"Inchoate Intuitions": A Constructive Empiricist Analysis of Black Hole Dynamics and Entropy

Nicholas Meadowcroft-Lunn

MPhil

The University of York

Philosophy

July 2024

Abstract

This thesis maintains that scientific realist conceptions about science actively hamper the potential of modern physics to be useful and productive by encouraging a perception that science's legitimacy is founded in giving us "truths" about the world, as well as leading physicists to do work that can't ever be empirically grounded. To replace this realist bias, I seek to make the case that if physicists adopt Bas van Fraassen's constructive empiricism as a framework for understanding why they do science, they will recover more usefulness from their physics practice. I test this conception in case studies in still unresolved physics problems, focusing on the Page-time paradox and Maxwell's Demon. I also make the case that physics itself often implies the superiority of constructive empiricism over realism, as well as make novel contributions to the constructive empiricist framework by extending van Fraassen's understanding of epistemic communities and observability heuristic. This work starts and ends by discussing the Page-time paradox, initially introducing it as an example of the style of problem that has its roots in misplaced realist intuitions. The thesis takes the example of the Page-time paradox and ends by using constructive empiricism, along with my novel extensions outlined in the thesis, to present pathways via which a working physicist could solve or ignore the paradox in ways the realist couldn't. By doing this, I state that the working physicist should adopt constructive empiricism to better allow them to do useful and productive work in physics.

Table of Contents

To Tom,

I hope we can discuss this work again

Authors Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

I acknowledge that I have received assistance from friends to proofread this thesis in line with the Policy on Transparency in Authorship in PGR Programmes.

Introduction – Some Thoughts on Physics

Physics knowledge and study is essential to the functioning of society because all our interactions with the external world are governed by physical laws. As I type this out on my computer, I am utilising the physical world via the action of my fingers on my keyboard, all the way through to the electrons passing through logic gates that cause photons to light up my screen with the correct symbols. There is no avoiding what is in essence a trivial observation: we are physical beings.

Because of this, the study of physics is one of the root ways we try to model and make useful the relationship between us and the world around us. Newtonian gravity is the description of a relationship between different bodies as they interact with each other that (at macroscopic, non-relativistic limits) explains not "why" an apple falls to the ground, but the velocity, acceleration, and path which it falls. At a fundamental level, physics doesn't need explanatory power to be powerful and useful, it simply needs to match the experiential data we can acquire through the observations we make as humans.

Physics is also an empirical subject. The process by which we acquire knowledge about the universe in physics comes via the production of theories that are either accepted or rejected via the way they model, describe, or predict the dynamics of systems around us. What metaphysical power you assign this knowledge acquisition and the consequences of what it is to accept a theory acceptance form the dividing lines between what makes someone in essence a realist[1] or an anti-realist[2]. However, I maintain that all analyses of what science is and does have at least a commonly acknowledged central point, a common denominator of purpose: physics practice and methodology derives from an empirical substructure of reasoning, where proof for any theories' successes are found in the observations they predict being present to us in the external world.

If we take this lowest common denominator, that doing successful science at least involves describing reliably the relationships we can observe, what does that mean for how physicists should do physics? Is there a "best approach" out there, by which physicists can achieve this simplest goal most efficiently and usefully? This thesis proposes that we can find such an

[1] There are many different forms of realism, but I take it to be the belief that by accepting a theory we believe it is at least approximately true about what exists in the universe (Chakravartty, 2017, §. 1.1)
[2] Which has as many different forms as realism does, but I define as fundamentally involving a negation of realism on semantic or epistemological grounds (Chakravartty, 2017, §. 4.1)

approach, and that that will allow us to help answer the further question: do realist intuitions about physics cause problems for physics and the physicist, and what can we do to fix them?

What's the point of this thesis?

In asking this question, it is important to outline why it would be attractive for physics to optimally model the world around us. In doing so, I begin with the central conception that we do physics, and all forms of human knowledge acquisition more broadly, to better improve our lives and the lives of those around us. That, as opposed to more philosophical or theoretical goals, is the purpose of all structured thinking. After all, what is the point of thinking about anything unless we can connect that directly to useful advances in our lived experiences? From this point, I maintain that realist intuitions and beliefs cause physicists to lose track of the purpose of physics: to improve our lives. Using case studies from black hole dynamics and statistical mechanics, I will show that realist intuitions and an overburdening desire to find truth can be shown to hinder physics research for no long-term gain. The connection between the case studies I will be introducing, the inability to study their systems and dynamics empirically, helps show why adopting realist beliefs is a problem when progressing physics. It also hints at how physicists can do more productive future work if they adopt the appropriate philosophical frameworks.

This thesis also begins with the assumption that the ultimate measure of success in physics is pragmatic. Metaphysical or ontological questions about the reality of our theories are not as important as the simple question of "does theory x reliably predict y, where y is something, we can perceive with our own senses". I take as a central assumption that our physical theories should above all allow us to do more and improve our lives in ways that were before not possible. This metric, which is necessarily subjective at its limits, can be best understood by taking a look at the daily occurrences we as human beings experience in the world. The physical theories underpinning the mechanics of the internet, from electrodynamics to statistical mechanics, allow me to write this thesis from the comfort of my own computer and have access to infinitely more research and philosophical source material, all at a much lower barrier to entry. It has improved my ability to do this thesis and, thus, the theories that underpin the dynamics of the internet are useful. This also extends out to less obviously technical applications. My oven has a temperature dial, created by an electronic engineer, which uses the basic relationships between voltage, resistance and subsequent heat generation and transfer that allow me to cook different foodstuffs at different temperatures, allowing for me to have a more varied, better quality of dinner. The relationships inherent in the design decisions made when crafting that oven were discovered generations ago, but are still relevant and

useful, and the physical theories describing these relationships still go into the engineering that takes place when designing new and improved ovens. While a 'better quality of life' is not an objective metric, for the purposes of this thesis we will be using it as the beginning point, a proto answer to “what is the purpose of doing physics”, as it grounds physics most firmly to the real world surrounding us.

As we are using the fluid and subjective concept of usefulness, we must be careful and accept the inherent qualities of said usefulness: what appear at times to be deeply non-useful discoveries can in fact provide great theoretical leaps that lead into even greater practical application (from Michelson-Morley to the invention of GPS). Importantly, usefulness is not connected to whether our theories are true or explanatory. To produce theories that can provide for new, better, practical applications we don’t need to know why they work or that they are true, we simply need to know that they work and give predictions that align with experimental data.

There are opposing views of what physics can and should do. Since the 1960s we have moved within the philosophy of science toward a broad orthodoxy for scientific realism, the belief that our scientific theories are (at least) approximately true about the world, and that we must have ontological commitment to the entities within our theories. “Electrons exist” insofar as the theories of physics we have, that we accept, state that they do. This ontological set of commitments is designed to broadly to achieve many goals, of which I offer three I find persuasive:

- Give strong motivational grounds for studying physics in the first place. Usefulness can mean many different things to different people, or even lead to no further progress, citing how complete and fulfilling our lives already are. A realist outlook gives us a solid end-goal to reach, truth, which we can strive ever closer toward, motivating physicists on to new and better theories.
- Consistency of approach. As mentioned above, prioritising usefulness can mean many different things to many different people. Truth is a far less subjective concept, and allows for all of humanity to have, in theory, a single stick by which to measure the success or failure of their theories and produce a consistent canon of physical theories. Theory X is more predictive, more elegant and provides greater explanatory power than Theory Y, thus Theory X is closer to the truth, thus we must now accept Theory X.
- Explanatory Power. A classic argument in favour of realism is the No Miracles Argument. Concisely put, if all our theories seem to so naturally align with each other to produce a

> great cosmic blanket of consistency covering all areas of physics, surely that indicates that our theories point us closely to how the universe truly functions? It cannot be mere coincidence, to think the opposite would imply miraculous odds. Thus, we should be realists, as "the theories we have are true" is the best explanation for why all our theories work so consistently together.

We have a dilemma here. This thesis will show that realist orthodoxies exist within physics, and that they are the cause of internal incoherence and an inability to solve nascent problems. However, as is outlined above, there are solid, sensible, and rational reasons for being a realist; it is orthodoxy for a reason. What is required is a better philosophical approach than either something ardently realist or ardently anti-realist, an approach that allows physicists to maintain their intuitions whilst producing good physics, even if those intuitions are wrong about what physics can achieve. I think this approach, a conciliatory one, offers the most practical pathway forward as it will meet the least resistance within physics itself. These solutions offer little if they remain on the page and don't extend out into the real world.

Above all, any new approach we endorse should be grounded primarily in the pragmatics I begin this section by endorsing. The goal of this thesis isn't to argue outright that realism is a failed philosophy of science, or that there are no rational reasons to adopt realism. I have already provided three reasons why one could. Instead, the goal is to prove that realist biases and intuitions cause practical problems for physicists within certain contexts, which can be understood via analyses of the physics practice underpinning the work. The "new approach" I am advocating for must be grounded in empiricism, sensitive to the working practice of physics, and as removed as possible from the metaphysical concerns that keep philosophers awake at night, but don't necessarily trouble physicists.

For this potential new approach, I propose that physicists adopt Bas van Fraassen's (van Fraassen, 1980) constructive empiricism as a more pragmatically useful philosophy of science within the areas of physics the case studies consider, as it allows us to defuse the problems created and worsened by realist intuitions by contextualising them within the framework of physics needs to do to be successful and can do empirically. In doing so in Chapter 4 of this thesis, I introduce a novel expansion of van Fraassen's understanding epistemic communities. This expansion better accounts for edge cases in what counts as observable, which better aligns both with the theories of physics we have (and endorse) as well as achieve the constructive empiricist aim of achieving as modest an epistemology as is possible to do science.

The Structure of the Thesis

The thesis is structured into four separate chapters and a conclusion. The first, Chapter 1, introduces the Page-time paradox and some of the grounding concepts relevant to it within physics. This is offered to establish a baseline level of understanding of the physics so that all potential readers, be they philosophers or physicists, can engage with the later philosophical content within whilst understanding the context into which these arguments are offered. It also serves the function of allowing us to understand *why* these "paradoxes" are perceived as "paradoxes" within physics, as well as highlighting some of the underpinning philosophical undercurrents in the work. In the spirit of clarity, I will note here (and at the start of that section) that the work isn't always fully technical, a choice made to ensure legibility for a philosopher with no previous physics training.[3]

Chapter 2 offers the central argument of the thesis: realist intuitions compromise physicists in performing physics, these compromises are numerous and pernicious, and working physicists should view their work and theories through a constructive empiricist lens. Realist intuitions in the philosophy of science cause physicists to incorrectly understand the scope of their subject, attaching unneeded metaphysical significance to the entities and properties they use. These intuitions, in turn, cause "paradoxes" to appear and cloud research fields, making resolving these problems much more difficult than they otherwise need to be. Chapter 2 also introduces a protagonist actor, the "working physicist", who I use to validate the reasonableness of any new approaches taken. This character's desires and intuitions form the framework of how we judge these new approaches, as if we are not sensitive to them this thesis stands little chance of achieving its stated aim: providing a better framework for viewing physics that is acceptable to physicists that will rid them of unproductive investigations. The pathway Chapter 2 adopts is constructive empiricism, which allow us to remove unwanted metaphysical or ontological worries in a rigorous, well-defined framework that accounts correctly for the behaviours, actions, and motivations of our "working physicist". To this end, the latter half of Chapter 2 offers a description of constructive empiricism, a defence against critics of the school, and reasons why our "working physicist" should consciously view physics and its theories using the framework offered by constructive empiricism.

[3] This occasional lack of technical accuracy is an unfortunate necessity. To give a full account of black hole dynamics and their associated information loss paradoxes would be a thesis unto itself, therefore I have given the best, brief outline of one problem and why it generates whilst maintaining conceptual accuracy. The goal is to provide adequate grounding for later philosophical claims, not give a history of black hole physics.

Chapter 3 and 4 both offer accounts of problems in physics that have confounded resolution and offers philosophical designed solutions or reframings that allow our "working physicist" to either exorcise or discount them as problems. They also both discuss how physics already points us in the direction of constructive empiricism and away from realism.

Chapter 3 focuses on two key themes: that the value-laden stances present in constructive empiricism are already hinted at by current understandings of physical properties (in this case entropy), and that a long problematic paradox at the core of statistical mechanical thought (Maxwell's Demon) can be exorcised via simple epistemic reframing of the issue, something granted to us naturally if we adopt constructive empiricism that the realist has to work to explain.

Chapter 4 considers similar themes in the paradigm of black hole physics, where again entropy is a dominant and issue-causing concept. This chapter also novelly extends van Fraassen's initial conception of what an "epistemic community" is, making it more practically useful in the context of both black holes, and also in any context where an absolute, natural gap in information transfer can occur within physics (e.g., crossing the event horizon of a black hole). Similarly to Chapter 3, Chapter 4 ends by considering the specific paradox introduced in Chapter 1, the Page-time paradox, and offers a selection of possible pathways the working physicist could take to either exorcise or solve the paradox. This will also show that the realist case is always inferior to the constructive empiricist case when considering the Page-time paradox. Realism's inability to understand the limits of physics correctly and accurately leads to unintuitive, confusing, and unresolvable problems that we simply do not have to concern ourselves with.

The conclusion to this thesis addresses the main question of future research and the general applicability of the ideas. This thesis is limited in both its scope and the definitive conclusions it is willing to offer; an entire academic career could be given over to thinking about all possible conclusions of this line of thinking. However, immediate areas of interest are apparent. Both case studies given in this thesis, in Chapters 3 and 4, revolve centrally around entropy and its usage. Whilst being careful not to conclude entropy is the singular property that causes these issues for the realist, the fact that it can be viewed solely as an analogue for information hints that the way a physicist views entropy can lead to epistemological consequences that need dealing with carefully. Are all unresolvable entropy paradoxes exorcisable in the same manner this thesis deals with Maxwell's Demon or the Page-time paradox? Can entropy be considered a physical property at all? Can a human, with a finite epistemic reach, ever develop a non-

problematic metaphysical framework for entropy, given the arbitrary choices an observer makes can directly lead to different experimental results?

Above all, the conclusion encourages physicists who work within black hole physics or statistical mechanics to adopt or apply constructive empiricist viewpoints regarding the work they are currently doing. If you're a physicist, is the research you're undertaking useful to the development of empirical understandings of the world, or is based on a set of "inchoate intuitions"[4] that will fundamentally lead nowhere and offer no meaningful progress when it comes to modelling, predicting and bending to our will the world around us? Once asking these questions of yourself is established as normal practice within physics, I contend that physics will be far better off and in a stronger place to understand the complex and unintuitive universe that surrounds us.

[4] A phrase taken from Erik Curiel (Curiel, 2019) when discussing other problems to those which are introduced by this thesis in black hole physics. I have found myself returning to the phrase repeatedly throughout this PhD project.

Chapter 1 – An Example of a Black Hole Paradox

This thesis will demonstrate that realist views about physics are often inappropriate. The claims the realist wants to make about the ontological power of physics are unsustainable, which is particularly acute in certain research fields. Instead of accepting realism, I will show that we are better off taking a constructive empiricist approach, which manages to "save the good" whilst eliminating problematic metaphysical questions from the scope of physics entirely.

In order to do this, we require a firm understanding of the state of play within physics. This Chapter will be focused on achieving that by giving an overarching view of the Page-time paradox in a way intelligible for a philosopher. As this revolves around mutual inconsistencies between thermodynamic/statistical mechanical approaches to black hole entropy and quantum physics approaches, it serves as a useful heuristic and conceptual relation to the other entropy paradox discussed directly in this thesis, Maxwell's Demon (which is explained and grounded within its own Chapter, starting in section 3.2.1). The Page-time paradox itself will be returned to in all chapters of this thesis and used to help motivate why we should be constructive empiricists, so outlining and clearly showing both that it *is* a problem and *why* it should be dealt with seriously will be particularly useful in our later analyses.

This Chapter starts by introducing the most necessary aspects of General Relativity (hereafter occasionally GR), before discussing concepts around black hole entropy introduced by Bekenstein. From that point it introduces the Page-time paradox and some related theories. This is done to introduce the version of the black hole information loss paradox I consider strongest, as well as demonstrate that the debate in physics surrounding black hole dynamics is intimately linked to metaphysical and epistemological questions.

The underlying physics work of these paradoxes is deeply technical physics that is constantly developing. As this thesis is written within philosophy, I will over this chapter seek to present how these paradoxes generate conceptually and why, according to modern physics analysis, they deserve the moniker "paradox". In doing this, I have occasionally chosen to use language that is understandable to a philosopher not trained in physics. Occasionally I will be using terminology differently to how a physicist would or explaining concepts in broader terms than a black hole physicist would, making points that are conceptually accurate, even if technically non-precise. Chapter 2, Section 2.1.2 seeks to give a defence for this approach across the piece (a natural consequence of explaining concepts in an understandable way to two subjects with occasionally mutually contradictory terminology). Another reason for this approach is

sheer intelligibility: it is simpler to explain the conceptual functioning of some mechanics via descriptions that are almost akin to analogy (i.e., not technically precise) where the full technical description may cause confusion in those without the pre-existing training. Some examples of this approach are pulled out of the text and explained via footnote throughout the chapter, to aid the reader's comprehension.

This Chapter also holds a heavy debt to the work of David Wallace, whose multi-part analysis of black hole thermodynamics and information loss paradoxes have been essential in helping to frame the discussion and presentation of what the Page-time paradox is and its consequences. Readers who want, in my mind, the best discussion of black hole thermodynamics and information loss paradoxes can read (Wallace, 2017), (2018a) and (2018b). The Stanford Encyclopedia of Philosophy page "Singularities and Black Holes" by Erik Curiel (Curiel, 2019) has also been an invaluable source, not just of philosopher appropriate descriptions of physics, but also of relevant future questions in philosophy. I hope I can help contribute to their answers.

