sky, with no hole in the center. Then, when we transferred these exact imaging settings onto M87 data we found that each imaging pipeline still produced a ring with a hole in the center". When the three pipelines were fed with the data set, they have never "seen" before – the 2017 EHT data – they produced the same fundamental structure: an asymmetric ring around a dark shadow. *Fiducial images*, i.e. 2017 EHT images, were produced. Daniel Palumbo, a member of the EHT Collaboration, adds that the three pipelines knew nothing about disks and crescents. All the algorithms were doing was what they were trained to do (Bouman, 2017, 2019, p. 41; Bouman, 2020b; Black Hole Initiative, 2019; EHTC 2019d, pp. 7–17, p. 22; pp. 32–33, pp. 37–41).

Socio-Epistemology. Fiducial images were reconstructed from the April 11, 2017 EHT data set. Reconstructed images of data sets from observations made on April 5, 6, and 10 were subsequently incorporated, while preventing the poorest reconstructions of 2017 EHT April 10 from dominating the outcome. The best 2017 EHT images, for each of the four observed days, from each of the three imaging pipelines, were subsequently selected, the averages of which were taken and restored to an equivalent resolution of the EHT array. The EHT Collaboration formed *a single image* (which they called the *Consensus Image*) from the

average of each day. This image "made the front page of newspapers around the world" (Bouman, 2020a). The images produced from the three pipelines are broadly consistent across all four observing days and show a central shadow and a prominent asymmetric ring having an enhanced brightness toward the south (EHTC 2019a, p. 5; EHTC 2019d, pp. 16–21, EHTC 2019f, p. 46).

Galison asks: why would one want a single image? He gives two reasons (Galison, 2019):

1) *A psycho-social reason.* A reason that has nothing to do with the science: nobody in the EHT Collaboration wanted to be left out of the 2017 EHT image. People didn't want the image to be made by one person or the other from the collaboration. Hence, the images were averaged and the single image was a kind of collective achievement. But Galison is not persuaded by this reason because it is not a scientific argument.
2) There is another reason. The 2017 EHT image that came as an average looked pretty much in a qualitative way like each of the different constitute images that were being averaged. Suppose we had two pictures, says Galison, one was Mount Everest and the other was a valley that looked like the opposite of Mount Everest. If we took the average and got something flat that looked like the desert, then that desert-like image would neither look like a mountain, nor would it represent a valley. Actually, this example only roughly represents the *Consensus image.* Galison therefore argues that what the average 2017 EHT image does is that it values that which is in each of the images. The *Consensus image* is a version of truth to nature because it is an attempt to extract a single picture that is an image that would hold our belief as reflecting something (Galison, 2019).

This leads to van Fraassen. For van Fraassen, images produced from data collected by radio telescopes do not reveal what exists behind the observable phenomena but rather produce significant new effects. Van Fraassen's basic idea is that a theory does not have to be true to be a successful theory. To accept a theory there is no need to believe it to be true. We do not even have to believe that the entities it postulates are real. What matters is only that the theory is correct from the observational and experimental point of view, that it is empirically adequate. We believe the theory is empirically adequate because observable entities are as the theory says they are. That is the reason why we accept the theory. It correctly describes what is observable and saves the phenomena. And so, it is the aim of science to construct empirically adequate theories. Scientists construct theories in order to represent the observable phenomena. If we accept a theory, we believe it is empirically adequate with respect to what is measurable and observable and not with respect to unobservable phenomena, which are indirectly accessible through instruments. A look through an optical telescope at the moon is considered observation because we can land on the moon and see for ourselves that there are craters.

If observable is visible to the unaided eye, radio telescopes create new observable phenomena. They do not reveal what exists behind the observable phenomena but are rather experimental arrangements that produce significant new effects for us to represent the world. Detecting the presence of things and the occurrence of events by radio telescopes does not generally count as observation because observation is perception, and perception is possible for us without radio telescopes. The EHT radio telescope array produces data. After processing the data images of an asymmetric ring around a dark shadow are reconstructed by computers (by imaging algorithms). Van Fraassen calls such images, *public hallucinations* and says that they could be an illusion or images of a thing. Yet the image itself is not a thing because it is being produced by an array of telescopes and computers. Accordingly, the *Consensus image* is like hallucinations because it is not a real thing and it is public: the 2017 EHT image of M87 has been shown everywhere and seen by all. Conversely, the painting *Mona Lisa* is a thing but images produced using a radio telescope are not things. We do not have the experience of seeing such images because they do not exist (based on van Fraassen, 1980, pp. 10–17; van Fraassen, 2001, pp. 154–155). The empiricist says that, although we see a reconstructed image by a computer showing a ring around a shadow, the *Consensus image* is not like the experience of seeing a real thing.

Doeleman has been asked: if you were somehow in a spaceship that was whirling around the shadow in the 2017 EHT image of M87, how long would it take to go around it? He replied: "First I would give anything to be in that spaceship so sign me up". And then he said: "If I can get wonky for one moment", the bright ring is the "innermost stable orbit in which matter can move around the black hole before it spirals in" (Doeleman, 2019c). But van Fraassen would probably disagree with this interpretation: rather than a thing, the 2017 EHT image is a public hallucination. We cannot be in a spaceship that is whirling around a shadow of a hallucination.

3. Coincidence and comparison

3.1. Coincidence and agnosticism

3.1.1. Coincidence, reproducibility and robustness

In 1981, Hacking invoked the *Argument of the Grid*. Imagine we create a circular copper disk, a grid (a wire mesh). We reduce the disk to a micro-grid and coat it with a layer of carbon on which the sample is examined. We look at the tiny disk through a microscope and see the same shapes/numbers/letters as were inscribed into the macro-grid. We know that what we see through the microscope is veridical because we made the grid to be just that way, we can check the results with the microscope and even check it with other microscopes. We may look at the grid using twelve different kinds of microscopes but still see the same structure as the original grid. This is the *Argument from Coincidence*, the idea that if experimenters see the same fundamental features of a structure using several different physical systems, then they have excellent reasons for saying, that is real rather than, that is an artefact. That is because it would be a preposterous coincidence if, time and again, different physical processes produced identical visual configurations which were artefacts of the physical processes (Hacking, 1981, pp. 314–315; Hacking, 1983, pp. 201–203).

The EHT Collaboration explains that the processing of data focusses on the use of "unbiased automated procedures, reproducibility, and extensive review and cross-validation" (EHTC 2019c, p. 2). Unbiased automated procedures and cross-validation were discussed in Section 2.1.4. Let us now define "reproducibility".

For Hacking, *repeatability* or repetitions of experiments are defined as attempts to do the same experiment better using different kinds of equipment. Rather than comparing among experimental results obtained from twelve different instruments, a single improved or a different instrument is used. Experiments are often repeated by sceptics when people do not believe an experimental result (Hacking, 1983, p. 231).