1.1.1 General Relativity and Classical Black Hole Spacetime

The theory of General Relativity, whilst not the origin point for the supposition for supermassive gravitational systems in physics, is the point of departure for how they are described in current physics. Albert Einstein proposed an extension to his Special Theory of Relativity in 1914 that incorporated gravitational non-flat spacetimes, giving us three important concepts in how we understand gravitational space-times, which in turn led to the theorising of black holes in the modern sense. These are:

1) The Equivalence Principle,
2) Gravitational Time Dilation, and,
3) Existent Singularities.

The Equivalence Principle states that no local experiment should be able to tell the difference between being in an object accelerating free from gravity at $X\,{}^{m}/_{S}{}^{-2}$ and being in an object being pulled gravitationally downward at $X\,{}^{m}/_{S}{}^{-2}$. The forces upon the observer should feel identical (a supposition we use in everyday life when referring to acceleration force on an object in units of "G", 1 G being the object's weight on the surface of the Earth) (Einstein, 1914, p. 291).

Gravitational Time Dilation is the concept that the rate at which time passes in a certain area is dependent on the space-time topology[5] of its surroundings. The closer you are to a gravitational source, the slower time flows, time speeding up or slowing down dependent on the distance from the gravitational source and yourself, i.e. what would be classically known as one's gravitational potential. This will impact how we can view black holes, given the fact they are large gravitational sources, and is invoked centrally in supposed solutions to black hole paradoxes (Einstein, 1920, §. XXIII).

The singularity of a black hole is a region of infinite density (but a finite mass) that is often conceptualised as either being viewed as a rip in the fabric of space-time or an example of pathological space-time topology (Curiel, 2019, §.1). It can be also viewed as the engine of the black hole, as its infinite mass creates the black hole's resulting structure, which as we will demonstrate throws up epistemic challenges for the physicist.

The structure of a black hole is defined by the underlying General Relativistic framework. There are two key components of note. The first is the singularity at the centre, which is the gravitational source for the entire system and is the root cause for the rest of the observable phenomena. The second is the black hole event horizon, a coarse-grained region of space-time which is the boundary between the part of spacetime with a curvature such that light can escape from the black hole and the part of spacetime where light must inevitably fall into the singularity. This is the boundary we "see" when we refer to the black hole and is in effect the surface area of a black hole, bounding the blackness when seen from outside. I will refer to region between the singularity and the event horizon as the "black hole interior" from here on out, and the region of space-time between the event horizon and infinity as the "black hole exterior".

One important aspect of the event horizon to note is that, due to the principle of equivalence, it is a fully non-physical boundary. There should be no marked change in events or space-time between being 1m outside the event horizon compared to 1m inside it, beyond a very small change in something akin to the gravitational potential experienced (and a change that should be no different, proportionately from being 3m outside the event horizon and 1m outside it). However, despite its lack of physicality, the event horizon does provide us with an epistemic gap. We can, theoretically and with enough energy, recover information from an object 1m outside a black hole, whereas there is absolutely no hope of recovering information from an

[5] You can also think of this as akin to a change in gravitational potential, which is the way you would describe this space-time curvature in Newtonian Mechanics.

object 1m inside the black hole. This lack of physicality, combined with the informational loss, will help motivate the first of our black hole paradoxes, the information paradox.

One last structural point to note is that black holes are, when taken as a combined system, quite simple to describe. As a generalisation, a classical black hole can be described in full by only three attributes: its charge, angular momentum, and mass. This characterization of black holes is called the "No-Hair Theorem"[6] (Bekenstein, 1995). This understanding, which is important historically in relativistic physics and has had eminent physicists amongst its adherents[7], presents problems for people wanting to resolve the information paradox if we assume its correctness, as it demonstrates the fact that black hole interiors and exteriors are completely disconnected systems informationally, meaning we can never hope to recover any information about what is inside a black hole once it has fallen into one.

In the next section I will introduce the work of Bekenstein and Hawking, which lays the groundwork for black hole thermodynamics. All information loss paradoxes are consequences of black holes, in effect, losing information, and the process by which we describe those losses is with regard to entropy changes. Therefore, I am choosing this moment to introduce concepts like black hole entropy and the temperature of a black hole as well as their mathematical descriptions.

1.1.2 Thermodynamic Systems and Black Hole Entropy

The concept of black holes being thermodynamic systems began when Hawking showed black holes have a temperature and was further evidenced by the theoretical acceptance of Hawking Radiation.

Hawking Radiation is the name given to the thermal radiation that emanates from black holes, causing black holes to shrink as energy is thus radiated away (Hawking, 1975). This radiation is caused by quantum mechanical effects that happen everywhere, happening near the event horizon. In such a situation the creation of a particle-antiparticle pair, where one goes away from the black hole and the other falls into it, would cause a loss in mass for the black hole and

[6] Another aspect of language difference between physics and philosophy. The "No-Hair Theorem" properly understood isn't a "theorem" or "theory", but more properly a conjecture about the supposed simplicity one can have when describing a classical black hole. However, within the literature it is referred to as the "No-Hair Theorem", so I will stay with convention and refer to it as such.
[7] Gravitation, by Charles Misner, Kip Thorne and John Wheeler famously states that "a black hole has no hair" in the hope that further physics progress will give a full mathematical proof, and that the No Hairs Theorem has "several highly technical assumptions" that may "seem physically reasonable and innocuous, but … might not be" (Misner, et al., 1973, p. 876)

produce thermal radiation. This gives a black hole its thermodynamic quality, by which a black hole's temperature could be measured, it being equivalent to equation (1) below:

$$T = \frac{\hbar\, c^3}{(8\pi \mathrm{G} \mathrm{M} k_b)}.$$ (Hawking, 1975, p. 199) (1)

Bekenstein further increased understanding of black holes by providing both a strictly mathematical description of both black hole entropy, and a generalized form of the Second Law of Thermodynamics incorporating black hole entropy. Bekenstein's black hole entropy is directly proportional not to the volume of the black hole, but its surface area, further hinting at a deep relationship between the knowable physics of black holes and their surfaces (or black hole exteriors). The equation is thus:

$$S_{bh} = \left(\frac{1}{2}\frac{ln2}{4\pi}\right) k c^3 \hbar^{-1} G^{-1} A$$ (Bekenstein, 1973, p. 2338) (2)

This clearly shows that as the surface area of black hole increases, its entropy (as measurable to an external party) will also increase in a directly proportional relationship, dependent on the constants h, G, c, and k (Planck's constant, the gravitational constant, the speed of light and Boltzmann's constant). From this, and to prevent any breakages of the second law of thermodynamics for closed systems (i.e. the universe)[8], Bekenstein also provides a generalized form, which follows:

$$\Delta S_{bh} + \Delta S_c = (\Delta S_{bh} + \Delta S_c) > 0$$ (Bekenstein, 1974, pp. 3296-3297) (3)

This shows that if we take the overall change in entropy of both the black hole (ΔS_{bh}) and the entropy of everything outside the black hole (ΔS_c) we still get a measure of entropy that increases for the entirety of the universe.

There are potential violations to this understanding of black hole entropy. In 1970 Geroch proposed at a colloquium at Princeton that any generalized second law on the model of Bekenstein's is open to a breakdown of the Second Law, which he demonstrated via a thought experiment: Imagine a massless box full of energetic radiation with a high entropy, prepared far away from a black hole. You could then lower this box toward a black hole; its energetic radiation being attracted via gravity to the black hole via the mass-energy relationship. We could then rig up a system to extract energy from this weight (via friction on a counterweight etc). If we arrange this system so that the mass-energy of this system is exhausted when it

[8] This is an assumption. I motivate this by stated that the universe is isolated (i.e. is not got another universe to transfer energy into or out of) at least from our epistemic perspective, and this has the consequences of leading us to a "closed" universe thermodynamically (further discussion on information theoretic views on entropy happen in Chapter 3)

reaches the event horizon, we can open the box and let the radiation fall in. The black hole should not expand here (as the mass-energy of the black hole won't increase), but the entropy of the surrounding universe will have fallen. This leads to a lower entropy over the closed system, which is undesirable (Curiel, 2019, p. 5.3)[9].

Bekenstein considered this problem and proposed that there must be an upper bound on the amount of entropy that can be stored within any region of space-time. Whilst no limit is given to us by current physics, Bekenstein maintained it would be revealed to us in any quantum gravity complete theory. We can also see an intuitive reason there must be an entropy upper limit in space-time. Imagine a region of space-time with more entropy than a black hole of the same size. One could then collapse this matter into a black hole, which would not be larger than the size of the original region (or we'd already have a black hole there). Here we have a violation however, as the black hole entropy would have to be lower than the entropy of the matter that formed it (Curiel, 2019, §5.4)

The very fact we can not only model black holes based on them being thermodynamic, but we seem to have found a deep physical relationship between two seemingly independent theory spaces raises interesting inter-theoretical questions for a physicist. What does it mean to say that a purely gravitational system is a "thermodynamic object", given it has no physical boundary? This question, amongst others, has prompted many different forms of "black hole information loss paradox". This thesis will not deal with all of them, for reasons of brevity and clarity. Instead, I will focus on one specific paradox, The Page-time paradox. I have chosen this paradox for its seeming intractability, though I assert that its troubling aspects are not unique to itself but are in fact shared with other information loss and entropy paradoxes. This will make it a good case study in which we can test my hypothesis that constructive empiricism is the most attractive philosophy of science to adopt in the context of entropy paradoxes.

The next two sections will introduce the Page-time paradox, originally described by Don Page (Page, 1993), and following this will discussions of two of, arguably, the most eminent proposed solutions. These two proposed solutions are black hole complementarity (as favoured by (Susskind, et al., 1993) and others), an example of the 4th style of resolution, and black hole firewalls (as favoured originally by Almheiri and others, being introduced in the 2013 paper (Almheiri, et al., 2013)), an example of the 5th. I introduce Black Hole Complementarity and firewalls here not because they are essential to describing the Page-time paradox, but because they further highlight that the debate within this area is driven largely by non-empirical, nearly

philosophical, beliefs about the world and physics. I offer this quote from Curiel who, despite talking about the Information Loss Paradox, offers an opinion that I think applies also to the physicists directly dealing with the consequences of Page's theory: "The attitude that individual physicists adopt towards [the information loss paradox] is strongly influenced by their intuitions about which theory, general relativity or quantum theory, will have to be modified to achieve a consistent theory of quantum gravity" (Curiel, 2019, § 6.2).

1.1.3 The Page-Time Paradox – Origins

The *Page-time paradox*, first described by Page (Page, 1993), is a form of black hole information paradox that derives from the mutual incompatibility of two mature and well-grounded theories. Consider a semi-classical black hole, cooling through Hawking Radiation. We can assume that, if the original state of the black hole is pure and evolves with unitary dynamics, the black hole and radiation from the black hole must also be pure. However, we equally know that as Hawking radiation is exactly thermal, it must itself have a highly mixed state. This implies that for the total system to be pure (as we would expect it to be), our highly mixed photons must be entangled with some other system so as to be pure. As no two emitted photons can be entangled with each other, all photons must be entangled with the system they originated from. This brings about an entropy of entanglement, the *von Neumann* entropy, which if it is to evolve with unitarity, must be bounded by the wider microcanonical entropy of the entire black holes, which is given by the inequality:

$$(1)\quad S_{VN}\big(E_0, E(t)\big) \leq S_{MC}\big(E(t)\big)$$ (Wallace, 2017, p. 12)

Where S_{VN} is the Von Neumann entropy, E_0 is the initial energy state of the system, $E(t)$ is the energy as a function of time and S_{MC} is the microcanonical entropy.

The upshot of this is, as the black hole cools and radiates, become smaller as it does so with its S_{bh} tending to zero, there comes as a time where the inequality above is saturated. However, as the von Neumann entropy is also bounded by the microcanonical entropy, at some stage the von Neumann entropy must stop rising. This, in a figure taken from (Wallace, 2017) shows the relationship well. The saturation point is called the "Page-Time".

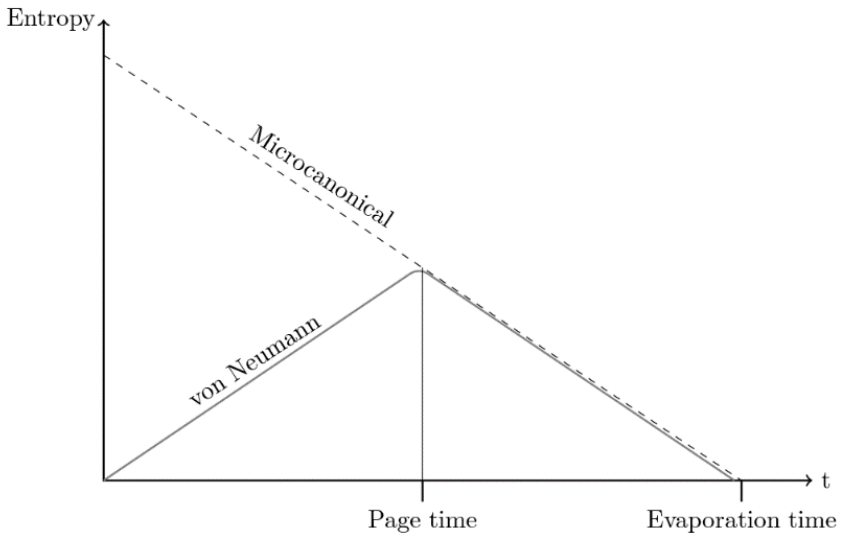

Figure 1 - The Page-Curve, showing the bounding of the von Neumann entropy by the microcanonical entropy of the black hole (taken from (Wallace, 2017, p. 14))

For a normal Schwarzschild black hole, its von Neumann entropy is very small compared to its microcanonical entropy (as much because things fall into black holes in the real universe, raising its microcanonical entropy as anything else). For an isolated black hole, the Page-time will be reached about halfway through the process of evaporating via Hawking Radiation, when half of its matter has evaporated away (as is shown in Figure 1). After this time, our Hawking radiation cannot be exactly thermal; instead, it should be maximally entangled with early time radiation rather than entangled with interior modes of the black hole. However, it *has* to be exactly thermal according to quantum field theory (hereafter occasionally QFT) calculations.

Here we have a paradox. In Page's analysis of the problem, we have an even worse issue than the original black hole information loss paradox. Hawking's original version of the black hole information loss paradox, the full evaporation of black hole through Hawking Radiation would lead to a situation where any pure quantum state that entered a black hole would necessarily become thermal, i.e. a mixed state, after it had evaporated (Page, 1993, p. 3743), breaking unitarity. This paradox is vastly more concerning to the physicist as we cannot be hoped to be saved by exotic, quantum physics. This breakdown between the two entropies is occurring at a macroscopic scale, in regions where semi-classical approaches to black holes should remain fine. The moment at which our pure/mixed quantum state mutual inconsistency occurs is now well before the black hole's evaporation (Wallace, 2017, p. 14). Again, this is a *true* paradox.

This is made all the more troubling by the fact it happens when the black hole is macroscopic[10] and that the fact that the two theories causing the paradox, statistical mechanics and quantum field theory are mature and well-grounded theories experimentally. If we choose to abandon the statistical-mechanical understanding of black hole dynamics, we are left with no good model that predicts the thermodynamic aspects of a black hole. If we choose to abandon QFT we have to explain why it does so in an area that it should otherwise work. This is a deep problem for both physicists and realist philosophers of science; the former want consistency of their own theories across the areas they should be applicable; the latter wants those desired consistent theories to form the basis for their metaphysical understanding of the world.

The strength of this paradox seemingly leaves us with two possible pathways:

1. Accept that QFT fails and retain the statistical-mechanical understanding of black hole thermodynamics, even though there is nothing particularly interesting about the region of space-time in which QFT seems to be failing here, or
2. Retain QFT and abandon black hole statistical mechanics. However, if we do this, we must find some non ad-hoc way to explain why our statistical-mechanical understanding of black holes successfully model the thermodynamic properties of black holes, as well as the *prima facie* quantitative consistency of statistical-mechanical calculations of black hole entropy. To paraphrase (Wallace, 2017), this would be miraculous. (Wallace, 2017, p. 16)

Neither of these pathways is good. Both are degenerative, giving physics and the physicist more work to do in areas that, empirically, work well. For physicists, the choice seems to be painful: accept one of the two pathways and work to solve the problems they generate, despite this paradox not thus far having experimental data to confirm its physical instantiation; or accept this paradox will remain paradoxical until some new theory fixes the problem, however stark and unresolvable it may currently appear.

All isn't lost however, as the Page-curve hints at something that is uncomfortable for those who hold to conjectures such as the No-Hair Theorem. A pathway to demonstrating that we *do* recover information from black holes is now present to us: can we demonstrate that the entropy of a black hole follows the Page-curve? Recent work by (Almheiri, et al., 2020b) proposes to demonstrate that not only is the Page-curve followed for idealised black hole models, but also proposes to show how the previous 'lost' information is returned to the universe. In this thesis I

[10] David Wallace points out that because of this "there seems no prospect of exotic quantum-gravitational effects coming to the rescue" (Wallace, 2017, p. 14)

occasionally discuss this research and some extensions to it and demonstrate that realism still isn't appropriate even in the context of these proposed solutions.

A sensible question to ask here is: "what happens to the black hole *after* the Page-Time?". We have outlined that the paradox centres around whether the Hawking radiation of the black hole is exactly thermal or not but not yet provided a description of what happens physically, in terms of both an external and internal observer, when a black hole reaches the Page-time. Section 1.1.5 introduces "firewalls", which are the most mature and well adopted approach for answering the question posed at the top of this paragraph.

Before we introduce them, in the next section I will introduce another theory, black hole complementarity. Although not a direct solution to the Page-time paradox (it was theorised before Page's introduction of the Page-time), I am introducing black hole complementarity to highlight its explicit centralisation of the observer, the philosophical consequences that arise for that from the realist if it is successful.