Here is Cartwright's point of view. For Cartwright, there is a world of difference between repeatability and reproducibility. *Replicability* is doing the same experiment again, i.e. you repeat the experiment by using the same instrument on the same phenomenon. We expect that the instrument would produce the same results wherever it is applied to look at the same phenomenon. *Reproducibility*, says Cartwright, is something different: "Philosophers commonly call the argument involved here 'the Argument from Coincidence'". For Cartwright, *Coincidence* is reproducing the same experimental result by *different* kinds of experimental methods in a variety of *different* experimental situations. The different kinds of experimental procedures require different skills and rely on very different assumptions. The *Argument from Coincidence* "is taken as an argument against a general skepticism about the existence of anything we cannot directly observe" (Cartwright, 1991, p. 147, pp. 149–150).

Suppose the result of the experiment is reproduced in a variety of

very different experimental situations involving very different procedures. Those methods require both independent skills and independent assumptions. We are using our best-confirmed theories and models and working at our most careful and precise level. Even so, it seems that one or more of these experimental designs (e.g. method 1) is wrongly built or mistaken. In this state of affairs, if each of the separate methods independently produced the same wrong result, it would really be a coincidence. That is why, says Cartwright, it is so compelling when studies using very different designs all agree. If that is the result of the experiment, we can read the phenomenon with just one instrument and there is no need for very different kinds of experimental methods in a variety of very different experimental situations. *Reproducibility/coincidence*, then, is nothing more than "a guard against errors in the instruments" and that is the reason why it is not necessary. "The more secure we are in the design of our instruments, the less need there is for reproducibility".

Note that there may be a justification for the use of *reproducibility* and Cartwright specifies in what terms we may accept it. An experiment is probably *coincidence*-worthy, if we convert the above entire picture into an argument saying that method 1 is mistaken because a certain physical process is responsible for the distortion. We then arrive at a kind of *reproducibility* that is called *robustness* across alternative and contradictory implementations (Cartwright, 1991, pp. 149–151, p. 154).

I now analyse the EHT experiment through the prism of Hacking's and Cartwright's versions of the *Argument from Coincidence*. The EHT Collaboration explains: "Direct comparisons between corresponding data products delivered by separate pipelines allow us to quantify the degree of confidence that we may have in their properties and their dependence on specific choices in calibration procedure" (EHTC 2019c, p. 21). Direct comparisons between the three pipelines (EHT-HOPS, rPICARD, and AIPS) had been performed and a single pipeline output was subsequently designated as the primary data set of the engineering data release. All three pipelines produced data with a slight anisotropy in observed amplitudes.

Recall from Section 2.1.3 that the EHT-HOPS pipeline performs a cross-validation LOOCV test. Following Cartwright's interpretation of the *Argument from Coincidence*, we may use the EHT-HOPS pipeline, and there is no need for very different kinds of pipelines (rPICARD and AIPS) to quantify the degree of confidence that we may have in the pipelines and the result produced by them. Indeed, the data set from EHT-HOPS was selected as the (fringe fitted and network calibrated) primary data set for creating images with the remaining two sets (from rPICARD and AIPS) available "for validation and direct data cross-comparisons" (EHTC 2019c, p. 14).

The key understanding is that although calibration was made by EHT-HOPS, rPICARD and AIPS, station-based gain errors nevertheless persisted and overwhelmingly dominated the remaining thermal noise. Recall from Section 2.1.3 that in order to deal with this problem, the EHT Collaboration designed three separate imaging pipelines which performed self-calibration. Designing imaging pipelines that would alternate between imaging and self-calibration was *the major* technological challenge of the EHT project. It therefore makes perfect sense that the rPICARD and AIPS data sets functioned as "guards against" errors, to use Cartwright's words.

Why then did the EHT Collaboration not use a single pipeline, the EHT-HOPS, for calibration? I suggest the answer is twofold, with the first part being Galison's *psycho-social reason*. Nobody in the EHT Collaboration wanted to be left out of the EHT enterprise. And the second part is *coincidence*: the EHT Collaboration reasoned that if three pipelines produced data with the same fundamental features, then we have excellent reasons for saying, that slight anisotropy in visibility amplitudes is not an artefact ("We knew we were onto something", Doeleman, 2019a). Although Hacking advocated manipulating micro-entities, the central feature of his *Argument from Coincidence* is the different methods and the conclusion that this is probably not an artefact.

In 2018 the three imaging pipelines produced images of an unresolved photon ring around a shadow. Although not exactly identical, the images produced from the three pipelines were all consistent. One method, DIFMAP, produced a ring which was fluffier than the one produced by the other two techniques, eht-imaging and SMILI because of the Gaussian restoring beam. The SMILI reconstructions were fainter than the eht-imaging reconstructions, the latter ones were generated in higher resolution. Doeleman said: "These are the final images that the three different techniques our collaboration came with from the April 11th data. This really showed us that we had a very robust finding" (Doeleman, 2020b). The EHT Collaboration has concluded: "This morphology is robust among the different analysis methods" (EHTC 2019f, p. 21) and a *Consensus image* was created from the best images generated by each of the three pipelines.

Cartwright says that here is a reasoning she does not understand. Pipeline 1 uses the CLEAN method and Pipeline 2 and 3 use RML methods, and the results are the same anyway. Since the results are very robust, says Cartwright, we think that there must be some truth in them. She disagrees with this claim. That is because "at the very best one and only one of these assumptions can be right". We may look at the images produced by the three imaging pipelines but pipeline number four may generate an image with no shadow at all. If "God's function" is this fourth pipeline, then the three other pipelines teach us nothing. The imaging pipelines are uncorrelated with a general argument because their output image has no bearing on what images the fourth, fifth and sixth pipelines would reconstruct. Cartwright agrees that it is a coincidence that all methods find the same results. But she does not see what reason we have to assume that the correct explanation for the coincidence is that each of the instruments is nevertheless reading the outcome correctly. She therefore suggests to "accept the coincidence just as it is" (based on Cartwright, 1991, p. 154).

3.1.2. *Technological agnosticism and scientific agnosticism*

Members of the EHT Collaboration have been asking themselves many times over: How do we prevent ourselves from seeing shadows everywhere? As the saying goes, to a man with a hammer, everything looks like a nail. That being the case, if you have got a shadow detector, then everything starts to look like a shadow (Doeleman, 2019a; EHTC 2019d, p. 1, p. 9).

The above question asked by the EHT Collaboration hinges on another question: what is the difference between observing an entity and observing that something? Van Fraassen explains that difference through a fable. Suppose one of the Stone Age people recently found in the Philippines (living under a rock) is shown a tennis ball. From his behaviour, we perceive that he has noticed it because he picks up the ball and throws it. But he has not seen that it is a tennis ball for he does not even have that concept. He would first have to learn a great deal about it. To say that he does not see the same things and events as we do is exploiting the ambiguity between seeing and seeing that certain thing (van Fraassen, 1980, p. 15).[7]

The intricate ambiguity between seeing and seeing that certain thing is most relevant to the EHT experiment. In generating images from the 2017 EHT data, it seems that the EHT Collaboration strived to be like the Stone Age people, i.e. like people who do not have the concept of a black hole. Bouman says in her TED talk: "How to take a picture of a black hole": we as twenty-first-century people can easily tell the difference between black hole simulation images and images taken every day on Earth (Bouman, 2017). But the Stone Age people are in uncharted

[7] During the 1970s a Stone Age "Tasaday" tribe living in the Philippines was discovered. They lived in caves and it was reported that they had no contact with people for a thousand years. The "Tasaday" grabbed a lot of attention and hooked the imagination of many people. But a decade later it was found that the whole 'Tasaday" melodrama was nothing but a hoax. In 1980, in the middle of the fuss over the Tasaday people, van Fraassen wrote his book, *The Scientific Image*. In his book, he sheds philosophical insight into the nature of observability using the example of these supposedly Stone Age people.

waters. They have never seen a simulation of a black hole, wherefore they cannot say that a certain image represents a black hole and another shows a terrestrial object.

Bouman further says in her TED talk: "This is a little bit like giving the same description to three different sketch artists from all around the world. If they all produce a very similar-looking face, then we can start to become confident that they're not imposing their own cultural biases on the drawings" (Bouman, 2017). Great efforts were made to get around that bias by applying a *Coincidence* strategy to identify the intrinsic features in the images.

The EHT Collaboration explains: "The features of the underlying image are then reconstructed either using agnostic imaging algorithms or by directly fitting model images to the interferometric data" (Psaltis et al., 2020, pp. 141104-4). In order to avoid introducing human bias into the imaging process, the EHT Collaboration construed *Coincidence/Reproducibility* – the same morphology of an asymmetric ring and a dark shadow that could be obtained equally well by each of the three imaging pipelines – as *Technological Agnosticism*, no preference towards any one of the three imaging pipelines. The EHT Collaboration ends the fourth letter, saying that imaging techniques and results for the 2017 EHT observations of M87's core have been presented. These are the first observations with sufficient sensitivity for reconstructing images that "can provide agnostic results" (EHTC 2019d, p. 29).

Technological Agnosticism does not provide justification for believing in the existence of unobservable entities. *Scientific Agnostics*, on the other hand, have no opinion about the existence of entities. Van Fraassen explains that he is not agnostic about theories which postulate unobservable entities. He is agnostic about the existence of these postulated entities. If you are agnostic about the existence of an unobservable entity, then you are also agnostic about its non-existence. Scientific agnostics purport to have no opinion about the existence of unobservable entities; they neither believe nor disbelieve in their existence; they have no opinion at all about whether these entities exist (van Fraassen, 1998, pp. 213–220).

3.2. Confirmation of experiment

3.2.1. Comparison of models with data

The next stage in the experiment was comparison of model images with the 2017 EHT data. The working hypothesis that M87's core is a black hole described by the Kerr metric was adopted. When fitting models to data, we ask what is the (likelihood) probability of getting the data given the model? We adjust model parameters to obtain best-fit to data. A GRMHD model has *adjustable parameters* which are defined by general relativity and GRMHD:

1) Black hole spin and magnetic flux: Recall that according to the *no-hair theorem*, black holes have two hairs. Black holes are entirely specified by their mass M_{BH} and angular momentum J. A GRMHD model is defined by two dimensionless parameters: *spin* a_* and *magnetic* flux ϕ:

$$a_* \equiv \frac{Jc}{GM_{BH}^2}, \quad -1 < a_* < 1 \quad \text{and:} \quad \phi \equiv \frac{\Phi_{BH}}{\sqrt{\dot{M} R_g^2 c}}, \tag{7}$$

where G is the universal gravitational constant, c is the velocity of light, Φ_{BH} is the magnetic flux, $\dot{M}$ is the mass accretion rate (the rate at which mass is accreted onto a black hole) and the (Schwarzschild) gravitational radius is:

$$R_g = \frac{GM_{BH}}{c^2}. \tag{8}$$

If the angular momentum J of the Kerr black hole is zero [in equation (7)], then it becomes a Schwarzschild black hole.

2) Image scale, position angle, and observer's inclination: Three additional parameters are defined: the *position angle* of the approaching (forward) radio jet PA measured east of north, the *observer's inclination* i, the orientation of the observer through PA, and the *image scale* θ_g :

$$\theta_g \equiv \frac{GM_{BH}}{dc^2} \approx \frac{M_{BH}}{d}, \tag{9}$$

where d is the distance to M87: $d = 16.8\,^{+0.8}_{-0.7}$ Mpc (Megaparsec) = 55 million light years and θ_g represents the angular gravitational radius.

3) Total compact flux density: F_ν is measured in Jy (Jansky). The average flux density of 1.3 mm (230 GHz) emission is between 0.5 Jy and 0.6 Jy.

4) Temperature ratio of electrons to protons: Photons at 1.3 mm wavelength observed by the EHT are believed to be produced by synchrotron radiation. Relativistic energetic electrons gyrating in magnetic fields, i.e. caught up by the electromagnetic field of the black hole, emit radio photons known as synchrotron radiation. The plasma is considered to be composed of electrons and ions that have the same temperature in the funnel (the strongly magnetized regions of the accretion flow), but have a substantially different temperature in the gas dominated regions (the middle of the accretion disk). Assuming the gas is composed of nonrelativistic ions with temperature T_i and relativistic electrons with temperature T_e, the ratio of the temperatures of the two species can then be imposed in terms of a single parameter R_{high}. If the synchrotron radiation is emitted from weakly magnetized regions then $\frac{T_i}{T_e} \approx R_{high}$. If the emission comes from strongly magnetized regions, then $T_i \approx T_e$ (EHTC 2019a, p. 6; EHTC 2019e, pp. 4–5, p. 12).

Two kinds of models for the Kerr black hole: A GRMHD Simulation Library of three-dimensional *Standard and Normal Evolution* (SANE) and *Magnetically Arrested Disk* (MAD) models was created. Each model is an approximation of the complex non-linear general relativitsic dynamical system of a turbulent, magnetized accretion disk flow orbiting compact objects, in our case study a Kerr rotating supermassive black hole. Models with $\phi \approx 1$ or $\phi \approx$ few in equation (7) are conventionally referred to as SANE where accretion is largely unaffected by the black hole magnetic field. Once the magnetic flux reaches $\phi \approx 15$, the large-scale magnetic field accumulated by the accretion disk stops the accretion flow. Models with $\phi \approx 15$ are conventionally referred to as MAD (Porth et al., 2019, p. 6).