1.1.4 Black Hole Complementarity

Black hole complementarity was first devised as a resolution to the original information loss paradox (not the Page-time paradox) by Susskind *et al.* in 1993. It uses the epistemic positions of the observer to make the case that there is a description of events that while, *prima facie,* is contradictory, actually resolves due to how black holes themselves create epistemic gaps. Simply put, due to the fact that photons lose energy when away from, but originating outside of, a black hole (gravitational redshifting), an external observer will see an infalling observer getting progressively slower and hotter to the point that the observer gets thermalised (Susskind, et al., 1993, p. 3758). Further to that, considering t'Hooft's "holographic principle", that the internal dynamics of systems such as a black holes can be viewed in surface perturbations, we can say that there exists some "stretched horizon" where this information is encoded, which exists a Planckian distance above the event horizon ('t Hooft, 2000, p. 13). The description of it given by Susskind *et al is:*

> "From the point of view of an outside observer, the stretched horizon is a boundary surface equipped with microphysical degrees of freedom that appear in the quantum Hamiltonian used to describe the observable world (Susskind, et al., 1993, p. 3745)".

In this sense, black hole complementarity is nothing more than an extension of the semi-classical description of the physics of the world external to the black hole to QFT and the black hole interior (Wallace, 2017, p. 21), but a key difference can be found with regard to what the

infalling observer experiences. As they get closer to the black hole, they also heat up, but they do not get thermalised (i.e. burnt up) at the stretched horizon and moreover continue falling into the black hole as if nothing has happened[11].

Here we seem to have two contradictory understandings of what happens to the infalling observer: in one accounting, we appear to have them reaching the stretched horizon and simply burning up, and in another we have the observer falling through the stretched horizon physically undamaged. The black hole complementarian says this can be resolved however, by considering the fact that thermalisation is a coarse-grained notion as opposed to a fine-grained one, and that we can perfectly reasonably expect the two observers to have very different expectation values when taking readings of the situation they find themselves in (Wallace, 2017, p. 22). Because these two observers now have permanently parted world-lines, we are never going to be able to recover any information from the infalling observer that can be compared with the external one. No true inconsistency arises because both accounts are "complementary" of each other, in a similar way (philosophically at least) to the Uncertainty Principle of Heisenberg. In this sense, they treat black holes in string-theoretic terminology, rejecting the concept that there is this "global state" that can be accessed in a way to view both observers at the same time. In the words of Susskind *et al*:

> "The assumption of a state ... which simultaneously describes both the interior and the exterior of a black hole seems suspiciously unphysical. Such a state can describe correlations which have no operational meaning, since an observer who passes behind the event horizon can never communicate the result of any experiment performed inside the black hole to an observer outside the black hole. The above description of the state lying in the tensor product space $H_{bh} \otimes H_{out}$ can only be made use of by a "superobserver" outside our Universe. As long as we do not postulate such observers, we see no logical contradiction in assuming that a distant observer sees all infalling information returned in Hawking-like radiation, and that the infalling observer experiences nothing unusual before or during horizon crossing" (Susskind, et al., 1993, p. 3744).

This accounting has received multiple critiques from multiple directions, one of which is significant enough to be given its own section (black hole firewalls). One attack line to deal with first, however, is that made by Belot *et al.* in 1999, who accused black hole complementarians

[11] David Wallace points out that this "heating up but not burning up" can be resolved by considering the relative blueshifting/redshifting of radiation that the infalling observer will experience, it being experienced by the infalling observers over very short time periods, therefore energy transference for this observer will be lower (Wallace, 2017, pp. 21-22).

of being at best verificationists/operationalists who seek merely to use discredited philosophies of science to wrap science in unnecessarily modest metaphysical clothing (Belot, et al., 1999, p. 214). Belot *et al.* also highlights that when the black hole complementarian states that it makes little sense to talk of black hole evaporation globally[12] it seems to be because they are unwilling to talk about *anything* to do with black holes globally. This is held to be problematically degenerative if we want to do physics about things beyond our mere epistemic ranges, i.e. do physics about black hole interiors[13]. To this end, Belot *et al* bring in thought experiment in which they challenge the central conception of black hole complementarity, that no external observer's observation can be contradicted by the observation of an internal observer:

> "An EPR pair of particles with anti-correlated spins is created. Particle 1 falls into the black hole, where its x-spin is measured. The black hole, obedient to the Complementarians' axioms, emits radiation from which particle 1's spin state can be determined. An external observer uses this radiation to determine the z-spin and of particle 1, then falls into the black hole where she receives a message telling her of the outcome of the interior x-spin measurement of particle 1. *Pace* the doctrine of Black Hole Complementarity, she is then in a position to simultaneously entertain assertions made from putatively complementary perspectives - her own past assertion, issued in the black hole's exterior, about the z-spin and of particle 1, and the assertion, issued in the black hole interior, about its x-spin" (Belot, et al., 1999, p. 215)

This is meant to challenge the core conceit that no one person should be able to access both spins simultaneously, and demonstrates there is fundamentally a global state in which one can observe both the exterior and interior of black hole whilst being within them (and not having to reduce oneself to observing the black hole interior via the exterior, *a la* the complementarians). I maintain this counterexample seems to miss the entire point of what happens to a person's epistemic position once they cross the event horizon, and how much a physical description of events relies on unspoken assumptions when making such thought experiments. I will discuss why I perceive the event horizon to operate as a full epistemic barrier in Chapter 4, when discussing black hole observability and the consequences this will further have for resolving the Page-time paradox, but we can note for now that the event horizon's property of permanently

[12] As they do when they claim their case is only problematic from the perspective of a "superobserver" (which humanity can't be) who can view the system globally.
[13] This is seen by Belot *et al.* as something physicists do, in fact, want to do. I will be strongly challenging this conception in later chapter as both unworkable and undesirable when attempting to do good physics.

severing those inside from those outside is central to many theories of black hole dynamics outside of just black hole complementarity.

There are also defences of black hole complementarity, though these mainly come in the terms of thinking Susskind *et al*'s theories are themselves wrong, but prompt better and interesting thoughts elsewhere. Peter Bokulich (Bokulich, 2005) makes the case that whilst he considers the arguments put forth by the black hole complementarian unsatisfactory, the research programme as a whole has interesting merit when it comes to delineating the limits of our semi-classical theories and their effective limits (Bokulich, 2005, pp. 1346-1348). He concludes that the "spacetime complementarity as offered by Kiem, Verlinde, and Verlinde, which talks about understanding different quantum states as being better understood via differing background geometries takes what the black hole complementarian is trying to do, softens it and makes it more palatable for the scientific realist who still holds out for global descriptions" (Bokulich, 2005, p. 1346).

In the next section, we will talk about what some people have described as the logical extension of and counter to the black hole complementarian account of the stretched horizon, which is taken to demonstrate deep flaws in the black hole complementarian's account: black hole firewalls. This is still an active area research, and all possible accountings would be a thesis unto itself. Instead, I intend to offer an account of what firewalls are, how they generate, and some criticisms of firewalls as a whole from Susskind and others.

1.1.5 Black Hole Firewalls

Black hole firewalls are a physics-based critique to the assumptions implicit in the black hole complementarity approach, and equally a demonstration of the failure of semi-classical physics alone to resolve the Information Loss paradox. Almheiri *et al.*'s introduction of firewalls starts from the statement that the initial axiomatic assumptions underpinning black hole complementarity are mutually incompatible. Specifically, they maintain that three of the assumptions necessary for the complementarian account cannot all be true simultaneously (Almheiri, et al., 2013, p. 1):

1) Hawking radiation is in a pure state (i.e., entangled with the black hole itself *a la* the von Neumann entropy in section 1.1.3)
2) The information carried by the radiation is emitted near the horizon, with the semi-classical approximation holding, and,
3) Any infalling observer experiences nothing out of the ordinary.

They maintain that the most *conservative* approach, i.e. the one that causes the fewest problems for wider physics as a whole is to simply expect for there to be a firewall at the event horizon after the Page-time. In doing so, they highlight the necessary incompatibility also highlighted in the Page-time paradox, that the Hawking radiation of a black hole cannot be simultaneously thermal (i.e., pure and entangled with the early time radiation) and entangled with the system, as predicted by QFT. To quote them directly: "To restate our paradox in brief, the purity of the Hawking radiation implies that the late radiation is fully entangled with the early radiation, and the absence of drama for the infalling observer implies that it is fully entangled with the modes behind the horizon" (Almheiri, et al., 2013, p. 5)

David Wallace in his 2017 piece (Wallace, 2017, pp. 22-23) presents another description of what he terms "the firewall paradox" via a thought experiment that I present here for the reader: Imagine a wavepacket on some photon mode B which has been emitted well after the Page-Time. This wavepacket will be in a thermal state equivalent to the age, size and evolution point of the black hole it has been emitted by. This creates two contradictory claims:

1. According to QFT, this photon mode B must be entangled with some photon mode $\tilde{B}$ inside the event horizon,
2. According to statistical mechanics, this photon mode B must be entangled with some earlier radiation released by the black hole before the Page-Time.

Given the monogamy of entanglement[14], this seems at best like a contradiction that needs resolving. At worst, it seems like even something akin to complementarity cannot save us; see this next thought experiment:

1. Gather up all radiation emitted by the black hole *up* to the Page-Time,
2. Carry out a series of complex complicated operations to distil this radiation into a single mode C, which is fully entangled with B,
3. Verify that that C is fully entangled with B whilst lingering close to the event horizon,
4. Jump into the black hole.

In the thought experiment above, assuming that quantum mechanics holds for the observer, they cannot consistently find B to be entangled with $\tilde{B}$; in fact, $B + \tilde{B}$ must be, in and of itself,

[14] The "monogamy of entanglement" is the concept that two particles maximally entangled cannot be also maximally entangled with some third particle. This is relevant here because both QFT and statistical mechanics suggest that our photon mode B is, *in extremis,* with some other particle. They can't both be right; one must be wrong. For a very readable deeper explanation of the monogamy of entanglement, see (Terhal, 2004)

in a product state. This undermines the QFT assumptions that underpin Hawking Radiation, which in turn undermines the statistical-mechanical understanding of black holes we originally had.

Consequently, given the massive break in entanglement that must occur at the Page-Time and the flux of Hawking radiation must be so large, there must be some physical "firewall" that forms at the event horizon of a black hole, thermalising all that enter it (Almheiri, et al., 2013, p. 3). This immediately leads to questions about compatibility with pre-existing general-relativistic understanding of black hole dynamics. To quote Wallace, "complete distentanglement of QFT modes across the event horizon corresponds to a Planck-scale wall of energy at the horizon — the 'firewall' — that seems physically inexplicable and quite at odds with the general-relativistic idea of an event horizon as a globally-defined, locally-inaccessible phenomenon" (Wallace, 2017, p. 23).

This is not to say that there are not alternative opinions. Susskind, in a response paper to the original Almheiri *et al* pre-print, presents another consequence of adopting firewalls as a description of what happens after the Page-time. If you adopt firewalls but also wish to maintain the principle of equivalence as sacrosanct, the singularity must move to the event horizon as a black hole hits the Page-Time. He proposes that this does not happen suddenly, through the singularity suddenly hitting the event horizon as the Page-time is hit, but a slower process over time where the singularity 'migrates' over time. The logical consequence of this, after the Page-time has been hit, is the sheer 'non-existence' of the space-time behind a firewall (Susskind, 2012, p. 7). This obviously presents a radically new way of considering the internal structure of black holes that has deeply troubling philosophical consequences for the realist.

Beyond this, he presents troubling metaphysical issues in terms of how and when this black hole singularity can move before and after a black hole's Page-time has been reached. Figures 2 and 3 below show how a black hole's event horizon is supposed to move under Susskind's proposal, and what area of space-time is lost due the singularity getting proportionately closer to the event horizon than would have been semi-classically expected. In Figure 2, we have a smoothed, logarithmic progression of the singularity from its traditional place as an infinitesimally small region to being present at the event horizon, the moment of which corresponds to the Page-time.

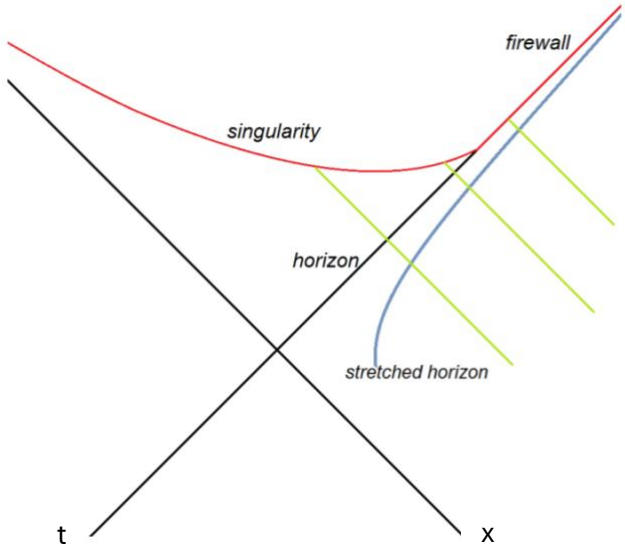

Figure 2 - Susskind's understanding of the migration of black hole singularity over time, in Kruskal co-ordinates. (taken from (Susskind, 2012, p. 9).

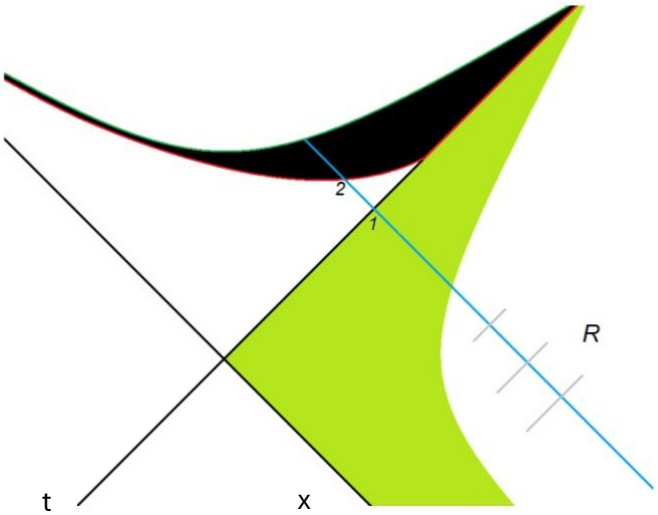

Figure 3 - The area of space-time 'lost' due to the firewall effect (area shaded in black), where the teal line bounding the black area represents the singularity as the black hole approaches the Page-time; and the red line representing the migration of the singularity according to Firewall theory. An infalling observer (the blue line) hits both "singularities" at different points. (taken from (Susskind, 2012, p. 11)).

This seems, if not intuitive, understandable within our everyday causal experience of reality. Before the Page-Time, a particle can enter through the event horizon and experience a shortening but still real time before hitting the singularity. At the Page-time however, hitting the event horizon is akin to hitting the singularity.

In Figure 3 we can explicitly see what area of space-time has been lost due to this new understanding. To again clarify, this area of space-time must be fully considered "non-existent", in the same way we consider the area within a singularity non-existent due to its inherent infinities which make it conceptually unphysical in a finite universe. This problem is

resolved for the singularity because it is also infinitely small, so the area lost is non-physical. I think that this cannot be maintained with macroscopic black holes.

Susskind ends up contending that the largest issue with black hole firewalls is that they seem to cause interesting causal events to arise when it comes to predicting the existence of a firewall. This includes suggesting that a state exists whereby an observer could, by measuring the flux in a certain area, predict with a reasonable degree of certainty whether she would be immediately thermalised on contact with a black hole or not, despite having no other way to tell its age (Susskind, 2012, pp. 23-24). From an external perspective[15] this seems like another example of how black hole firewalls seem to cut against the principle of equivalence, which is central to Einstein's understanding of general relativity. Whether this problem is to be solved via more, better physics or a philosophical shift in how we consider black holes is still up for debate.

Another problem for the person advocating for firewalls is presented by Harlow in his Jerusalem lectures on Black Holes and Quantum information. In this, he gathers 6 possible proposals for how firewalls could be physically instantiated all of which have deep problems to resolve, which I present with my own framing below (Harlow, 2015, pp. 115-117):

1) **Full Strength Firewalls**: Considering firewalls in their strongest sense, fully accepting that the interiors of black holes must be considered non-existent and advocating for a moving singularity. Problems with this come with explaining how and/or why. the singularity moves in a non ad-hoc sense, and how black hole thermodynamics makes any sense on this analysis.
2) **Typical States Only Firewalls**: An argument that firewalls apply as is stated above in theory but could never be firmly realised in reality due to the complex nature of our universe. This standing also potentially allows for microscopic black holes due to the fact they are exceedingly rare as a proportion of black holes and that they have different dimensionality. However, the same caveats about Hawking radiation and black hole thermodynamics from above still hold.
3) **S-Wave Firewalls**: Black holes having firewalls that only affect low angular momentum modes. This has the benefit of potentially reclaiming chunks of black hole thermodynamics by limiting the scope of firewalls but suffers from the issue of seemingly being an ad-hoc solution. No proposed mechanism for the formulation of S-Wave Firewalls has been produced.

[15] Bearing in mind the event horizon is a non-physical boundary.

4) **Non-violent Nonlocality**: This proposal suggests that we should consider the concept that firewalls form in a more diffuse way throughout the black hole atmosphere, rather than just at the event horizon. This seems to require large changes in how we model black hole thermodynamics, which could potentially be saved by equally large changes in Schwarzschild geometries outside the black hole (which could be experimentally detected). Harlow makes the case that when we allow such large modifications of effective field theory outside the horizon, we could see similar modifications being needed in other place. To quote, "'small' violations of causality tend to have a way of not staying small" (Harlow, 2015, p. 117).
5) **Fuzzballs**: For higher-dimensional black holes that seem to exist in string theory there appear to be "fuzzball solutions" that look like black hole solutions when approaching infinity, and there are suggestions there is a fuzzball solution for every black hole solution to the general relativistic field equation. These may provide ways to allow for firewalls that don't conflict with our other theories, but this is still very much a work in development, as there still exist no "fuzzball" solutions equivalent to an uncharged, semi-classical black hole, and even some theories that prohibit them existing.
6) **"Shut-Up-And-Calculate"**: String theory gives candidates for quantum gravity. Shouldn't we simply take one and calculate what it predicts for what occurs at the event horizon/ the structure of the black hole interior? Harlow endorses this approach philosophically but concludes that string theory simply isn't well enough understood for us to acquire clear answers about what it predicts. Beyond this, I maintain that via such approaches we are still left with metaphysical questions. Is a non-empirical response to what is surely an empirical question solid enough ground on which to base our *physics*. Theory alone, as for many of these examples, may not prove solid enough ground.

The research surrounding black hole firewalls is active, changing, and fast-paced. Any description of them will necessarily be a snapshot of time, with considered opinions being shown to be incorrect in a matter of months. Since this beginning of writing this thesis, revolutionary work has been done in this field which this section won't cover. Examples include (Almheiri, et al., 2020a), which introduces "entropy islands" and "replica wormholes" as pathways to resolving the problems of firewalls and the Page-time paradox (and which, in my understanding, help resolve the problems that Susskind identifies in a migrating singularity) by explicitly describing how information and in a sense entropy can be physically transferred from the black hole interior to the black hole exterior, i.e. be recoverable. (Busso & Penington, 2023)

extend this further, showing mathematically that these entropy islands could extent to atom-length scales outside the event horizon, far beyond the stretched horizon and thermalisation.

Rather than deal with all these strands at once, this section has provided an example in the Page-time paradox of the "entropy paradoxes" central to this thesis' contention that overly realist approaches to physics cause the generation and proliferation of such paradoxes. It has also made the case that the debate and discussion within this field of physics research is bound up with philosophical argument and intuition far more deeply than is traditionally expected of debates in physics. In the next section, we will focus on these conceptions and argue that better ways forward are only visible once physicists abandon realist metaphysical commitments about their theories.

1.1.6 Takeaways

In this Chapter, we have introduced a problem considered a "paradox" within physics, the status of which seems to force us into uncomfortable choices about which physical theories we choose to adopt. The direct conflict between QFT and statistical mechanics is, to quote Wallace, a "deep puzzle arising from enormously-plausible yet apparently-contradictory lines of reasoning within quantum gravity, and at present it is completely opaque how it is to be resolved" (Wallace, 2017, p. 24). This should give cause for concern to the realist, whose belief system is predicated on the idea that physics can provide for them a consistent ontology, wherein they can use science to perform their metaphysical work for them. By showing, in some depth, that physics cannot absolutely provide this internal consistency, producing paradoxes that are seemingly unresolvable, the realist is left in an unenviable position where a "science-first" approach doesn't allow for one single description of "what exists". Further to this, any realist-inclined physics approach seems to have to wrestle with this paradox fully, definitively demonstrating which of QFT or semi-classical statistical mechanics we have to be willing to throw out simply to re-achieve consistency despite their enormous plausibility.

Also to note here is the practice of the physicists themselves. Investigation in theoretical physics is often at a remove from direct empirical observation of the systems under question, at least in initial theory development. However, as is hinted at by black hole complementarity, this is a research field dealing with not only the usual issues present in theoretical physics, but also with macroscopic regions of the universe that are potentially permanently inaccessible to external observers.

However, despite this paradox existing, all hope is not lost for the physicist attempting to do good science. In order to provide physicists with a framework for solving paradoxes like the Page-time paradox, this thesis will do the following:

1) Analyse the root causes of the paradox and ask whether they are explicitly empirical in nature, or the consequence of theoretical extensions of our current physics to new, yet untested (or maybe untestable) areas.
2) Analyse the foundational philosophical positions or biases that may prevent us from accepting otherwise pragmatically useful pathways to solutions. For example, is there a burning need to have a universal ontology driving research down specific paths, which physics *per se* doesn't require?
3) Take these higher-level thoughts and reapproach the paradox, in this case The Page-time paradox (other paradoxes are also considered throughout this thesis) and see if novel approaches, outside of our previous self-imposed restrictions, are available.

Chapter 2 attempts to show that points 1) and 2) are not only achievable via using a specific anti-realist philosophy of science (constructive empiricism), but also naturally generate from certain approaches to how and why we do physics. It also seeks to make the case that for any given "working physicist", it is necessary to remove any "inchoate intuitions" about realism or metaphysics before one is successfully able to proceed on this journey. In achieving this, Chapter 2 will start of by establishing terms of engagement for the discussion: a simple heuristic for understanding why we do physics, an approach to language and terminology throughout the thesis, and a protagonist against whom we can analyse the reasonableness and likelihood of acceptance of any suggested philosophy of science that we encourage physicists to adopt. After that I will discuss the issues realist biases and leanings cause physics at a foundational level when it comes to *doing good physics*, along with a defence that these biases are given to us both pedagogically and by our lived experience in society. Following this we will reintroduce the Page-time paradox and analyse it again using the structure the Chapter will have outlined by that stage, before finally introducing what I believe to be the best approach our working physicist can take, constructive empiricism.

We will return to the Page-time paradox in Chapter 4, where this thesis hopes to show that, at its core, the debate above is doomed to either deadlock and failure, or resolutions that have dire metaphysical consequences for those wishing to remain realists. In any case, I will demonstrate that the constructive empiricist case is superior to the realist case given current physics, irrespective of your leanings either for or against theories like firewalls.

Chapter 2 – The Working Physicist and Constructive Empiricism

2.1.1 The Problem of Defining Physics

This chapter will argue that physicists should become constructive empiricists, at least within the context of analysing and understanding entropy paradoxes. It does so on the basis that adopting constructive empiricism as a framework allows physicists fewer headaches and more productivity, i.e. being a constructive empiricist is more *useful* than being a realist. In doing so, it is important to outline the scope of my argument forthrightly: I am, for the purposes of this thesis, only advocating constructive empiricist viewpoints when directly dealing with entropy-centric paradoxes. Any further extension of this to other areas of physics may be valid but is not rigorously defended within this Chapter or thesis.

In establishing constructive empiricism as a superior philosophy of science to realism, we will begin with setting the stage and defining the metrics and standards which are necessary for a rigorous analysis of the argument. To begin, we will begin by looking at Cartwright's understanding of what constitutes "physics" work and how that can inform our future conclusions. I will also, within that same area, discuss my approach to terminology throughout the thesis, defending the need for a looser terminological approach on the basis of communicating between disciplines; this thesis should be equally readable and understandable to both physicists and philosophers, both of whom occasionally use the same words to mean different things.

After this, I will introduce a central character to the narrative, the "working physicist". I do this in order to give us a metric against which to frame the reasons for being a constructive empiricist: there isn't much utility to the thesis without outlining who the sort of individual is that we are trying to convince. In doing this, I will also show that said "working physicist" is pushed into the direction of realist intuitions both explicitly and implicitly through language and normative behaviours.

From this point, I will again look at the Page-time paradox and analyse what our working physicist would think when faced with such a problem, and the pathways they can follow in trying to resolve the issue. From there, I will introduce constructive empiricism, introducing it along with some philosophical grounding on stances and pragmatic scientific realism debate. I will take this introduction and show that the arguments against constructive empiricism are

either poorly formed or are also true of realist approaches, before finally showing the ease by which our "working physicist" can adopt constructive empiricist principles.

In the later chapters, I will use this introduction to constructive empiricism and the working physicist and apply them to specific areas of physics. In doing so, I will show that physics has already presupposed the validity of constructive empiricist approaches, as well as offering more optimistic viewpoints on paradoxes such as the Page-time paradox, whereby physicists can either solve them or ignore them, depending on how broadly we choose to view the scope of physics.

2.1.2 Cartwright and The Task Ahead of Us

Firstly, we should look at an example of how physics functions using an example. I take the following from Nancy Cartwright's *A Philosopher Looks at Science,* which provides us a perfect type-cast modern physics experiment (Cartwright, 2022, pp. 110-120)

On the 20th of April 2004, Stanford Gravity Probe B was launched. The purpose of the mission was purely scientific; to test the fundamental principles of General Relativity and confirm the work of Einstein and others regarding frame-dragging and the geodetic effect. They did this by testing how the Earth's gravitational field "drags" space-time around it, causing relativistic effects to occur. By mission's end in 2011, a paper published in *Physical Review Letters* reported good agreement between the experimental findings and the predictions of relativistic theory, helping bolster arguments that General Relativity is a good theory that accounts for some knowledge in the world, at least within the scope of what physicists consider knowledge.

This experiment follows all the standard principles that any sort of physical test should, at least as the layperson would perceive it. A theory predicted X results, lots of work went into producing a circumstance by which we could test those predictions, and then those predictions were verified. This is seemingly another success story for both science and physics, showing how we can acquire knowledge via the synthesis of theoretical ideas and practical measurement.

Nancy Cartwright, in her book *A Philosopher Looks at Science,* takes issue with this naïve reading. Whilst not denying the success of the experiment, she claims that considering this a success of some statuesque "physics" is both misleading and reductive. In Cartwright's mind, even the most seemingly "pure" physics experiments call on the work of a vast number of different fields, some even non-scientific, in their pursuit of knowledge. In her own words, "for predictions of even the purest of physics results, physics must work in cooperation with a

motley assembly of other knowledge, from other sciences, engineering, economics and practical life" (Cartwright, 2022, p. 110). To make this case about the Stanford Gravity Probe B, she points out the gyroscopes had to be as perfectly smooth as human ability could allow for the data recovered not to be lost to noise, which requires a huge amount of engineering knowledge beyond the scope of physics as a subject. Equally, the construction of said gyroscopes, and the probe as a whole requires yet more logistical and economics works, political knowledge to acquire funding for the project and the skillset to correctly project manage such a vast enterprise. The expertise required to pull off such a project also isn't solely limited to physicists; the human resources engaged on the project can be seen via its array of co-investors, from NASA to the US Air Force, organisations that have interest in physics but are equally, if not more so, enthused by the prospect of the practical applications of the knowledge the experiment can generate. For Cartwright, this all leads to a great example that "even physics isn't all physics", and that when we think about physics we shouldn't consider it some lone entity, acquiring knowledge about the universe solely within its bubble; physics, like all other areas of research, is a product of the society that surrounds it (Cartwright, 2022, pp. 110-117).

What relevancy does this have to a thesis discussing paradoxes in modern physics and the philosophical issues that help perpetuate them? By looking at this we can begin to understand the scope of "physics" and the language we should use when discussing the subject. This shouldn't be taken as the final word on the subject, many more theses could be written on this topic alone. However, when setting the groundwork for later discussion, some care and attention has to be given to what it means to be a "physicist" and to do "physics". Rather than blindly walk into a de facto definition that may neuter the message of the piece before we begin, the goal of this opening chapter and the first few sections is threefold:

1) Introduce a core understanding for the meaning of terms used within the entire thesis. Speaking across two subject areas, philosophy, and physics, it becomes necessary to clearly define what is meant by terms that can having differing meanings or implications across different subjects, especially if they have everyday usage as well. Terms like "true", "false", "knowledge" and "proof" all have definite and strict meanings within philosophy that may end up not being shared by either the layperson or the physicist, so the establishment of a linguistic framework is necessary.
2) A "definition" of physics. Following the work of Cartwright and others, I don't seek to provide an absolutely definitive understanding of what "physics" is. This task I take to be impossible, or at least beyond the scope of this thesis. Instead, I intend to present a way

of viewing physics that is acceptable to as many differing philosophies of science as possible. This "definition", as vague as it may end up, should be something both the realist and the anti-realist can agree captures at least *some* of the point of physics, allowing us a conceptual framework on which to analyse the success and failures of both fields.

3) The introduction of our protagonist, "the working physicist". Given the aim of this thesis, in practical terms, is to provide physicists with a philosophical underpinning that helps them avoid paradoxes and provide a pathway for more constructive physics work, establishing who this physicist is will be useful. This "working physicist" will be presented as someone with an expert level of knowledge in some arbitrary physics field, but with general physics knowledge akin to any other well-educated physicist. They will be assumed to have limited knowledge of formal philosophy, and a "layperson's" approach to language in all other areas. Within this structure, the idea of a working physicist having realist desires about their work will be explored, along with the idea that they remain in some sense fundamentally agnostic on the questions of formal philosophy.

Once these three issues have been clarified, the rest of this chapter will focus on the how realism has become the default orthodoxy within societal understanding of physics, the problems with such an orthodoxy and how the philosophy of science, and how constructive empiricism, provided by Bas Van Fraassen in his 1980 work, *The Scientific Image,* aligns more closely with the practice of our average "working physicist". By adopting constructive empiricism, we will suffer fewer internal contradictions than anyone adopting a realist philosophy of science, as well as the ability for physicists to be more productive when doing physics, being less bogged down by worries surrounding paradoxes.

2.1.3 Language and a Working Definition

This section covers two necessary starting positions this thesis: an outline of the how philosophical and scientific terminology will be approached throughout the thesis and a bare definition (more like a working heuristic than a definitive answer) of "what physics is". The former is required to attempt to speak cogently to both physicists and philosophers, as this thesis aims to be understandable and coherent to those who work solely in physics (who would be implementing its recommendations). The latter is to help provide a conceptual framework of physics that both realists and non-realists can agree is representative of at least the most core aspects of physics.

When attempting to synthesise work in both philosophy and physics, it becomes important to define our terms as clearly as possible. Physicists, philosophers, and the average layperson in the street all use the same words, constructed with the same sounds produced via our larynxes which can mean, or at least imply, slightly different concepts. These different concepts can in turn cause confusion, which can lead to problems in cleanly imparting information from one party to another. In considering this, consider the classic question "is a tomato a fruit?". For the average layperson, the answer to this question is probably no: when I go to make a fruit salad as an addition to a nice lunch, I never consider chopping up a tomato and throwing it in the bowl, whereas I would if I was making a salad constructed of various vegetables. By this culinary metric, the answer to our question is no. If, however, I was to ask someone with a keen interest in biology, they may answer yes, as they would be aware of the technical definition of "fruit":

> *"A fruit is a mature, ripened ovary, along with the contents of the ovary. The ovary is the ovule-bearing reproductive structure in the plant flower. The ovary serves to enclose and protect the ovules, from the youngest stages of flower development until the ovules become fertilized and turn into seeds"* (Kelly, 2014).

Under this definition, a tomato is absolutely a fruit, irrespective of its use in the kitchen. Furthermore, we could consider the legal precedent established within the United States Supreme Court by Nix vs Hedden (1893), where it was established for the purposes of tariff law that a tomato is a vegetable rather than a fruit, following the "ordinary" every day linguistic conception. The answer to our simple question is less obvious than at first glance and is obviously context dependant. This example is adapted from Englehardt's paper on the linguistic division of labour, and serves as an good example as to how words can have differing meanings in differing contexts (Englehardt, 2019, pp. 1860-1863).

For this thesis we can run into similar issues regarding words like "true" or "false", and concepts like "knowledge" and "proof" can mean different things to different people working within different disciplines/contexts. Following the work of Jeff Englehardt on the "division of linguistic labour", I am seeking to define these terms both within their broadest possible context and with respect to their own areas of specific expertise. For example, for words like "knowledge" I will use the layperson's understanding of the term, i.e. "knowing (a) thing(s)", as opposed to a philosophical definition referencing justified true belief or any other epistemological construction. I do this not to cause to confusion, but to reduce the philosophical complexity of the term for the physicists this thesis is equally aimed at. Equally, the quantity and technical depth of physics terms will be kept as low as possible, to help the

inverse problem. All these terms have deep levels of situational context that goes into their division of linguistic labour and following Englehardt again I seek to define my contexts here explicitly, so as not to accidentally stray later from one pathway to another. In short, ordinary language will be used as default throughout this thesis.

Another useful and important part of this viewpoint is trying to define how we define the area of study that is "physics". In this, I take as broad a brush as possible again, defining the scope of physics to be 'all things researched in "Physics" departments at universities and taught by "physicists" to other "physicists"'. This broad definition is intended to neuter arguments about whether we should consider, for example, theoretical physics the same functionally to experimental work, and where we draw the line between physics and applied mathematics. The questions are beyond the scope of this thesis, and this thesis at least begins by taking physicists at their own word: if they self-define as a physicist, for the purposes of this thesis they are a physicist.

The next important task to accomplish is to present a view on "what physics is" in a methodological sense. Is physics 'the study of the universe that allows us to access the objective reality around us via empirical methods?'. Is it 'the process by which we create models that allow us to explain and describe the functioning of the universe?'. Or is it in fact some middle ground between these two positions, one not metaphysically committed but still explicitly grounded in the knowledge we can acquire from the world around us? This question is a thesis in and of itself, so the task here is to merely present an initial structure from which we can analyse later work, one which is as uncontroversial as possible and can be accepted as at least representing some of what physics is to both the scientific realist and scientific anti-realist. In doing this I define realism and anti-realism as broadly as possible, as is shown in footnotes 1 and 2, and show both parties can agree to it at least within the limited scope of it being 'somewhat close' to what physics is.

Before doing so, it is important to understand what is meant here by "definition". Nancy Cartwright makes the case that definitions are themselves built from concepts that can have multiple different interpretations, and this leads to a vagueness in even the most technically rigorous definition (Cartwright, 2022, pp. 21-31). Rather than seeking to fix this problem, which operationalists and others have sought to do for decades, this thesis simply accepts that it can't be done, and therefore any 'definition' presented within this thesis cannot be fully grounded. Instead, the concept here is to provide a working definition which can be seen to be

pragmatically successful, i.e. it matches the day-to-day activities of physicists well, and isn't completely incompatible with realist or anti-realist schools in the philosophy of science.