Forty-three high-resolution, three-dimensional simulations covering well the physical properties of magnetized accretion flows onto Kerr black holes were performed by varying a_* and ϕ. From these simulations, many snapshot images were produced by three general relativistic ray-tracing codes. The snapshot images show the black hole with a variety of accretion flows and jets. They depict a yellow-orange asymmetric bright ring around a central dark shadow (a crescent). As we saw in Section 2.1.2, a perceptually uniform colourmap was used. Since each GRMHD model simulation can be used to describe several different physical scenarios, the Simulation Library was used to generate more than 420 different physical scenarios. Each scenario was then being used to generate a library of more than 60,000 synthetic snapshot images (Doeleman, 2019a; EHTC 2019d, pp. 8–9, p. 14).

The comparisons between the snapshots and the data were performed with a single data set, namely the 2017 EHT April 6 data set because of its largest number of scans. The 2017 observations were scheduled as a series of scans and model-fitting requires a large number of scans. The number of scans obtained of M87's nucleus each night ranged from seven on April 10 to twenty-five on April 6 (EHTC 2019a, pp. 3–4, p. 6, p. 10, p. 36; EHTC 2019e, p. 8).

To test each GRMHD model (SANE and MAD) quantitatively against the 2017 EHT April 6 data set and find the best-fit parameters F_ν, θ_g and

PA for each snapshot image, images were generated from each model at each of $R_{high} = 1$, 10, 20, 40, 80, and 160. Two independent software packages were employed: THEMIS using a DE-MCMC sampler and GenA using a DE algorithm for minimizing χ^2_ν. Recall from Section 2.1.2 that THEMIS produces Bayesian posterior estimates for the three parameters F_ν, θ_g and PA. In the course of minimizing χ^2_ν, the three parameters F_ν, θ_g, PA and the gains at each VLBI station are varied. Variations in the three parameters F_ν, θ_g, PA approximately correspond to variations in the accretion rate, black hole mass, and orientation of the black hole spin, respectively. Each parameter affects the structure of the snapshot image differently. In fitting GRMHD snapshot images to data, θ_g is adjusted (i.e. the mass to distance ratio $\frac{M_{BH}}{d}$ is changed) by stretching the image, F_ν is varied by rescaling images from the standard parameters in the Simulation Image Library, and PA is changed by rotating the image. Varying the parameters a_*, ϕ and R_{high} can change the width and asymmetry of the photon ring and introduce additional structures exterior and interior to the photon ring (EHTC 2019e, p. 5; EHTC 2019f, p. 14).

But the χ^2_ν –test provided a fit with a minimum $\chi^2_\nu = 1.75$ over the entire set of more than 60,000 individual images in the Simulation Library. THEMIS-AIS was subsequently used. The models were tested by asking whether they were consistent with the data according to THEMIS-AIS. After performing the AIS test, only very few models were rejected. What that means is that the majority of the simulations could explain the 2017 EHT image. That is, many of the GRMHD models were acceptable and fitted the 2017 EHT data. At this stage, therefore, the experiment revolved around the interplay among SANE and MAD models and different parameters F_ν, θ_g, PA, i, R_{high}, ϕ, a_*. Even the most advanced computer algorithms were helpless in the face of choosing between the different SANE and MAD models. The ensuing steps required narrowing down the range of models by imposing three constraints (EHTC 2019e, pp. 5–6; EHTC 2019f, p. 10; Black Hole Initiative, 2019).

In his book *Laws and Symmetry*, van Fraassen deals with the relation between the constraints and the posterior probability. He argues that our posterior opinion is dependent on the imposed constraints. The prior probability, opinion, and the constraint change a conditional probability (van Fraassen, 1989, pp. 340–342).

The EHT Collaboration adopted priors called *Astronomer's priors* and imposed three constraints:

1) Radiative equilibrium: the radiative efficiency η of accretion flows is given by: $\eta = \frac{L_{bol}}{\dot{M}c^2}$, where L_{bol} stands for the bolometric luminosity emitted by the accretion flow (luminosity at all wavelengths). $\dot{M}c^2$ represents the rate at which rest mass energy is accreted. The radiative efficiency η of disk accretion depends only on the black hole spin a_*. As a result of calculating η, only a few SANE and MAD models with low R_{high} were eliminated.
2) X-ray luminosity constraint: the EHT collaboration simultaneously observed M87 with the Chandra x-ray observatory and the Nuclear Spectroscopic Telescope Array (NuSTAR). X-ray luminosity contemporaneous Chandra and NuSTAR data from these observations likewise rejected only a few models. That is, imposing the above constraint, it was found that only a few SANE and MAD models with low R_{high} were inconsistent with the observed X-ray luminosity (EHTC 2019e, pp. 11–13).
3) Minimum jet power: based on previous measurements, M87's jet power P_{jet} was assumed to be large: $P_{jet} > 10^{42}_{erg\ s^{-1}}$ (EHTC 2019e, p. 13). Van Fraassen points out that if the constraint is "simply a new certainty", then "it is indeed as if experience has simply handed us some proposition E on a platter – a 'total new evidence' – and has spoken as if with the voice of an angel" (van Fraassen, 1989, p. 321). Unfortunately, this does not hold for the P_{jet} condition. The estimate of P_{jet} is based on both physical uncertainties in the models used to estimate P_{jet} and uncertainties in the observations. But it was the jet power constraint that rejected the largest number of models. That is because the behaviour of jets and accretion flows is studied by GRMHD simulations, which should produce the same P_{jet}. P_{jet} is correlated with the black hole spin a_*. GRMHD models with $P_{jet} < 10^{42}_{erg\ s^{-1}}$ and zero spin $a_* = 0$ were outright rejected because non-spinning black holes cannot produce sufficiently powerful jets.

The EHT Collaboration arrived at two conclusions:

1) Models with low values of R_{high} should be eliminated. Most of the SANE models were therefore rejected and MAD models with small R_{high} failed. MAD models launch jets powered by the black hole spin a_* with wide opening angles and large P_{jet}. Three members of the EHT Collaboration conjectured that since the measured P_{jet} is large and the opening angle observed in M87's jet is wide, then "there is reason to suspect that M87 has an MAD at its core". That is, given the P_{jet} constraint, the probability of an MAD model is high (EHTC 2019e, p. 13; Chael et al., 2019, p. 2874).
2) In models that produce sufficiently powerful jets and are consistent with the 2017 EHT data, P_{jet} is driven by the extraction of black hole spin energy through the Blandford-Znajek process (EHTC 2019e, p. 5, pp. 11–15, p. 19; Doeleman, 2019a). When a rotating black hole is threaded by magnetic field lines, rotational (spin) energy is extracted from the rotating Kerr black hole: $P_{jet} \propto \Phi^2_{BH}\Omega^2$, where Ω^2 is the angular velocity of the horizon (Black Hole Initiative, 2019).