For a generic scientific realist, we should take a positive epistemic attitude to the work of science, and therefore recommend to people belief in both the observable and unobservable content of our theories. Extending this, a realist would be comfortable with the assertion (at least in its reduction) with the statement "physics is the subject in which we can, via belief in our theories about the word, know the structure and dynamics of the universe". The statement contains within it metaphysical content, i.e. we can say X is existent in the universe because X is contained within our theories, and we believe those theories.

The anti-realist doesn't share this positive epistemic attitude and doesn't accept that we can acquire beliefs about unobservables through our theories. For them, the subject 'physics' doesn't necessarily have a privileged metaphysical position with regard to informing our metaphysical beliefs about the universe, and any attempt to do so is flawed because we do not have a method as humans for successfully believing in central aspects of our theories. For our 'definition' of physics to be baseline, therefore, it must come somewhere between these two positions.

Following the work of Cartwright on definitions and concepts[16], I propose the initial working definition of physics for this thesis:

"Physics is the subject in which we use empirical skills to form theories that allow us to make falsifiable predictions about future events in the universe around us".

This vague definition allows us to focus future discussion around physics' success as a knowledge gaining or practical tool, whereby we use our epistemic abilities to produce theories that give empirical reproducible date experimentally. If a theory allows us to predict accurately what will happen under a set series of circumstances, it is a good theory. There are no implied metaphysical conclusions we can, or should, draw away from such predictions, merely that the theory is currently acceptable to those using it. This definition should be acceptable to both the realist and anti-realist and is compatible with the methodology of science as practiced by physicists every day.

[16] I must accept that any 'definition' presented here will be subject to the same issues any other definition would be with regard to its reliance on pre-existent, vague concepts, and that any discussion on the definition of physics is so broad as to be a thesis in and of itself. Like most definitions presented within the thesis (such as "observability" in constructive empiricism), it is best to think of this as a good working heuristic, as opposed to an objectively way of describing what physics is.

2.1.4 The Working Physicist

The final part of this section takes the work of the previous two sub-sections and uses them to define a generic, idealised "working physicist", against whom we can measure the success or failure of the project attempted by this thesis. The following assumptions are made about this working physicist:

1. *Their understanding of "physics" and "science" is at an expert level, but their understanding of philosophy, and the philosophy of science, remains non-formalised, or at least unacademic.*

This principle means that we can assume that our protagonist would hold an expert level of understanding about any term used in physics, the scope of which we can describe arbitrarily. For example, they would be able to use, within their work, concepts like "entropy", and equally utilise fully all knowledge that derives from the concept of "entropy". However, their lack of formal philosophy training or education means we can assume knowledge no greater than that of a layperson when it comes to philosophical concepts. This assumption is grounded and made reasonable primarily from a pedagogical basis: it is necessary for all physicists to have been trained in physics to be physicists, but it is not necessary for them to have been trained in philosophy. Consequentially, we can assume that the philosophical tendencies of our working physicist derive more from general societal beliefs about the philosophy of science than formal philosophy of science academic work[17].

2. *The working physicist has 'realist intuitions' about their work and the work of physics more generally.*

Following on from the previous assumption, given our physicist has something closer to a layperson's understanding of the philosophy of science than a grasp on formal academic philosophy, they are likely to hold 'realist intuitions' about the work of physics. This means that they perceive their work, and the work of physics more generally, has access some aspects of existence or truth about the universe, via the theories of physics. This physicist may not fully understand the distinction between observables/unobservables within philosophical discourse, but when they see images on a screen from an electron microscope, they believe that they are *seeing* entities that *exist* within the universe. If you were to ask such a physicist

[17] Obviously, some physicists have and will continue to interact with academic philosophy work. The "working physicist" introduced here is meant not to account directly for those individuals, but to describe a proto-typical physicist educated purely in physics, operating as a sort of 'lowest-common denominator'.

“do viruses exist?” they would answer in the affirmative and give the question no greater thought: it’s obvious to them that viruses exist because they can be seen though imaging technology.

This ‘realist intuition’ I take to be something caused both by societal and pedagogical pressures. Nancy Cartwright talks about the three different incorrect views on science that are pervasive within society and inform this almost accidental slip into realism, the most important one for this problem being that all the sciences are reducible to physics (Cartwright, 2022, p. 8). If we believe that wildly diverse subjects such as psychology, economics and biology can all be derived from physical first principles, we necessarily imply that physics has some form of privileged standing within empirical knowledge-seeking, which in turn can only derive from physics’ stated direct connection to the most fundamental workings of the universe. Furthermore, the only viable reason that this matters is if physics has some privileged metaphysical status, i.e. physics is the only subject (of the sciences at a minimum) that allows direct metaphysical access to the universe and can talk about things “existing”. Is it such a surprise that physics has a realist lean when these unstated but deeply important chains of logic are not just present within the orthodoxy, but aren’t frequently interrogated by physicists themselves?

Pedagogical pressures toward realism also exist. These pressures stem not from the conscious decisions about the anti-realist/realist debate, but from unconscious choices made by teachers, examination boards and governments when it comes to presenting or discussing the qualities of being a good physicist or scientist. As such, declarative examples that “physics education is realist or realist leaning” are hard to identify directly; the very manner in which they present a realist orthodoxy makes finding strong statements for or against this realist *milieu* either non-existent or abstracted from the strong point we are looking for. However, despite this, hints are available within syllabi that show a consistent usage of language that, when considered philosophically, helps to form the picture that the way we are taught physics in schools encourages realism as a default position.

Take for example, the AQA 2016-2018 GCSE Physics course overview. This document, which forms a guideline for physics education of 14- to 16-year-olds in England, is traditionally amongst the first interactions young future scientists will have with physics as a separate science. I contend that the framing of the course will inform the core beliefs a future physicist will have about the subject. The future physicists who took this course would at this stage (Q2 2024) be starting on postgraduate physics research, probably having no structured philosophy

of physics training at any stage. This course outline contains learning goals that outright lean realist when it comes to analysing data and understanding empirical experimentation. In Chapter 4 of the AQA GCSE Physics course overview, "Working Scientifically", states that skills student should learn over the course include "[being] able to … [be] objective, evaluat[e] data in terms of accuracy, precision, repeatability and reproducibility and [identify] potential sources of random and systematic error". To demonstrate this skill set, the AQA course guideline also states that students should be able to understand concepts such as "An accurate measurement is one that is close to the true value" and that "Systematic error is due to measurement results differing from the true value by a consistent amount each time." (AQA , 2015, p. 15). These statements are themselves fairly uncontroversial; good physics practice requires consistency of data analysis and a reproducible methodology, but the language present, with words such as "true" and "objective", silently encourages a philosophical outlook that only veers realist. You could replace the word "true" with "empirically accepted" and be no less valid a physicist methodologically. "True" is a word that is metaphysically loaded, strongly in philosophy but also in natural, layperson's usage; having it present within a course structure will affect how metaphysically loaded the pedagogical presentation of a subject is.

Because these issues aren't interrogated within adolescent physics education[18], this innate set of beliefs and intuitions is left to form and build, with no structured philosophy education that provides differing contexts or ways of understanding physics. The AQA A-level course (active from June 2017, for 16–18-year-olds, which would be the next natural progression for our GCSE students) has no mention of any philosophy of science (AQA, 2017). The institution under which this thesis is being written, The University of York, at the time of writing offers no structured philosophy teaching in its MPhys (Hons) Physics course (The University of York, 2024). An incoming PhD candidate, educated in physics in a perfectly orthodox manner (AQA GCSE, AQA A-Level, The University of York integrated master's degree) could have no philosophy of science education and thus no context in which to place their understandings of what a "true value" is or the competing views on "objectivity" within physics. This example, which this thesis takes to be an orthodox and common manner in which people become educated research physicists, demonstrates the pedagogical pressures encouraging realism in otherwise non-philosophical physicists.

[18] The word "philosophy" appears just once in the 104-page course syllabus given by AQA here, describing their 'philosophy' that science education is for all and that they have a range of courses for all educational needs (AQA , 2015, p. 5)

Now that the case for our working physicist having realist intuitions has been made, it is important not to overstate the point. Our working physicist may have these realist intuitions, but they are not necessarily full-blown scientific realists. Their commitment to these ideals do not have to be fully fleshed out, they are not philosophers after all. Instead, what they are almost certainly committed to, by the very nature of the subject they are studying, is that physics allows insight into the dynamics and functioning of the universe. Rather than this being a metaphysically loaded concept, this at is simplest can function as a purely a belief that we should, at a minimum, expect our physical theories to have predictive power, and that if they successfully predict observable events, they are currently acceptable. We can use $F = ma$ in a variety of different contexts despite Newton's work being superceded by Einstein's purely because it still holds strong predictive powers within non-relativistic contexts. This, of course, returns us to our definition of physics in 1.1, which I take to be the very minimum definition of what physics is and does for both the philosopher of science and the active working physicist.

2.1.5 The Problem at Hand

Now we have established a way to frame the language within this thesis as well as the way key terms are used, a necessary starting point for defining what physics is and a protagonist/yardstick against to measure future work, the central argument of this thesis can begin in earnest: realist philosophies of science don't work within modern physics. Further to this, they can never work whilst maintaining any level of metaphysical commitment. Realist philosophy of science lead, through the maintaining of realist orthodoxy from realist philosophers to intuitively-realist-physicists, to worse physics. A subsequent goal of this thesis is to demonstrate the benefits of constructive empiricism over realist philosophies of science. By adopting constructive empiricism, our working physicist can maintain their realist intuitions and everyday practices whilst removing any pernicious metaphysics that may lead to the creation of problems within technical physics work. This chapter, Chapter 2, is initially focused on the task of introducing this problem and grounding it both within the language of scientific discourse, the practice of philosophers and demonstrations of places where physicists themselves stumble upon this problem, often without correctly identifying it. After doing this, Chapter One then introduces Bas van Fraassen's constructive empiricism, a less dogmatic anti-realist school, which I contend our working realist may already be, even if they don't realise it.

2.2.1 Realism – The Hidden Axiom

In Sections 2.2.1 through 2.2.6, we will address the problems of having realist intuitions, how it constrains the thinking of physicists and enters into their worldviews through the language, educational priorities, and moral choices we use about science.

Realism is the school of thought that holds that we can, through our scientific theories, understand what does or does not exist within the universe, and that our theories are approximately "true" because they accurately describe the world around us. Such a mindset is epistemically optimistic about the ability of humans to know whether things are true or false; for a realist, human beings can answer questions like "do electrons exist?" accurately, and our theories are the framework within which we answer these questions.

This optimism, however, is not something that can operate within a vacuum. Our everyday experience of the world is structured around beliefs about the value of concepts like 'truth' and 'falsity', which we can see from areas as far apart from each other as politics and organised religion. The "search for truth" is always something to be prioritised, irrespective of the actual epistemic abilities of an individual to perform that search successfully.

Physics, and the practice of science more generally, is not immune from these social norms, as physics operates within the society in the same way any other subject does. The practice of physics does not operate with on a day to day basis with these philosophical questions or problems hanging overhead, and therefore I maintain that these unconscious biases towards realist intuitions both exist within physics and affect its practice, leading eventually to worse physics being done as we enter realms beyond the epistemic reach of humankind, where the epistemic optimism of realism can no longer hold.

To begin this part of the Chapter, we will discuss two examples where the actions of physics are considered to have failed, one in terms of "wasted work" and the other in terms of a "failed experiment". I will then discuss the "moral case" for realism, and how that further affects the practice of physics in ways that aren't in and of themselves physical. After this, I will introduce some examples of paradoxes in physics I think are affected by this set of problems, before concluding that our best solution is to find a philosophy of science that allows for these realist intuitions to co-exist with a rigorous removal of both an epistemic optimism about human ability to know things to be "true" and any metaphysical content within our scientific theories.

2.2.2 Negative Language in Science

The language we use when discussing science in ordinary contexts has the potential to be evaluative. As human beings, we see "success" as being linked to being "right" about a certain concept or theory, and "confirmation" of any arbitrary theory is emotionally more satisfying than the "negation" or "failure" to produce expected outcomes. This is a hard position to prove via the words and thoughts of physicists themselves; either they have not considered the problem and therefore have nothing to say on the matter; or they have they claim to operate in such a way that these implicit biases have no effect on them[19]. However, we can observe aspects of these unconscious (or at least unspoken) biases in the history of physics and the pedagogy that underlines physics education.

Firstly, when going through the classic stories told about progress in physics, we are more frequently informed about success that led to theories still in use, than work done down dead-ends and rabbit holes. Isaac Newton's description of gravity is still taught widely, *Principia Mathematica,* is still an important text that training physicists are encouraged to read. We even have a well-known piece of apocrypha about what led to his work on gravity originally (the apple falling from the tree). Second to Newton's work on gravity, people are still taught his work in optics, where his text *Opticks: or, A Treatise of the Reflexions, Refractions, Inflexions and Colours of Light* is still considered a founding text. What people are not taught about, at least in the context of the physics classroom, is Newton's decades of work in alchemy, attempting to produce gold from lead, without any success. The Michelson-Morley experiment of the late 19th century was intended to observe and discover properties about the "luminiferous ether", the medium in which light-waves propagated through space[20] (Encyclopedia Britannica, 2024). Instead of demonstrating the ether existed however, the experiment failed to produce anything close to the results expected, demonstrating that there was no ether in space, merely a vacuum, and that our theories on light were incorrect. In a letter from Michelson to Lord Rayleigh in 1887, Michelson himself says "the Experiments on the relative motion of the earth

[19] It is of course also possible that a physicist has considered this issue and endorses these evaluations. However, our "working physicist" by our definition will not have, as to do so involves extra consideration outside the scope of their direct physics education. In this case we take the layperson's understanding of terms such as "truth" and apply them onto the working physicists understanding of the world, which is conceptually similar to the concept of "is real" or "is how things are in the world".

[20] The idea that light waves needed a medium through which to propagate is based in pre-wave-particle duality theory of light, and was based on the fact that 1) light is a wave and 2) waves in all other contexts require some medium in order to travel between points, in much the same way that soundwaves require some form of substance to travel from A to B. This led to the theory that there must be some entirely non-viscous substance across all of space that explain why we could see light-waves from the Sun, the "luminiferous ether".

have been completed and the result is decidedly negative" (Shankland, 1964, p. 32). The immediate reaction to this experiment at the time, even though it is now considered one of the most important milestones on the way to our currently successful theories on the speed of light, was that this was somehow a failure of science. This is still considered one of the most important "failed" experiments in the history of science, and one of the few still to be taught.

What do these two examples of an unspoken attitude toward realist dogmas within the teaching of physics tell us about the intuitions and beliefs of physicists? Both share in the problem of confirmation biases. Newton's work on gravity and optics advanced human knowledge and understanding *positively* by providing a set of theories that have strong predictive powers about future events. Newton's work on gravity allowed us to calculate what happened to an apple when it dropped from a tree mathematically, which we could measure up against the things we actively observed when an apple dropped from a tree. For Michelson-Morely, the immediate reaction to the experiment was that the experiment failed to provide any answers as to how light moved through space; in doing so it had seemingly undermined current best working theories. Lorentz, in an 1892 letter, wrote "I am totally at a loss to clear away this contradiction [between the Michelson-Morley experiment and Fresnel's ether hypothesis], and yet if I were to believe we were to abandon Fresnel's views, we'd have no adequate theory at all" (Shankland, 1964, p. 32). In retrospect, we view the Michelson-Morley experiment as an important catalyst for developing new theory and knowledge, but leading physicists at the time did not see it that way.

Both examples talk to the realist intuitions of physicists past and present. We still, despite our methodological desires, prioritise theories, work and experiments that advance knowledge in some productive way, and either disregard or dismiss work that fails to do so. Newton's work on alchemy could have been performed methodologically in full keeping with scientific practice, but as it did not manage to demonstrate how one could turn lead into gold, it is considered "failed" or "wasted" work. However, an equally valid interpretation of this work is that it continued the gathering of knowledge, through repeated failures, in a way that allows us now to know that alchemy is a failed area of research. Evidence gathered through negation or failure has the power to be as powerful as evidence gathered via a "successful" experiment.

What, however, is the issue here? The realist may argue that these biases do not prevent the creation of good physics. In fact, every day we see the progression of physics and science as new theories are made, new knowledge discovered, and better, more complete models that are empirically adequate. The implicit realism in our language of science may exist, but its practical effects don't harm physics and may in fact motivate it, giving reason behind the search for

knowledge that is physics. This argument isn't entirely wrong, physics still evidently works and provides positive utility in our search for knowledge about the universe. However, this thesis maintains that the progress of science could be *better* if we weren't encumbered with the metaphysical baggage that realism necessitates. Instead, what is required is an approach outside of the realist/anti-realist paradigm, which allows our realist intuitions and motivations to exist whilst not letting realist intuitions take us beyond what is required for physics to be a successful subject, and which may in fact make it less successful.

Let us introduce one area of physics, discussed in more depth in Chapter 4 and initially explained in Chapter 1, where the problems outlined above contribute to the creation of confusion and "paradox" within the work of physicists: black hole dynamics.

2.2.3 The Page-Time Paradox – Intuitions and Pathways

Here we return to the Page-time paradox. As we have already covered this material in some depth in Chapter 1, this section and the following sections will address primarily the philosophical problems at hand, reintroducing physics content when necessary.

> "According to one definition (Quine, 1966), a paradox is an apparently impeccable argument to an impossible conclusion — such as a pair of apparently impeccable arguments whose conclusions contradict each other. By this definition, the Page time version of the information-loss paradox is a true paradox" (Wallace, 2017, p. 16).