M87's P_{jet} is therefore inconsistent with a compact object with a surface (CS) (EHTC 2019f, p. 22). Alternative hypotheses to the Kerr black hole model had been checked before the EHT experiment was performed. These alternatives are compact objects that have no horizon but have a surface (such as very dense neutron stars and supermassive stars). The choice between a CS and a Kerr black hole is weakly underdetermined by evidence because a CS is not in principle unobservable. Unlike a hypothetical observer near a black hole, an observer standing just outside of a CS would be able to observe it. But by the time the CS's radiation would reach a distant observer on Earth, the CS would become practically unobservable. A black hole and a CS are therefore only empirically equivalent for distant observers.

As a matter of fact, even after imposing the P_{jet} constraint there were still many snapshot images that fitted the data. The EHT Collaboration did one more thing. They narrowed down the number of models that fitted the data as follows. The SANE and MAD snapshot images were divided into two groups: the *spin-away* models, the black hole's spin a_* points away from Earth ($i > 90°$ and $a_* > 0$, or $i < 90°$ and $a_* < 0$), and the *spin-toward* models, the black hole's spin a_* points toward Earth ($i > 90°$ and $a_* < 0$ or $i < 90°$ and $a_* > 0$). In each group, the accretion flow either moves with the black hole's spin or against it (*prograde* and *retrograde*). If the accretion flow's angular momentum and that of the black hole J are aligned, the accretion disk is *prograde* ($a_* \geq 0$) with respect to the black hole spin axis. On the other hand, if the black hole's angular momentum J is opposite that of the accretion flow, the accretion disk is *retrograde* ($a_* < 0$).

The approaching jet's PA was found by the fitting procedure (THEMIS and GenA) to be consistent with the spin-away models. Chael explains that applying the right-hand rule in physics and taking into consideration the approaching jet's PA, the two chosen cases (prograde and retrograde) are the spin-away models, the ones in which the spin of the black hole a_* must be pointing into the page and it is always going into the page. And the bright section is at the bottom part of the ring. The EHT Collaboration concluded that the black hole's spin a_* is moving clockwise. The alternative GRMHD model images having a bright section of the ring at the top – the spin-toward models, a_* is moving anticlockwise – were rejected. The snapshot images with the bright section of the ring at the bottom – the spin-away models – captured the

qualitative features found in the 2017 EHT 2017 April 6 image.

The EHT Collaboration explained that the ring is brighter at the bottom because the plasma is moving toward us. Due to Doppler beaming, at the top, the ring is less bright because the material is moving away from us. While the approaching side of the plasma of the forward jet is Doppler boosted, the receding side is Doppler dimmed, producing a surface brightness contrast.

The EHT Collaboration placed three images one next to the other: A (spin-away) GRMHD model snapshot image of what we would expect to obtain if the EHT were able to have an infinite resolution; a blurred (spin-away) GRMHD model snapshot image (blurred to the angular resolution of the EHT array, $20\mu as$); and the (blurred) 2017 EHT April 6 image. The reader of the fifth letter of the EHT Collaboration can see for themselves that the two blurred images are the same (EHTC 2019a, p. 6; EHTC 2019e, p. 2, pp. 5–6, p. 8, p. 10, p. 15; Black Hole Initiative, 2019).

3.2.2. Coincidence and mass estimation

The EHT Collaboration writes: "The four EHT observations of M87* in 2017 spanned about a week, corresponding to a timescale of ~15 M. With such a short time span, we cannot exclude a transient origin for the source morphology using 2017 data alone. However, such a feature would need to attain the geometry and apparent size expected of a shadow of a black hole (of independently measured mass-to-distance ratio [θ_g]) through an unusual coincidence" (Wielgus et al., 2020, p. 19). The EHT Collaboration set itself to calculate the mass M_{BH} of the core of M87. For this purpose, three independent different methods were used. M_{BH} (and θ_g) was then shown to correspond to a previous value arrived at by yet another different method. The three different methods are:

1) Geometric crescent model fitting: Two kinds of crescent models were developed (xs-ring and xs-ringauss) and compared directly to the 2017 EHT data. The crescent models were collectively named, the generalized crescent model (GC). The xs-ring model was compared to data using dynesty while the xs-ringauss was fitted to data by THEMIS. The GC model, however, does not provide any scientific explanation for the 2017 EHT data because no underlying mechanism based on physics is responsible for the structures in the GC model. The parameters of the GC model were therefore calibrated, i.e. fitted to the parameters of the GRMHD models. The crescent models were associated with the emission surrounding the shadow of a black hole and the first physical parameter was obtained, equation (9). If the geometric crescent is formed by gravitational lensing, then its diameter obeys the equation:

$$\theta_d = \alpha\theta_g, \tag{10}$$

where, α represents the gravitational lensing factor.

Both THEMIS and dynesty produced values of the crescent diameter of $\theta_d \approx 42 \pm 3\mu as$ for the two geometric crescent models. Those values were then compared with the known value of θ_g that went into the GRMHD simulations. This led to two mean values of α for the two crescent models. The two θ_d measurements were combined with the two α values to arrive at values of θ_g using equation (10) (Doeleman, 2019a; EHTC 2019f, pp. 10–13).

2) GRMHD model fitting: Variations in M_{BH} are associated with changes in the diameter of the photon ring θ_d, a generic feature found across all of the images in the GRMHD image library, see equation (12) below. THEMIS and GenA produced a value of θ_d that is consistent with the estimate obtained by fitting crescent models to data (EHTC 2019f, p. 14).

3) Image domain feature extraction: Imagine that hot gas surrounds the black hole. The black hole casts a shadow on that background. Recall from Section 2.1.2 that in 1973 Bardeen calculated "the precise apparent shape of the 'black hole'" (shadow) seen by a distant observer. He showed that the diameter of the apparent shadow depends on the gravitational lensing around the black hole. For a Schwarzschild black hole: $D_{sh} \approx 2 \bullet 3^{3/2} M_{BH} \approx 10.39 M_{BH}$, where $2 \bullet 3^{3/2} = 2\sqrt{27}$ is the gravitational lensing factor α. "However, the effect of the frame dragging induced by the angular momentum of the Kerr black hole is quite apparent" (Bardeen, 1973, p. 231, p. 233). According to models, in the vicinity of a rotating Kerr black hole, effects of frame dragging due to black hole rotation act to compress the shadow with respect to the rotation axis while the quadruple moment of the rotating black hole causes an oblate shape of the shadow. The two effects approximately cancel each other out and we are left with a nearly circular shadow. Thus, for a Kerr black hole:

$$D_{sh} < 2\sqrt{27}R_g \approx 9.8R_g \text{ or } R_{sh} < \sqrt{27}R_g \approx 4.9R_g. \tag{11}$$

The radius R_g is given by equation (8). Combining equation (11) with equation (9) gives the angular diameter of the shadow θ_d:

$$\theta_d \approx \frac{2\sqrt{27}R_g}{d} \approx \frac{9.8R_g}{d} \equiv \frac{9.8GM_{BH}}{dc^2} = \alpha\theta_g. \tag{12}$$

Dominic Pesce, a member of the modelling group, explains feature extraction in a picturesque way. The ring diameter θ_d is estimated directly from the 2017 EHT image domain by an edge finding algorithm. A radius that is moved around different points of the ring measures the location of the peak brightness which changes as the ring is swept by the radius. Twice the distance to the peak brightness gives the ring diameter. The measurements are aggregated and a histogram is obtained whose mean gives a certain knowledge of the ring diameter. The spread in that probability distribution gives information about the uncertainty in the measurements.