The Page-time paradox is a problem of physics that highlights the ever-widening gap between differing specialisms in physics and how the intuitions of physicists can become the central motivation for theory creation, in the manner of philosophy. As the complexity of the science has increased, becoming both less intuitive and increasingly fine-grained, different areas of physics have increased their knowledge bases to the extent that one cannot be an expert in all things. This is not a fully modern shift. Increasing specialisation has been a process of hundreds of years, but it can lead to problems for physics *as a whole* when different areas within physics collide and produce mutually inconsistent results, where we cannot simply arbitrarily prioritise one area over another. The Page-time paradox is an example of one of these issues, specifically between inconsistencies in the quantum mechanical understanding of black hole dynamics, and the statistical mechanical approach.

In order to explain both the paradox and the issues for our working physics, let us revisit what a black hole's entropy is with regard to an external observer. As is shown in Chapter 1, Eqn 2. the

entropy of a black hole is stated by the Bekenstein Entropy equation for black hole entropy (Bekenstein, 1973, p. 2338), which is simply:

$$S_{bh} = \left(\frac{1}{2}\frac{ln2}{4\pi}\right) kc^3\hbar^{-1}G^{-1}A, \quad (4)$$

Where S_{bh} is the entropy of the black hole, A is the black hole's surface area, and $kc^3\hbar^{-1}G^{-1}$ is a constant. We can conceptualise this simply by stating that the entropy of a black hole is directly proportional to its surface area[21], and thus as the black hole takes in mass (via things falling into its singularity) its size and surface area will increase, and thus will its entropy. This is an implicitly informatic understanding of entropy, because the entropy we calculated outside the black hole via this equation is directly related to the amount of mass we cannot view, due to the laws of physics that describe the event horizon.

The other important entropy property a black hole possesses is related to the process of Hawking radiation, where over time due to particle/antiparticle creation occurring at the black hole event horizon, a black hole can lose mass, in effect "evaporating" away. As time passes for an isolated black hole (i.e., a black hole with no infalling material via which it increases in size), the number of these quantum-physics described particles will increase. Each of these particle/antiparticle pairs, one inside the event horizon and one outside, will have an entanglement entropy that helps describe their relationship. As time passes, the cumulative entropy of these entanglements will also increase, in a manner directly related to the increase in these particles.

The paradox is generated by the inversely proportional relationship between these two entropies. As the black hole "evaporates" away, the surface area of the black hole decreases, meaning the Bekenstein, or "microcanonical", entropy of the black hole also decreases. However, during this evaporation, the number of entangled particles increases, leading to an increased entanglement, or "von Neumann", entropy for the black hole. Importantly, the microcanonical entropy of the black hole always acts as an upper bound for what the von Neumann entropy of the black hole can be (as discussed in Chapter 1). Thus, as the black hole continues to shrink via Hawking Radiation effects, the von Neumann entropy of the black hole must also decrease in line with the microcanonical entropy. Page calculated (Page, 1993) this occurring as the black hole reached half of its initial size, which is equivalent to time $\frac{t}{2}$, where t

[21] Talking about the "surface area" of a black hole, much less a 4D object, can be confusing, but it is much easier to understand it as the surface area of a sphere, where this area is defined by distance between the singularity at the centre of the black hole and the black hole's event horizon, which in effect acts as the "surface" for the black hole for the purposes of understanding black hole entropy.

is the time taken for the black hole to fully evaporate[22]. The challenge for the physicist is explaining what occurs after this time and how we recover the lost information within a setup obviously macroscopic and unlikely to be the domain of exotic quantum physics. There is a direct clash between the predictions, theories, and axiomatic premises of statistical mechanics and QFT. To reject one in favour of the other would be to undermine one or either of the areas of physics entirely, and thus it currently remains unresolved.

This doesn't mean that there have not been attempts to say precisely what occurs after the Page-Time. As discussed in Chapter 1, two competing and distinct theories exist in "black hole complementarity" and "firewall theory". These competing theories precisely demonstrate the dynamic in physics that this thesis argues causes problems for physicists when it comes to producing theories that have predictive power, fulfilling the core requirement of physics. Here is an example of work that is classified as "physics", done by "physicists", that more closely reflects the rationalist work of Enlightenment philosophers. Rather from empirical principles, as you can't within the context of black hole information loss paradoxes, here physicists are engaged in taking a selection of stated axioms about the functioning of the world and following them deductively to their conclusion. The idea of two competing theories having different axiomatic beginnings is not in and of itself disqualifying in terms of producing good theory, but how unmoored these theories are from empirical evidence[23] again demonstrates how far these physicists have gone from doing "physics". And yet, they remain physicists all the same, in terms of how society defines them.

Erik Curiel makes a similar case regarding other problems in black hole physics, seeing the work being done (in this when discussing theories on singularities) as closer to the work of philosophers of science as opposed to physicists, stating:

"Here again, as with almost all the issues discussed up to this point in this entry regarding singularities and black holes, is an example of a sizable subculture in physics working on matters that have no clearly or even unambiguously defined physical parameters to inform the investigations and no empirical evidence to guide or even just constrain them, the parameters of the debate imposed by and large by the intuitions of a handful of leading researchers. From

[22] It is important to note other times for the Page-time have been calculated. This is discussed again in Chapter 4.

[23] We have not seen what happens to a black hole beyond its Page-Time, and we could never, according to the basic principles of general relativity and the event horizon, see any effects (at least from outside the black hole) that happen *within* the black hole event horizon. We could simply say that the reason for this is our theories about the Page-time were wrong, but doing so would mean dramatically reassessing work in QFT and stat mech that is strongly evidenced by other sources, which seems unlikely.

sociological, physical, and philosophical vantage points, one may well wonder, then, why so many physicists continue to work on it, and what sort of investigation they are engaged in. Perhaps nowhere else in general relativity, or even in physics, can one observe such a delicate interplay of, on the one hand, technical results, definitions and criteria, and, on the other hand, conceptual puzzles and even incoherence, largely driven by the inchoate intuitions of physicists." (Curiel, 2019, § 4.0)

The example of the Page-time paradox, and the arguments it produces, help throw into sharp relief the problem: Physics, and physicists, have a bias toward realist language, which itself produces realist beliefs about their theories. These realist beliefs, however, do not always fully align with the epistemic capabilities of the subjects or its operators, meaning we end up with beliefs about our theories that cannot be backed up purely through physics work. This further means that they end up accidentally straying into pseudo-philosophy (or other non-physics methods for producing knowledge). Despite this, however, we cannot simply undo these realist intuitions: they are baked into us via society's desire for "truth" and the very teaching of the subject. Further to that, we may not even *want* to remove these realist intuitions, as they provide strong psychological motivation in our search for knowledge about the universe (finding "truth" being a powerful motivating force).

The next subsection will reintroduce our working physicist and lay out for them the possible pathways they could take when addressing this problem: transition from *de facto* realists to *de jure* realists; abandon realism in favour of anti-realist schools; or a third way, in which we attempt to account for these realist tendencies whilst not maintaining any metaphysical connection between our theories and the universe at large.

2.2.4 Our Working Physicist's Best Approach

Let us return to our working physicist. This individual operates with realist intuitions but is not necessarily tied to them outright, but if the question posed to them was "given your knowledge of science, do you believe that your current physical theories are true?" they would be put in an unenviable spot. The section 2.2.3 shows how taking a naïve realist view is *prima facie* impossible (two theories we take to be true having consequences that contradict each other). Yet to deny their own realist intuitions would be to understate the importance of both the societal forces that produce such realist intuitions, and the fact those intuitions are often predicated on the fact that empirically adequate theories, given enough time and reconfirmation, become short-handed for "true" within our paradigms. We could consider it "true" (in a loose, non-philosophical sense) that apples that fall from trees, on this earth, with

an average acceleration of $9.8m/s^2$, *ceteris paribus*. This isn't because it is "true" (in the accurate, philosophical sense of the term) that apples fall from trees at that rate, but because our experiences lead us in the direction of observing a system that always has, always is, and always will be that way in the average human experience.

For this reason, if we are to continue with the concept that our working physicist shouldn't just, by default, end up a realist because of their intuitions, there needs to be problems within the realist worldview that are immediately evident to those without philosophical training. I make the crossing the bar high for this premise for two reasons:

1) This thesis, whilst based in technical philosophical work, is aimed primarily at either convincing physicists to adopt certain viewpoints or explain that they naturally hold them in the first place. To do so, this thesis will require both technical discussion of philosophy (to maintain the consistency and validity of the thesis) as well as more high-concept approaches, designed to apply to those physicists directly. The problems and solutions should follow logical reasoning that is as simple as possible, and,
2) The sensitivities and motivations of said realist-inclined physicists must be correctly accounted for. A substantial number of the consequences of the best possible solutions for this problem involve compromises or paradigm shifts that might be initially uncomfortable to the physicist, especially with regards to observation or the practicality of solving individual problems in physics. To reduce the prominence of these worries, the problems at hand have to be so severe as to make the compromises or discomfort worth it.

This, and the next, section, therefore, will return to the Page-time paradox specifically with regard to black hole dynamics and entropy, demonstrating that whilst our *current* physics has good workarounds to potentially tricky issues, a general lack of consensus (or, in the case of some thinkers, a lack of *ability* to find any consensus) means some problems lack even a conceptual process whereby we could solve them.

2.2.5 The Problem Facing the Working Realist

Continuing from the last section, for the intuitively-realist physicist, an initial question to consider is "Why does the Page-time paradox arise?". For this, there are three immediately apparent possible routes:

1) Wallace is wrong. This isn't a true paradox, and as time goes by and more productive work is done, this "paradox" will be resolved by some, as yet, undiscovered work within our current understandings of statistical mechanics and QFT.
2) The foundation axioms of physics are wrong. The Page-time paradox is built on competing axiomatic claims that cannot be breached in either stat mech or QTF, because doing so fundamentally undermines the structure of either field. Perhaps the Laws of Thermodynamics, the "no-cloning principle", or "unitarity principle" are incorrect, and future work within physics will demonstrate this, allowing a resolution (this is similar to route 1, but with more generalised effects for physics as a whole).
3) We can't do good physics here. We have reached, and attempted to surpass, the empirical limits of human understanding, meaning we cannot apply the methodological principles of science to the entities in the Page-time paradox. Rather than being a failure of physics *per se*, this paradox is the result of physicists reaching the fundamental limits of physics as a way of acquiring knowledge, and the paradox is nothing more than a symptom of trying to conceptualise a system we don't have the requisite tools for within physics.

1) and 2) are related problems to different extents. In both, there is some flaw within physics as we currently understand it that is fundamentally resolvable *within the current methodologies of physics* given enough time, thinking and resources. They broadly outline that given that our current theories taken together imply a contradiction, they can't both be simultaneously true, so something must be wrong with at least one of them. This can, and should be, solved by the constant process of science, creating, and destroying theories as it progresses until the point that it is internally coherent (even if both cases lead to differing levels of change and discomfort for the physicist). 3) is altogether different, as it proposes that it is the *application* of physics thought in and of itself that leads to the paradox arising. In 3), there is some aspect of the theorised entities, space or objects involved in the Page-time paradox that are beyond the scope of physics, and the introduction of these things cause a fundamental breakdown of our understanding of the system, leading to the paradox.

For our working physicist with realist intuitions, neither 1) or 2) should present many earth-shattering challenges to their confidence in scientific realism. We know that as time progresses theories we consider nigh-on axiomatic can in fact be wrong, but the realist already has ready-made argument structures for solving this quandary. Further to that, our realist-inclined physicist could say that, if responses 1) or 2) were correct, generating paradoxes such as the Page-time paradox is an example of *good* physics. Without discord within the subject, there are

no new challenges to overcome, and no new knowledge to be discovered. If 2) is correct, the process by which we discover this new knowledge will be more arduous and challenging, one that could potentially upset the knowledgebase of physics holistically, but that isn't to say it couldn't be done.

It is only pathway 3) that seemingly holds a strong, defeating response to the realist-inclined physicist. Rather than the paradox causing a new revolution in physics thought, a process that has occurred many times throughout history, if 3) were correct the realist-inclined physicist would have to wrestle with the idea that this problem is not inherently *physical*, but instead an insight into the limits of physics itself. Is there any reason to suspect 3) to be correct over 1) and 2)? Why should we suspect that the Page-time paradox, rather than being another example of a problem in physics which will be resolved, allows us insight into something more fundamental about the methodology and limits of physics?

We can investigate the limits of physics by examining the boundary conditions of our universe. Physics is fundamentally a subject based within human observations of the universe around us that limits it to both the epistemic range of human observation and the "physical" world which we inhabit. This range is naively very large; it contains within it all of human external experience and any concept, process or structure which is potentially knowable to the human mind. A law of physics, if correct and accepted, should be generalisable across all human experience and equal across all space-times, giving qualitatively identical observations with qualitatively identical initial conditions. Physics is above all an *empirical* subject. Theory acceptance or refutation is predicated on the acquisition of knowledge accrued via our senses.

It therefore follows that physics has tangible limits with regards to the theories it can talk about. We can develop solely empirical theories about why an apple may fall from a tree, but we cannot do the same about entities outside the scope of our observable universe (for example, God). We can intuitively categorise most entities and concepts that can be subjected to the methodologies of physics' investigation, the simplest boundary line being "spatiotemporally located things within the universe". Modern physics, however, interacts with both smaller and smaller entities (within quantum physics) and more complex macroscopic entities that fit uncomfortably within initial, naive intuitions about a finite universe (e.g. black holes, singularities). Black holes hold an interesting position within this framework. Rather than simply being some macroscopic object, like the supermassive stars that may have created them, the consequences of general relativity dictate certain interesting properties they must hold. For example, the event horizon, the boundary of the 'black' area of a black hole, must both be some

form of boundary condition upon which some physics operates[24] but also, due to the Principle of Equivalence, must be unobservable to any observer passing through it. To an observer external to the black hole however, for example one on Earth, the event horizon is the limit of our epistemic range. Any infalling matter (and all information about its construction etc), that passes the event horizon is no longer accessible to an external observer and, assuming that barrier is permanent, will never be accessible again. The "internal" region of a black hole (i.e. the region of a black hole inside the event horizon) is inescapably inaccessible to a human observer external to the black hole as long as the laws of General Relativity hold.

The thesis, whilst agreeing with Curiel that one of the largest problems facing black hole physics is physicists' inability to deal with the "conceptual puzzles and even incoherence" they themselves produce, takes an optimistic view about what can be done. Rather than simply decry this work as unimportant, I believe that it is deeply relevant not only to physics, but an often unstated and unrecognised aspect of theory creation. Physicists should be allowed to maintain their "inchoate intuitions", both as an acceptance that they are humans who operate in the world and as a way to drive and create new and better theories. However, "physics" as this monolithic process for finding information about the universe cannot work without being self-aware as to its limits; it has to recognise them and understanding the pragmatic choices present in its construction. The next section will demonstrate that current strongly anti-realist philosophies cannot, or will not, be adopted by physicists as a whole, because of these valid realist-intuitions. What is needed is a better approach, in which we have a philosophy of science that allows for these intuitions and biases to exist but reduces the problems they can cause by correctly demonstrating the limits of physics and science.

2.2.6 Anti-Realism, and The Need for Better Approaches.

Anti-realist schools on the philosophy of science, like realist schools, come in many different forms and have many different functions. For the purposes of this thesis, we will be addressing anti-realism as the school that is, in effect, the antithesis of the realist case, i.e. where realism is epistemically optimistic about science, anti-realism is epistemically pessimistic, and so on. The big issues with our intuitively-realist working physicist adopting an anti-realist approach are:

[24] This is not *strictly* true. Certain physical theories about how a black hole is instantiated post-Page-time (as shown in Chapter 1) are conditional on the existence of a "stretched" horizon. This is a region one Planck length outside the event horizon, mostly introduced due to concerns about these theories being consistent with the Einsteinian Principle of Equivalence (Susskind, et al., 1993, pp. 3749-3750)

1) They are definitionally realist-inclined. This point is self-evident, but the setup of our protagonist physicist is that they have and maintain realist intuitions about their work, and those intuitions are in and of themselves valid. Any attempt to convince them of anti-realist positions, which are in explicit conflict with these intuitions, means having to undermine the unconscious ways in which they view their own subject. I think this prospect is unlikely to have practical utility when it comes to the real end goal of this thesis: helping physicists understanding their subject better within their own paradigm. I am not seeking to be King Canute, attempting to hold back the tide of the expectations and assumptions of a whole subject.
2) Motivational problems. The strongest argument, in the view of this thesis, for realism as a philosophy of science is that it works *with* the desires and hopes of people wanting to generate knowledge from empirical study, not against it. Epistemic optimism is not in and of itself a naïve position. It has its utility in proving a framework for believing we can know more about the universe around us. Without a belief that we can, and should, engage in further and deeper study, new physics does not get done and the whole subject grinds to a halt. What is required is not epistemic pessimism, but full epistemic optimism *within an accepted and rigorous understanding of what humans are capable of*.

Both of these issues prevent us from fully adopting any form of ardent anti-realism. A full pessimistic epistemology that denies the individual strengths of the sciences, and specifically physics, will not be easy to accept for many working physicists. This will not be conducive to providing a philosophy of science that physicists can take forward and use to better their work in science whilst not undermining their day-to-day operation in what is, fundamentally, a useful knowledge generating subject. What this thesis is attempting to offer is not so much a *defence* of anti-realist worldviews as much as a pragmatic case that realism, unconscious or otherwise, is present in the way physics is done, and this is unhelpful.

To make this case, we require a philosophy of science that is anti-realist but otherwise modelled on analysing science pragmatically, i.e. focused on what scientists *do*, not what they *should be able to do*. This philosophy of science should both match a physicist's everyday practice and give them a framework they can use to untangle the pernicious problems that philosophers like Curiel can identify in scientific discourse. We need a philosophy that has a motivating framework for why we do physics, whilst limiting the consequences of extending physics to metaphysics incautiously. Above all, we need a philosophy of science that is firmly grounded in what I maintain to be the core of physics, empiricism, and that can provide a

structure for thinking about science that only relies on being an empiricist. This thesis maintains that that school is Bas van Fraassen's constructive empiricism.