Here is a more technical explanation. The following parameters were measured in the image domain of the 2017 EHT April 11 fiducial images from each imaging pipeline (DIFMAP, SMILI and eht-imaging): the ring diameter θ_d, the width, the asymmetry, the orientation angle, and the fractional central brightness (the ratio of the mean brightness interior to the ring to the mean brightness around the ring). The parameters were estimated from the 2017 EHT images for each of the four observed days. Unlike the SMILI and eht-imaging images, the DIFMAP images were restored with a Gaussian $20\mu as$ beam. DIFMAP therefore produced larger measured widths. But the measured image domain ring diameters θ_d were consistent among all imaging pipelines. Across all days, the DIFMAP images recovered $\theta_d = 40 \pm 3\mu as$ when restored with a $20\mu as$ beam. SMILI and eht-imaging recovered $\theta_d = 41 \pm 3\mu as$.

All three different techniques yielded $\theta_d \approx 42\mu as$ consistent with a mass of $M_{BH} \approx 6.5 \pm 0.7 \times 10^9$ solar masses ($\odot$) (EHTC 2019d, pp. 27–30; EHTC 2019e, p. 2; EHTC 2019f, pp. 4–6, pp. 7–20, p. 31; Doeleman, 2020b; Black Hole Initiative, 2019).

The EHT Collaboration points out that prior measurements of M_{BH} (based on stellar dynamics) are consistent with the above estimations of M_{BH}. Two teams of scientists obtained estimates of the mass of the core of M87. A team of scientists led by Jonelle Walsh used the Space Telescope Imaging Spectrograph (STIS) from the Hubble Space Telescope (HST) to study the emission lines from the motion of gas dynamics around the center of M87 (Walsh et al., 2013). A team led by Karl Gebhardt modeled the observed stellar dynamics (dispersion of stellar velocities around the center of M87) and measured the mass from the Near-infrared Integral Field Spectrograph (NIFS) data (representing infrared spectroscopic observations of stellar emissions) (Gebhardt et al., 2011). Significant discrepancies between the results obtained by the two methods were found. The gas dynamics led to $M_{BH} \approx 3.5 \times 10^9 \odot$ while the stellar dynamics gave almost double that value, $M_{BH} \approx 6.6 \times 10^9 \odot$. Both Walsh et al. and Gebhardt et al. measured a mass-to-distance ratio θ_g rather than directly M_{BH} (EHTC 2019f, p. 39).

The EHT Collaboration found "a striking level of consistency across all measurement methods" converging on the same value of M_{BH} and arrived at two key conclusions:

1) The consistency between stellar dynamics measurements of M_{BH} and EHT measurements of M_{BH} "strongly support the hypothesis that the central object in M87 is indeed a Kerr black hole". According to Hacking's *Argument from Coincidence*, experimenters measuring the same mass M_{BH} using different kinds of methods, have excellent reasons for saying, that is real. Unlike masses of black holes, however, experimenters according to Hacking measure the properties of micro-entities and only have excellent reasons for saying that electrons and other micro-entities are real. But "A black hole is as theoretical an entity as could be" (Hacking, 1989, p. 561).
2) "The mutual agreement that we see among the multiple measurements indicates that the θ_g value we have converged upon is robust" […] (EHTC 2019f, p. 9, p. 21, p. 23). Statistically robust is defined as follows: an estimate is robust if it is drawn from different kinds of distributions. The mass M_{BH} was calculated with five independent methods – geometric crescent model fitting, GRMHD model fitting, image domain feature extraction, and stellar and gas dynamics measurements – four of which led approximately to $M_{BH} \approx 6.5 \times 10^9 \odot$. Estimations based on stellar dynamics and gas dynamics do not agree. Suppose we conclude that the gas dynamics measurements produced a value of the mass with a distortion: $M_{BH} \approx 3.5 \times 10^9 \odot + X \approx 6.5 \times 10^9 \odot$. The experiment which gave a wrong estimation is thought to be not all that wrong, so that it can be expected to give an estimation of the mass with just a little distortion. That is why the bad method gives qualitatively similar results to the other methods. Now all five methods agree. We find an astrophysical explanation for why all five methods give qualitatively similar results. Cartwright explains that when that picture is cashed out into an argument, *robustness* across alternative and contradictory implementations is a kind of *reproducibility* "that argues for the truth of the outcome" (Cartwright, 1991, p. 154).

What the members of the EHT Collaboration did was exactly the opposite. They subsequently showed that the gas dynamics measurements produced a wrong result. Recall that the EHT Collaboration asked two questions: Is the core of M87 an $M_{BH} \approx 3.5 \times 10^9 \odot$ black hole or an $M_{BH} \approx 6.5 \times 10^9 \odot$ black hole? Is it a black hole at all? (see Section 2.1.2). According to equation (11), for a $3.5 \times 10^9 \odot$ black hole we would get a smaller shadow than for a $6.5 \times 10^9 \odot$ black hole, both compact objects yield $R_{sh} \approx 4.9 R_g$ (the experimental interpretation of the *no-hair theorem*). According to computer simulations, a $6.5 \times 10^9 \odot$ rotating wormhole has half as big a shadow radius as a $6.5 \times 10^9 \odot$ black hole: $R_{sh} = 2.7 R_g$, which is about the same size as the shadow of the $3.5 \times 10^9 \odot$ black hole.[8] A $6.5 \times 10^9 \odot$ naked singularity has an even smaller shadow radius of $R_{sh} = R_g$. Members of the EHT Collaboration put the 2017 EHT image on top of these different shadow sizes and ruled out these exotic possibilities. They were "left with a ring that nearly perfectly matches that of the stellar dynamics measurements of the black hole" mass (EHTC 2019e, p. 18; Bouman, 2019; Johnson, 2020).