2.3.1 Stances and the Need for a Pragmatic Approach

The better approach we will be using and adapting will be constructive empiricism, but the meta-philosophical structure that will allow us to view it as something outside both strictly realist or strictly anti-realist approaches comes from Boucher and Forbes, in the 2024 (forthcoming in *Synthese*) paper *The pragmatic turn in the scientific realism debate* (Boucher & Forbes, 2024). I will be using this to give a framework to concepts already presented within Chapter 2, and it will be returned to again throughout the thesis:

- Our *realist inclined physicist* can be understood to have their realist inclinations not because of the schools of philosophy of science *per se*, but due to the epistemic *stances* they may take toward foundational questions about the purpose, point, or aim of physics.
- Rather than directly engage in debate as to the successes or failures of one 'anti-realist' school vs a 'realist' school, I can outline the *stances* taken by both the constructive empiricist and myself at any stage of the argument. This should allow us to bypass the realism vs anti-realism debate for the most part, as we will be using the "pragmatic turn" to reframe the argument as a whole.
- Once we have established this reframe, we can use the new framework provided to us by a new understanding of both *stances* and *pragmatic justification* to correctly identify issues that are native to 'physics, the search for knowledge', vs 'physics, the study of what exists in the universe'. This will allow us to parse more concretely why the 'paradoxes' of Chapters 3 and 4 emerge, and what solutions are available for them.

So, what is a *stance?* And what is *pragmatic justification?* In this section I will introduce the concepts Boucher and Forbes present in their paper and directly relate it to actions and beliefs of our realist inclined physicist, whose motivations we have already outlined. This will then allow us to demonstrate why a constructive empiricist approach is superior for our realist inclined physicist at least within certain contexts, despite potential competing desires or biases.

Boucher and Forbes highlight both the existence and the need for the pragmatic turn in the philosophy of science on the basis of the failure of past approaches to firmly put the question to bed despite over a century of literature within the field. This need, to find more constructive discussions within realism/anti-realism debates has brought about the existence of this turn,

which rather than directly addressing either the arguments of realism or anti-realism functions as a meta-philosophical discussion as to *how* and *why* we should have these debates. To take examples, if neither the No Miracles Argument[25] (for realism) or the Pessimistic Induction[26] (for anti-realism) are fully defensible[27], it is impossible to definitively tell an individual why they should choose, for example, the No Miracles Argument and become a realist. Justification for doing so seems to derive from something other than pure belief. (Boucher & Forbes, 2024, p. 4)

Boucher and Forbes further this point by showing that in many cases the realist and anti-realist can't even agree on the battleground for their argument. The structure of the current debated between the NMA and PI seem to adopt the same process for evaluation, one that is naturalised and quasi-scientific (i.e., uses the framework of science to either confirm or deny claims via empirical confirmation). This framework for how philosophers perform this debate isn't itself directly related to the *reason* they hold their positions in the first place, which comes down to what your stance on the meta-argument are. In fact, this approach is not in and of itself unbiased as to which side it prefers: the realist position *is* explicitly naturalised, and even the framing of holding anti-realist positions via 'belief' can be problematic (see more below on stances). (Boucher & Forbes, 2024, pp. 4-5)

These issues generate sufficient pressure in the discourse to create 'the pragmatic turn', which Boucher and Forbes endorse. Its differences with the traditional scientific realism debate can be most adroitly summarised with their own words:

> *"The traditional scientific realism debate is concerned with determining which position is true, warranted, correct, or most rational. The pragmatic scientific realism debate, by contrast, can be thought of as being concerned with determining which position best serves certain values, i.e. is useful, preferable, prudent, or most practical."* (Boucher & Forbes, 2024, p. 6).

This view is, on their account, predicated on the usage of stances and frameworks rather than theories or beliefs; a rejection of the prior established quasi-scientific approach to the debate;

---

[25] The No Miracles Argument is, broadly, that all our theories of physics work so well as a structured whole that it would be 'miraculous' for it not to be pointing toward some truth about the construction of the universe. We shouldn't believe in miracles, thus, realism.

[26] The Pessimistic Meta-Induction is, broadly, the argument that all past physics theories have been shown to be wrong and not predictive of all the observables we now see. Why should we, from our current temporal location, not think that 50 years down the line we are just as wrong as we were 50 years previously? If we think we will be shown wrong now, our current physics cannot hold the metaphysical weight the realist wants it to. Thus, anti-realism.

[27] Directly litigating both the No Miracles Argument (NMA) or the Pessimistic Meta-Induction (PI) would be a thesis unto itself, so again for the purposes of this thesis we are stating it for the purposes of the argument here. Further reading on the topic can be found in Worrall's 1989 work "Structural Realism: The Best of Both Worlds?" (Worrall, 1989).

a centralisation of the pragmatic and values-based aspect of choosing one approach over another and a strongly permissive understanding of what constitutes "rational" within decision-making within the argument (Boucher & Forbes, 2024, p. 6). This then culminates into two centralised themes, which I will discuss here next, that become essential to both understand the pragmatic turn and apply it, *stances,* and *pragmatic justification*.

A *stance* can be understood mainly as a solution to a problem of the empiricist's own creation. Van Fraassen introduced the concept of a stance as a way to resolve a central problem of self-reference within the empiricist's framework of understanding: If all factual beliefs about the world are found via observation of the world (or, more precisely, are contingent and *a posteriori*) and empiricism is a factual belief, it must also itself be developed from our understanding of the world. However, the motivation to believe that all factual beliefs are *a posteriori* is reliant on empiricism. Here we find an issue, where in the words of Boucher and Forbes: "For the empiricist, empiricism seemingly must be both unquestioned presupposition, and vulnerable empirical hypothesis." (Boucher & Forbes, 2024, p. 8).

Further to this, for van Fraassen, a *stance* is not a *belief* but in fact an attitude one takes toward a particular issue. This can be viewed most clearly by relating a stance to, for example, a policy position you can take toward a particular issue in politics. Unlike a belief, which invokes a metaphysical position on the nature of the claim you are making, a stance can be a value-led judgment which one can freely choose to adopt as long as it's rational[28]. For van Fraassen, this helps us solve our current issue by simply reframing the empiricist account: we don't need to *believe* that empiricism is factual in the self-defeating manner we were troubled by, we simply have adopted the empiricist stance, in counter position to the metaphysical stance (van Fraassen, 2004, p. 173).

In his 1995 essay "Against Naturalised Epistemology", van Fraassen points out that this is not only desired but necessary when litigating realist/anti-realist arguments. On his account, "What empiricists [have] shared over the centuries… is not most obviously been a set of beliefs. More in evidence has been denials, in the sense of refusals to believe" and that "[making claims about the success or failure of realist dogmas] is not a belief, but an attitude toward the role of beliefs – a proper aesthetic distance to be preserved" (van Fraassen, 1995, p. 83). When we

[28] Van Fraassen here defines "rational" via a voluntarist framework which is very permissive and non-restrictive, i.e. a position is rational as long as it's not logically inconsistent or necessarily self-defeating. For the purposes of this piece we will accept the entire van Fraassenian position, including voluntarism, but an analysis of it is beyond the scope of this thesis, for more check out (van Fraassen, 1989, pp. 171-174).

have two countervailing world views that don't currently have a way of discussing their differences of opinion, we not only require a way to verbalise that this dynamic is occurring but also a rigorous and well-defined framework, by which we can categorise the different value-judgements that, say, a realist and anti-realist are coming to the table with. Thus, van Fraassen's introduction of stances.

Pragmatic justification adds to the concept of stances by placing them within a larger framework for assessing their success or failure. We may want to choose which stance is more successful out of the metaphysical or empirical ones outlined by van Fraassen, but if we are to remain voluntarist with regard to the rationality of choosing either one, we could be stuck. There is seemingly no way to compare either of these stances directly as they're not simply opposed on the arguments within the scientific realism debate, but on the value judgements that lead into taking one stance or the other.

This problem, of undecidability, is solved via the introduction of pragmatic justification. If we can outline some common goal for "what science is for", even very broadly defined, that both stances can fundamentally agree on, we can begin to assess the success or failure of any singular stance or framework pragmatically. Forbes and Boucher outline their understanding of this thus:

> *"Whether they defend scientific realism or anti-realism, most philosophers of science are "pro-science," epistemologically speaking. Their theories of scientific knowledge are not developed to question the value, actuality, or primacy of scientific knowledge, but rather only to understand its character, nature, extent, and the logic of its methods and development."*
>
> (Boucher & Forbes, 2024, p. 14)

Pragmatic Justification therefore can be summarised as assessing both realist and anti-realist stances or frameworks based on the benefits they provide in reaching some value based or practical goal. In our analogy, we would be asking which of the metaphysical or empirical stance is better if our goal is to extend the knowledge we have about the world, which we acquire through the methodology of science. This then theoretically allows us to answer the question "which is superior, realism or anti-realism?" on an entirely pragmatic basis, removing the theoretical arguments of the No Miracles Argument or the Pessimistic Meta-Induction entirely from the debate.

One potential flaw with this approach can be seen in the ungrounded statement made by Boucher and Forbes in the previous quote. Are "most philosophers of science 'pro-science'" in

the manner described, or is this simply an assertion that lacks substantive weight? For the purposes of this thesis, I will be accepting that statement as stated as I think it matches both the normative and quasi-philosophical approach of *physicists* themselves (I can think of no eminent physicist who would claim that the research they do is not in aid of furthering knowledge in some way, shape, or form), but potential future rebuttals to Forbes and Boucher's "pragmatic turn" from philosophers of science may start here, and further literature will be required in order to answer this question.

From this point onward, we will be using the arguments of Boucher and Forbes to address the central question of this thesis: should our working physicist adopt anti-realist/constructive empiricist worldviews in order to do physics more successfully? Boucher and Forbes handily cover this exact topic in sections 2.2.1 of their paper, entitled "*Should Working Scientists be Scientific Realists, or Anti-Realists (or does it even matter)?"* (Boucher & Forbes, 2024, pp. 9-11)

When it comes to analysing the success or failure of a particular school of philosophy or thinking, establishing the success conditions and context for decision making is vital. For our problem, i.e. given the existence of 'unresolvable' paradoxes within physics, is there a better way of conceptualising the subject, both philosophically and practically, that allows physicists to either resolve them or ignore them? Successfully answering this question centrally requires there to be: a) paradoxes that aren't resolvable, and b) a way of reframing these paradoxes via a specific pathway in the philosophy of science such that these paradoxes are either defused or are practically irrelevant. This thesis chooses to do this with explicit reference to a central 'realistic' character, our already established "working physicist", who is *motivated* to solve these problems without formal *education* in the philosophy of science or philosophical norms and terminology. To convince them we have solved this problem, we require some other success condition that directly relates to their world and their work; for example: "does this new framework allow for me to develop better theories, which are more predictive, with less confusion or compromise?"

Chapters 3 and 4 seek to demonstrate that paradoxes of the kind we are searching for do exist within the physics and are not simply issues of needing to do 'more physics' or 'more thinking', and we will demonstrate that our new pathway can be found via an application of constructive empiricist thoughts. However, when setting the stage for this problem, it is worth considering the history of similar attempts to do what this thesis sets out to do in the literature, the methodologies for doing so, and where this thesis sits within that.

When considering the question "*Should Working Scientists be Scientific Realists, or Anti-Realists (or does it even matter)?",* Boucher and Forbes present two broadly different approaches in having this debate, theoretical and empirical. The theoretical discourse mainly revolves around the moral or ethical choices of physicists and is hugely stance dependent. They present as an example the debate between Ernst Mach and Max Planck from the turn of the 20th century, which (in summary) reduces down to both sides accusing the other of finding validation for their assumptions within the biases toward realism they both already hold (Boucher & Forbes, 2024, pp. 17-18). In many ways, it is an example that demonstrates that stances are deeply relevant to theoretical discussion of both realism and anti-realism, and that the two opposed philosophies will never be able to convince each other on the merits of their logic, when their values are so fundamentally different. This style of argument is being dispensed with in this single paragraph. We will not be engaging with it going forward, as we have already chosen a more directly pragmatic approach in which this argument does not hold much sway, either for or against any points made.

The other approach is the empirical approach. This form of argument begins by taking some chosen success condition for science, i.e. 'more progress achieved over the same time period' or 'better predictive powers in theories generated' and analyses the philosophical underpinnings of the scientists within that field, producing some qualitative metric by which we can say if either realism or anti-realism is the better field. Note that, whilst we have defined some form of success condition for science, we haven't been methodologically prescriptive in and of itself. *How* the success is achieved is irrelevant to this question *if* the success *has* been achieved. Boucher and Forbes introduce the work of Robin Hendry, who using this style of analysis argued that in some contexts realists perform better than anti-realists in some specific types of scientific research, whereas in others it is the anti-realists who are more productive (Boucher & Forbes, 2024, p. 19).

This was extended by (Forbes, 2017) into a methodological analysis that was designed to help both historians of science and current scientists better identify if 'working scientists' (words theirs) should adopt either realist or anti-realist beliefs in different research contexts. Forbes directly looked at 19th century European work in electrodynamics, but this research concept hasn't been extended into other, more current, research contexts since said 2017 paper (Boucher & Forbes, 2024, p. 20). In the 2017 paper, Forbes makes the case that whether one should be a realist or anti-realist for pragmatic reasons depends on what work is needed to be done and the context in which its offered. In the case of empirical stance, he suggests that it may be more useful in the circumstances of "when attempting to produce and explore novel

phenomena in the laboratory" (Forbes, 2017, p. 3344), which seems to me to be close to that purest conception of physics I adopted in the introduction, and therefore that which best covers as many areas of physics as is possible. He also encourages us to undertake similar projects and determine "which epistemic stance is most appropriate for science policy makers, historians of science, or science educators, in a variety of contexts" (Forbes, 2017, pp. 3344-3345)

One of the goals of this thesis it to approach the question presented here by Boucher and Forbes in (Boucher & Forbes, 2024) (Forbes, 2017), and make the case that in the context of work surrounding entropy it is better for the productivity of physicists to adopt anti-realist stances, and beyond that adopt philosophies of science such as constructive empiricism. Showing that this approach, one of pragmatically accepting anti-realist stances in order to do better physics, can be workable philosophically and necessary from the perspective of modern physics. The productivity of modern physics is limited and restricted by paradoxes that only exist, from mine and thesis perspective, because of the realist stances that have become unspoken dogmas within modern physics practice.

This, however, is a grand project. The next section of this Chapter is dedicated to introducing the underlying philosophy of science that underpins *how* we can help dispose of the paradoxes that infect areas of physics, in the case of this thesis specifically entropy paradoxes. constructive empiricism, first developed by Bas van Fraassen allows us to introduce and develop a larger framework in which to apply our anti-realist stances that also isn't entirely dismissive of realist concerns. Van Fraassen does this via removing any and all metaphysical content from the application of science, simply asking "do our theories match the things we can perceive about the world" and making no claims, pro or anti, about whether this physics can be used to generate metaphysical claims about the world. This doesn't interest van Fraassen, whose central motivation for developing constructive empiricism comes from a desire to match the methodologies of practical science to a similarly practical school of philosophy, rather than make decisive metaphysical claims about what science can achieve as a tool for the metaphysician.

2.3.2 Constructive Empiricism

Our search for a philosophy of science that both matches the day-to-day practices of our working realist whilst also allowing them to remove the metaphysical concerns they may have should start and end with constructive empiricism. This philosophy of science, first introduced by Bas van Fraassen in 1980, is classified as an "anti-realist" school, but unlike other anti-

realist schools it attempts to construct a philosophy of science that first and foremost models the actions of scientists whilst being entirely agnostic on metaphysical questions. This approach leaves it in the position of not *denying* the importance of those metaphysical questions in and of themselves, but stating that those questions are not relevant to, or answerable by, science.

Simply stated, for the constructive empiricist, for a theory to be accepted by scientists it must be empirically adequate, which means correctly accounting for all observables. Van Fraassen formally gives it as:

> *"Science aims to give us theories which are empirically adequate; and acceptance of a theory involves as belief only that it is empirically adequate"* (van Fraassen, 1980, p. 12)

This structure seems naively immediately acceptable: if we wish to accept a theory, X, X must correctly predict what we observe with our own eyes given the conditions described in the theory. As an example, take GR itself. As a law about gravitation[29], taken to its Newtonian boundary conditions, it adequately predicts that apples fall from trees, that the Earth orbits the sun as it does, and that a cricket ball goes from one end of a wicket to another (accepting other theories of fluid dynamics also play a role). Under such a basis, it is an *empirically adequate* theory.

Empirical adequacy, however, does not involve, require, or even desire our scientific theories being true. At no point is the constructive empiricist concerned with whether general relativity is "true" or not. Such questions are fundamentally irrelevant to the constructive empiricist who, coming from the long empiricist traditions, doesn't think that such questions are answerable by merely science alone. The progression of theory development, at its core, goes from replacing one theory with another as more known observables are added to the pool, with each accepted theory being empirically adequate at any given time. By viewing science this way, we can fully understand and model the practice and history of science whilst avoiding commitment to any metaphysical constraints.

The interesting aspect of constructive empiricism, and the main ways it is attacked by realists, comes from how van Fraassen and the constructive empiricist understand and use the word "observable". For the constructive empiricist, an observable object is one that can be seen entirely veridically by the observer without the need for third party tools or interpretation.

[29] Again, strictly GR is not a "law about gravitation" (though it functions as such within certain context), it is a description of the topology and geometry of space-time *via which* we perceive things like gravity. I again defend my usage of such terminology on an explanatory basis.

Bluntly, for the constructive empiricist the observables about which we must be empirically adequate are only the ones that can be seen via our pure senses alone, i.e. the things we can see with our eyes. This means that, for the constructive empiricist, our theories don't have to adequately describe the workings of the microscope world, moreover they *can't* adequately describe the microscopic world, and that for the purposes of science entities such as "electrons" merely serve as convenient abstract tools via which we can more adequately describe the macroscopic world.