Finally, Bardeen's equation $D_{sh} \approx 2\sqrt{27} M_{BH}$ [equation (11)] rings a bell for those who are familiar with the fable about van Fraassen's tower. Van Fraassen told a story about a tower and a shadow, the length of which is proportional to the height of the tower (van Fraassen, 1980, pp. 132–134). The diameter of the shadow D_{sh} is proportional to the mass M_{BH} of the Kerr black hole, thereby a black hole of mass $M_{BH} \approx 6.5 \times 10^9 \odot$ is the reason why the shadow has a certain diameter D_{sh}. But along the lines of van Fraassen's scientific agnosticism and interpretation of images (see Sections 2.1.4 and 3.1.2), one neither believes nor disbelieves in the existence of Kerr black holes and the 2017 EHT *Consensus image* is either an illusion or an image of a thing.

[8] In 1998 Edward Teo obtained an exact stationary axisymmetric solution of the Einstein field equations describing a rotating traversable wormhole (Teo, 1998).

4. Concluding remarks

This paper discusses some philosophical aspects related to the recent publication of the 2017 EHT experimental results.

In Chapters 2 and 3, I meticulously analysed the EHT experiment through the prism of the philosophy of experimentation of Hacking, van Fraassen, Cartwright and Galison. In the EHT experiment, images were reconstructed from observations and models were fitted to observations by cutting-edge algorithms, several of which were purpose-developed for the experiment.

The EHT collected sparse and noisy data. The EHT Collaboration was therefore faced with the need to develop sophisticated new imaging pipelines that would also perform self-calibration. The imaging pipelines would alternate between imaging and self-calibration and also manipulate images and blur them. In the course of fitting models to observations, images were virtually manipulated. What does that mean? Parameters are adjusted and this, in turn, affects the structure of the image: changing one parameter stretches the image; variation in another, causes rescaling and changing another parameter rotates the image. Data are likewise manipulated, processed and calibrated. Although the above manipulation does not conform to Hacking's definition of manipulation, in this paper I showed that the black hole experiment is consistent with major elements from Hacking's list of elements of laboratory science.

The EHT Collaboration used three separate calibration pipelines, three different imaging pipelines, and three independent methods to estimate the mass of the black hole. Great efforts were made to get around the human bias by applying a *coincidence* strategy to identify the intrinsic features in the images. In order to avoid introducing human bias into the imaging process, the EHT Collaboration construed *Coincidence/Reproducibility* as *Technological Agnosticism*. *Technological Agnosticism*, however, does not provide justification for believing in the existence of unobservable entities.

As the EHT Collaboration points out, the crescent morphology is robust among the three different pipelines. Cartwright argues that if we say that the results are robust it does not mean that there must be some truth in them. I will end this paper with Bouman's words: "One question I get a lot is: Did you prove Einstein was right? Well I like to say: the short answer is no, we didn't. But we also didn't prove he was wrong, which is a pretty deal too" (Bouman, 2020c).

Acknowledgments

I want to acknowledge the funding for this study provided by ERC advanced grant number 834735. I would like to thank Gil Kalai for helping with this research project. I thank the anonymous referee for their useful suggestions and I thank James Ladyman for his insightful comments. This paper is based on my essay "Observation and Simulation: The First Black Hole Image" which earned me a honourable mention in the "2019 Philosophy of Cosmology Essay Competition", organized by the project "New Directions in Philosophy of Cosmology", funded by the John Templeton Foundation. I would like to thank Chris Smeenk and James Weatherall for granting me this award and Deborah Fox for her administrative help.

References

Bardeen, J. M. (1973). Timelike and null geodesics in the Kerr metric. In C. DeWitt, & B. S. DeWitt (Eds.), *Black holes Les Astres Occlus* (pp. 216–239). New York: Gordon and Breach Science Publishers.

Black Hole Initiative. (2019). *Black hole initiative science center talk on discovery of black hole*. Lecture at Harvard University. May 30.

Blandford, R. D., & Znajek, R. L. (1977). Electromagnetic extraction of energy from Kerr black holes. *Monthly notices of the royal astronomical society, 179*, 433–456.

Bouman, K. (2017). *How to take a picture of a black hole*. TED talk. April 28.

Bouman, K. (2019). Written testimony before the committee on science, space, and technology United States house of representatives, may 16. In *Hearing (2019)* (pp. 38–44).

Bouman, K. (2019). Imaging a black hole with the event horizon telescope. In *A lecture at the stanford center for image system engineering*. Stanford University. April 17.
Bouman, K. (2020a). The inside story of the first picture of the black hole. *IEEE Spectrum*. January 30.
Bouman, K. (2020b). *"Sensing to intelligence: From human body to black holes." Lecture in the Sensing to intelligence: From human body to black holes in space. 83rd Annual Seminar Day. Caltech seminar day session II. May 16*.
Bouman, K. (2020c). *"Computational imaging SPACE webinar series: Katie Bouman, caltech." lecture in sps-space*. June 30.
Bouman, K. L., Johnson, M. D., Zoran, D., Fish, V. L., Doeleman, S. S., & Freeman, W. T. (2016). Computational imaging for VLBI image reconstruction. In *Lecture in 2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 913–922).
Broderick, A. E., Gold, R. G., Karami, M., et al.(Event Horizon Telescope Collaboration). (2020). Themis: A parameter estimation framework for the event horizon telescope. *The Astrophysical Journal, 897*, 1–38.
Cartwright, N. (1991). Replicability, reproducibility, and robustness: Comments on harry collins. *History of Political Economy, 23*, 143–145.
Chael, A. A. (2020). *"The event horizon telescope photographing a black hole." Lecture in scientific computing with Python virtual conference, july 6-12, 2020*.
Chael, A. A., Narayan, R., & Johnson, M. D. (2019). Two-temperature, Magnetically Arrested Disc simulations of the jet from the supermassive black hole in M87. *Monthly Notices of the Royal Astronomical Society, 486*, 2873–2895.
Doeleman, S. (2017). *EHT status update, June 30 2017. EHT website*.
Doeleman, S. (2019a). *"The event horizon telescope: Seeing the unseeable." lecture in the colloquium at harvard university*.
Doeleman, S. (2019b). 2020 breakthrough prize winner shep Doeleman – how EHT imaged a black hole. *Talk in World Science Festival*. September 6, 2019.
Doeleman, S. (2019c). *Inside the black hole image that made history*. TED talk. May 10.
Doeleman, S. (2020a). *"Seeing the unseeable: Capturing an image of a black hole." Free science lecture, virtual April meeting*. American Physical Society. April 18.
Doeleman, S. (2020b). *"PSW 2421 the event horizon telescope. First ever images of a black hole." The 89th Joseph Henry lecture at the philosophical society of Washington science*.
Event Horizon Telescope Collaboration. (2019a). First M87 event horizon telescope results. I. The shadow of the supermassive black hole. *The Astrophysical Journal Letters, 875*(L1), 1–17.
Event Horizon Telescope Collaboration. (2019b). First M87 event horizon telescope results. II. Array and instrumentation. *The Astrophysical Journal Letters, 875*(L2), 1–28.
Event Horizon Telescope Collaboration. (2019c). First M87 event horizon telescope results. III. Data processing and calibration. *The Astrophysical Journal Letters, 875* (L3), 1–32.
Event Horizon Telescope Collaboration. (2019d). First M87 event horizon telescope results. IV. Imaging the central supermassive black hole. *The Astrophysical Journal Letters, 875*(L4), 1–52.
Event Horizon Telescope Collaboration. (2019e). First M87 event horizon telescope results. V. Physical origin of the asymmetric ring. *The Astrophysical Journal Letters, 875*(L5), 1–31.
Event Horizon Telescope Collaboration. (2019f). First M87 event horizon telescope results. VI. The shadow and mass of the central black hole. *The Astrophysical Journal Letters, 875*(L6), 1–44.
Falcke, H., Melia, F., & Agol, E. (2000). Viewing the shadow of the black hole at the galactic center. *The Astrophysical Journal, 528*, L13–L16.
van Fraassen, B. C. (1980). *The scientific image*. Oxford: Oxford University Press.
van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.
van Fraassen, B. C. (1998). The agnostic subtly probabilified. *Analysis, 58*, 212–220.
van Fraassen, B. C. (2001). Constructive empiricism now. *Philosophical Studies, 106*, 151–170.
Galison, P. (1988). Philosophy in the laboratory. *The Journal of Philosophy, 85*, 525–527. Eighty-Fifth Annual Meeting American Philosophical Association, Eastern Division (October 1988).
Galison, P. (2019). *"Philosophy of the shadow." CMSA YIP lecture, black hole initiative colloquium at*. Harvard University. April 18.
Gebhardt, K., Adams, J., Richstone, D., Lauer, T. R., Faber, S. M., Gültekin, K., Murphy, J., & Tremaine, S. (2011). The black-hole mass in M87 from gemini/NIFS adaptive optics observations. *The Astrophysical Journal, 729*, 1–13.
Hacking, I. (1981). Do we see through A microscope? *Pacific Philosophical Quarterly, 62*, 305–322.
Hacking, I. (1982). Experimentation and scientific realism. *Philosophical Topics, 13*, 71–87.
Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge: Cambridge University Press.
Hacking, I. (1988). On the stability of the laboratory sciences. *The Journal of Philosophy, 85*, 507–514. Eighty-Fifth Annual Meeting American Philosophical Association, Eastern Division (October 1988).
Hacking, I. (1989). Extragalactic reality: The case of gravitational lensing. *Philosophy of Science, 56*, 555–581.
Hacking, I. (1992). The self-vindication of the laboratory sciences. In A. Pickering (Ed.), *Science as practice and culture* (pp. 29–64). Chicago, IL: The University of Chicago Press.
Hacking, I. (2002). *Historical ontology*. Cambridge, Massachusetts: Harvard University Press.
Hacking, I. (2014). *Why is there philosophy of mathematics at all*. Cambridge: Cambridge University Press.
Hearing (before the Committee of Science, Space, and Technology House of Representatives one Hundred sixteenth Congress). (2019). In *First session. Event horizon telescope: The black hole seen round the world*. Washington: U.S. Government Publishing Office. May 16, 2019.
Hitchcock, C., & Sober, E. (2004). Prediction versus accommodation and the risk of overfitting. *The British Journal for the Philosophy of Science, 55*, 1–34.
Holland, J. H. (1973). Genetic algorithms and the optimal allocation of trails. *SIAM Journal on Computing, 2*, 88–105.
Johnson, M. D. (2020). Photographing a supermassive black hole with the event horizon telescope. In *Lecture at black hole initiative seminar*. Harvard university. April 12.
Kerr, R. (1963). Gravitational field of a spinning mass as an example of algebraically special metrics. *Physical Review Letters, 11*, 237–238.
Luminet, J.-P. (1979). Image of a spherical black hole with thin accretion disk. *Astronomy and Astrophysics, 75*, 228–235.
Michelson, A. A., & Pease, F. G. (1921). Measurement of the diameter of α Orionis with the interferometer. *The Astrophysical Journal, 53*, 249–259.
Millikan, R. A. (1911). "The isolation of an ion, a precision measurement of its charge, and the correction of Stokes's law. *Physical Review, 32*, 1–58.
Morrison, M. (2015). *Reconstructing reality: Models, mathematics, and simulations*. Oxford: Oxford University Press.
Penrose, R. (1969). Gravitational collapse: The role of general relativity. *Rivista del Nuovo Cimento, Numero Speziale, 1*, 252–275. General Relativity and Gravitation 34, 2002, pp. 1141–1165.
Porth, O., Chatterjee, K., Narayan, R., et al.Event Horizon Telescope Collaboration. (2019). The event horizon general relativistic magnetohydrodynamic code comparison project. *The Astrophysical Journal*, 1–40. Supplement Series 243.
Psaltis, D., & Doeleman, S. S. (2015). The black hole test. *Scientific American, 313*, 74–79.
Psaltis, D., Medeiros, L., Christian, P., et al.(Event Horizon Telescope Collaboration). (2020). Gravitational test beyond the first post-Newtonian order with the shadow of the M87 black hole. *Physical Review Letters, 125*, 141104-1-141104-9.
Psaltis, D., Özel, F., Chan, C.-K., & Marrone, D. P. (2015). A general relativistic null hypothesis test with event horizon telescope observations of the black hole shadow in sgr A*. *The Astrophysical Journal, 814*, 1–14.
Putnam, H. (1984). Guilty statements. *London Review of Books, 6*, 5. May 3.
Schwarzschild, K. (1916). Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. In *Über das Gravitationsfeld einer Kugel aus inkompressibler Fliissigkeit nach der Einsteinschen Theorie* (pp. 189–196). (Berlin). Sitzungsberichte: Königlich Preußische Akademie der Wissenschaften. pp. 424–435.
Sądowski, A., Narayan, R., Tchekhovskoy, A., & Zhu, Y. (2013). Semi-implicit scheme for treating radiation under M1 closure in general relativistic conservative fluid dynamics codes. *Monthly Notices of the Royal Astronomical Society, 429*, 3533–3550.
Teo, E. (1998). Rotating traversable wormholes. *Physical Review D, 58*.
Walsh, J. L., Barth Aaron, J., Ho, L. C., & Sarzi, M. (2013). The M87 black hole mass from gas-dynamical models of space telescope imaging Spectrograph observations. *The Astrophysical Journal, 770*, 1–11.
Wheeler, J. A. (1981). The lesson of the black hole. *Proceedings of the American Philosophical Society, 125*, 25–37.
Wielgus, M., Akiyama, K., Blackburn, L., et al.(Event Horizon Telescope Collaboration). (2020). Monitoring the morphology of M87* in 2009-2017 with the event horizon telescope. *The Astrophysical Journal, 901*, 1–28.