As an example of how this works in practice, consider again the theory of general relativity. For the constructive empiricist, it succeeds as being more likely to be empirically adequate than Newtonian mechanics not due to any descriptive power general relativity has in questions of microscopic entities, but due to the benefits it provides the things we see around us. The orbit of Mercury is not Newtonian, and in fact this was a great problem for Newton and others applying his theories. We now know that the orbit of Mercury is affected by relativistic effects, perfectly described via general relativity, meaning it is empirical adequate in a way Newtonian physics no longer is.

Another key question for the constructive empiricist is how we use, understand, and consider the testimony of others when it comes to thinking about what is observable. Science is a collaborative exercise, so the constructive empiricist needs a rigorous basis for establishing what we as individual actors can trust from other, third-party observers within our community. To this end, van Fraassen introduces the concept of our "epistemic community", which he judges to be a grouping of thinking actors that all share, fundamentally, the same epistemic potential, i.e. where in aggregate all members of the community can make the same quality of observations. This allows the constructive empiricist the ability to not have to make every single observation themselves and gives a pooling of testimony that the scientist who believes in constructive empiricism can fall back on. Van Fraassen goes on to explain that for us, right now, this community is effectively the entirety of what we would describe as humanity. (van Fraassen, 2005, pp. 114-115).

Further to this, van Fraassen's understanding of what an epistemic community is allows for an easier response to a common rebuke posed about van Fraassen's view of observability: what about those who require glasses or other aids to improve their epistemic skills? Are these individuals' testimony not to be trusted, and can they perform science in the way van Fraassen would require them to? The answer to this is, of course, yes, they can. By relativising the concept of what is observable to our epistemic community and understanding that community

by its *potential* epistemic skills, we allow for differing epistemic skills within individuals whilst setting a hard boundary condition on what is possible. There is only so small an object that even the best seeing actor can make out; what is observable ranges from those individuals through to the fully blind, all contained within one single epistemic community. Another consequence of this understanding of what our epistemic community is that we can only define our epistemic community via an understanding and application of scientific theory. For example, when we set our hard limit on the smallest 'thing' we can consider macroscopic, we inevitably must refer to theories on wavelengths of light that the cones and rods in our eyes can register, which is in and of itself a scientific theory. Observability cannot be derived *a priori*, and thus there can be concerns about the circularity of this definition. In Section 2.3.3 we will return to this issue, along with a greater analysis of suggested failures of constructive empiricism.

2.3.3 Problems with Constructive Empiricism

Constructive empiricism isn't a school without criticism. There are concerns from realists about nearly every aspect of its construction, from its view on what observation is to a suggested inherent circularity. I maintain that the main criticisms of constructive empiricism fall into three camps:

1) Philosophers misreading or misunderstanding the metaphysical scope of van Fraassen's work and intuiting understandings about what a constructive empiricist should be from mistaken realist beliefs. These problems tend to be intellectual in nature and are fundamentally addressed by van Fraassen in the essay collection *Images of Science (1985)*.
2) Problems focusing on van Fraassen's, and constructive empiricism's, accidental fall into realist understandings of the world, meaning a constructive empiricist by virtue of their own beliefs must accept some aspects of realist dogma. A good example of this is James Ladyman's (2000) work, *What's Really Wrong With constructive empiricism? Van Fraassen and the Metaphysics of Modality* (Ladyman, 2000).
3) Questions about the inherently circular nature of the constructive empiricist approach, up to and including whether the entire programme is fundamentally based on circular reasoning. These questions have no direct answers, but an appeal to the fact that there may be no way, or need, to resolve this within the framework of what the constructive empiricist is trying to do.

This section will address a concern from each of these sub-categories, showing how van Fraassen resolves them, strengthening the case for constructive empiricism.

To address concerns originating from category 1), a sensible place to start is questions arising from Churchland's (Churchland, 1985) critique of the what the limits of observability are for the constructive empiricist. For the constructive empiricist, spatiotemporal location does not fundamentally affect our ability to observe something, i.e. our theories still have to be empirically adequate about things that we cannot currently see but could if we were close to them. A good example of such an object is a supermassive star many thousands of light years away: we could not, even on the clearest of nights, see such a star with only our eyes, but if we were to get into a spaceship and travel toward it, at a certain new spatiotemporal location it would in fact be visible and present to eyes in exactly the same way as our sun is from any arbitrary point in the solar system. This is because the supermassive star is fundamentally macroscopic, a property it possesses in contrast to microscopic entities such as bacteria or viruses, which can never be seen purely with our eyes alone[30]. For Churchland, this distinction is only a contingent fact based upon the size human beings happen to be. For example, it would be entirely within our current best understanding of optics to say "humans can resolve images up to X resolution because of the construction of our eyes. If this were to change, say by humans being physically smaller, we could see more detail and have a greater pool of observables. Therefore, our understanding of what it is to be 'observable' cannot be based in these mere contingent facts about human existence". In Churchland's own words, the difference between things that haven't been observed yet but are observable and things that are called outright "unobservable" is "only very feebly principled and is wholly inadequate to bear the great weight that van Fraassen puts on it" (Churchland, 1985, p. 40).

Van Fraassen's response to this is to be astonished about how much Churchland can miss the point. Yes, the fact humans are the size, shape, and construction they are is a mere contingent fact, but our entire understanding of the world around us is constructed from those contingent facts. We can only understand the world, and our perceptions of it, through an understanding of what it is to be human, have the eyes we have that function in the way they do. For van Fraassen "[scientific] realists tend to feel baffled by the idea that our opinion about the limits of perception should play a role in arriving at our epistemic attitudes toward science" (van

[30] This is also a way in which science progresses for the constructive empiricist. If, to accept a theory, we agree it must be empirically adequate about *all* observables, discovering new observables out in the universe we could not access previously that do not align with our theory removes that pre-existing empirical adequacy, meaning we need a new, better, more empirically adequate theory. If the constructive empiricist believed empirical adequacy only covered *known* observables, we lose A) motivation to search for new observables ('the theory works as good as it must why search for more') or B) a simplicity about what it is to be observable ('macroscopic = observable, microscopic = non-observable').

Fraassen, 1985, p. 258). Churchland's point misses the idea that the observable/unobservable distinction is not a question of metaphysics, but merely a question of the epistemic attitude philosophers, and physicists, should take when approaching the world. Knowledge of the world around us is the collected sum of our experiences and abilities, metaphysical questions never enter the picture.

This dialogue, over the course of the essay collection *Images of Science* (Churchland & Hooker, 1985), gets to the root of several critiques of constructive empiricism. Rather than addressing van Fraassen on the level he is talking, many people who are either scientific realists, or lean toward scientific realism, start approaching constructive empiricism as if it were a metaphysical thesis, when van Fraassen clearly and definitively does not intend it to be. The observable/unobservable distinction is only troublesome if one begins to attach questions about the "reality" of the entities to the question of whether they're unobservable or observable. For van Fraassen, and the constructive empiricist, the only difference that manifests between an unobservable or observable object is whether we need to be empirical adequate about them. Anything beyond that simply isn't required for the constructive empiricist position to be successful, and any critique of constructive empiricism that goes beyond that simple position is not going to tarnish the success of constructive empiricism.

There are many critiques similar to Churchland's by eminent philosophers like Teller (Teller, 2001), Alspector-Kelly (Alspector-Kelly, 2004) and Hacking (Hacking, 1985). This thesis won't address them in detail here because they all run into this exact same issue: people applying metaphysical concepts and intuitions to places that they don't apply. However, there is another form of criticism that is worthy of discussion, such as the concerns that broadly fall under category 2): that constructive empiricism actually *does* contain metaphysical beliefs about the world, despite van Fraassen's position to the contrary. Our example of this comes from James Ladyman's work, *What's Really Wrong With constructive empiricism? Van Fraassen and the Metaphysics of Modality* (Ladyman, 2000).

Ladyman was concerned with the end result of a constructive empiricist believing in the statement "*x* is observable". One consistent way to think of this statement is to think in terms of the counterfactuals used to establish the observability of x within any given circumstance, such that "x is observable iff, under specific circumstances C, x could be observed" (Monton & Mohler, 2021, §. 3.5). Because many people understand and accept counterfactuals being understood through a "many possible worlds" framework (i.e., we judge the truth conditions of any counterfactual statement via an appeal to other, 'possible' worlds where the statement

could be variously true or false), it would be easy for the constructive empiricist to fall into some form of modal realism, and for many philosophers this would be the logical result of simply being a constructive empiricist. This is a direct and problematic challenge to the constructive empiricist, because their notion of observability cannot be compatible with having knowledge of other possible worlds: observability is built upon van Fraassen's epistemically modest empiricist worldview, however if via it can have knowledge of all possible worlds, we can no longer achieve the epistemic modesty van Fraassen requires (Ladyman, 2000, pp. 849-852).

In (Monton & van Fraassen, 2003)this question is directly attacked. Unlike the previous style of critique, van Fraassen doesn't attempt to deny the direct problem this viewpoint causes to the constructive empiricist viewpoint, and instead seeks to argue forcefully that we don't have to, and shouldn't, view counterfactuals within a "many possible worlds" paradigm. Instead, we should think of a counterfactual being judged against a contextual set of facts established at the creation of the counterfactual which we can already accept via our best empirical evidence. The statement "the keyboard in front of me is observable if it is on top of the desk and I am sat at my desk" is in and of itself entailed by facts about the world in front of me, the writer, right now. I have epistemic evidence that the desk is in front of me, it has a keyboard on it, and I can observe it with my eyes from my current location. If said keyboard were to be placed at my feet, outside of my direct eyeline, from my current location (circumstances) I would not be able to see it. I could change the context of my scenario, roll my chair back and, as if by magic, my keyboard would again be observable to me. Counterfactuals are entailed by the facts of the world around us, the observability of the keyboard (and the counterfactuals we can produce about its observability) are predicated on those facts, which can be disclosed by empirical research. For things we have not directly observed yet (but in theory can be observed), or where the empirical evidence is incomplete, we can only rely on our best theories to determine the facts of any given situation (Monton & van Fraassen, 2003).

This solution to Ladyman's problem is not without its faults. To start, it necessitates an inherent circularity within constructive empiricist thinking, where what we accept to be observable is dependent on our best theories, which we can only understand through an understanding of what is, or is not, observable. More discussion on this issue will take place later within this sub-section. Ladyman further extends this worry when considering solely "observable" objects that we have never yet observed. By wishing to consider them observable and adding them into our counterfactuals and theories as if they are, we have to lean on a belief that we can extend our best theories appropriately in contexts where we lack full empirical evidence. Ladyman thinks

that this is an impossible task. For him, the only way we could generate enough power to rise to a set of beliefs in the as-yet-not-observed observables is through a belief that the relationships that underpin our own acceptance of any given theory are in fact objective laws. Without such a belief system, which the constructive empiricist doesn't want to accept, the question of whether counterfactuals imply modal realist beliefs or whether we can fully contextualise them is irrelevant in the case of the as-yet-not-observed observables. Instead, we must commit ourselves to some *stronger* forms of realism, which is anathema to the constructive empiricist.

This follow up critique from Ladyman is hard to challenge directly. Much like questions of circularity, it devolves down to a set of value judgements any given philosopher can take regarding the varying strength of our empirical evidence required to justify claims about as-yet-not-observed observables, and whether any given evidence is *truly* empirical. The constructive empiricist should not be kept awake at night by such concerns, on the basis that we should take a more fine-grained approach to understanding observability and that this circularity is in and of itself unanswerable via any naturalistic approach.

We can also fallback to our work on stances here. When we understand the reason to become, for example, a constructive empiricist, is based not on a naturalistic analysis whereby we think the No Miracles Argument fails, but on an understanding in which we prioritize an empirical view of science over a metaphysical one, questions of circular logic become less important. We are choosing to be constructive empiricists because the constructive empiricist framework works better, is *more useful*, when it comes to the goal of providing good predictive power. How the constructive empiricist world view is grounded becomes less relevant than its ability to answer questions and produce good physics. If all our approaches, even realist ones, are implicitly value-based (as (Boucher & Forbes, 2024) indicates), then the realist has less justification for challenging the constructive empiricist on the grounds of circularity.

Furthermore, each individual object, either as part of a counterfactual or premise within a theory, can only be judged on the empirical merits defined by its context within science. If, purely as an example, we took black holes to be observable objects 15 years ago based on the empirical evidence we had *up to that point* acquired, it would have been reasonably orthodox position at least within context of astrophysics work. We had a well fleshed out set of theories, which helped describe and explain the motion of galaxies and massive bodies we could see with our own eyes from earth, despite having never a) photographed one or b) directly sensed the effects of one. If today we analysed the same statement, "black holes are observable objects", the position of accepting that statement has only been strengthened by the work of

LIGO, which detected their gravitational waves (LIGO Scientific Collaboration and Virgo Collaboration, 2016) and the EHT team, who photographed a black hole and its surrounding system (Event Horizon Telescope, 2022). If you had not believed black holes were observable objects 15 years ago, you may now, as a constructive empiricist, due to the changing level of empirical evidence. Yet, importantly, no scientist has *directly*, with their own eyes and just their own eyes, seen a black hole without resorting to huge telescopes. Ladyman argues that for the constructive empiricist, no level of empirical evidence up until seeing with their own eyes is enough to make them accept the observability of the as-yet-not-observed. I argue that constructive empiricists must make these value judgments, as part and parcel of being a constructive empiricist, and therefore Ladyman's concerns are effectively already "priced-in".

Ladyman's worries, and the arguments he makes, foreshadow the most challenging sort of problem the constructive empiricist faces, that of constructive empiricism's inherent circularity.

Let us start by again considering the process by which a constructive empiricist creates their own set of acceptable theories. They would start by laying out their understandings as to what is, and is not, observable. For the constructive empiricist, this is in and of itself not a theory laden notion. Indeed, it is meant to avoid as much theory as possible by making the definition as narrow and understandable as possible: if it can be seen merely with the eyes, it is observable. If not, it is unobservable. This is obviously the cause of much contention within both philosophical and scientific circles, as it goes against a naïve understanding of the progress of science. We have already addressed these concerns and won't go over them again here. For now, let us accept this definition of what observability is and consider it consequences down the line.

For the constructive empiricist, this understanding of what is observable is in and of itself a scientific theory: we create this theory as we do any other theory, and it is still based on a selection of scientific premises (we observed things via our eyes as photons hit our retinas, the signal from which goes to our brain, which is processed therein, etc.) This, however, is the theory that guides all others, and the theory by which we analyse each subsequent scientific theory for its own empirical adequacy. The question then arises: how do we analyse this theory of observability, its empirical success and failure? The answer, for the constructive empiricist, is unfortunately that we must use the theory to judge itself, leading to an epistemic circularity. The constructive empiricist must use their own theory of observability to judge the success of said theory.

This is another problem, and another one not easily solved. For the constructive empiricist, the best hope of providing a defence against charges of circularity may be no defence at all, but an acceptance of circularity as an unfortunate but unavoidable state of affairs. For the constructive empiricist, it could be argued that the search for some guaranteed grounding in the philosophy of science is functionally impossible, as the very subject and axioms even the most ardent anti-realist would accept it holds to are in and of themselves socially constructed. "Science" did not rise out of the universe like the physical world around us, it instead is the selection of thought processes we human beings use to analyse and interrogate the physical world. That it is circular could, indeed, be a necessary consequence of the human inability to be truly objective. We cannot remove ourselves from epistemic limitations that come with being human, even whilst attempting to analyse those limitations. The problem, therefore, is not one worth thinking about, due to our inability to resolve it[31]. Should that at any stage change, our understanding of observability would and, importantly, so would constructive empiricism as a whole.

To echo the start of this section, these three styles and formulations of critique demonstrate that constructive empiricism is not a philosophy of science without its own problems to resolve. However, some of these errors come from philosophers misapplying realist standards to a non-realist philosophy of science, some are defeasible by constructive empiricists themselves by expanding and explaining the position in greater detail. The issue of circularity is unavoidable, and we don't have easy counterarguments against it. The latter is the sort of issue that should trouble the constructive empiricist, and us, most keenly. Despite this, we can and should maintain a belief in the constructive empiricist position, if and only if it better explains the practice of scientists and provides us with *fewer* issues that its competing schools. On this count, the constructive empiricist merely has to reckon with the fact that we human beings, the beginning and endpoint (as far as we are currently aware) for thinking about the universe, cannot be super-observers with a full level of objectivity. This problem, present within the circularity critique above, affects all schools of philosophy to some extent or another, and therefore we can tentatively discount it and move on to the strengths of being a constructive empiricist for our realist-inclined physicist.

[31] Whilst this is true, in doing so we must maintain a strict policy to change any given theory should new empirical evidence comes to light, as we would with any other theory in constructive empiricism and indeed constructive empiricism itself.

2.4.2 Why be a Constructive Empiricist?

Constructive empiricism offers many attractive benefits to the realist-inclined physicist. As stated above, it allows for a philosophy of science solely focused on science's own internal, pragmatic success conditions. This distinction from the realist view is both important and subtle, in that the realist would themselves perhaps not recognise the distinction between what they are attempting to do, and what the constructive empiricist claims they are more successful in doing. The constructive empiricist claims that their approach, to analyse the structures, the aims, of science first and then acquire philosophical backing for their approach is far less likely to run into contradictions between what the philosopher of science *wishes* science could do versus what they actually do.

In order to address this claim, which would seem to neatly answer the worries of our working realist, we can approach the problem not from the work of philosophers, but from the practices and abilities of working physicists. In order to do this, let us analyse the "point of science" from the view of scientists themselves, and see which view for sciences success conditions greater matches the everyday approach of scientists themselves.

2.4.1 Truth or Adequacy?

The process of science is reasonably simply described: an observer will develop a theory about the operation of some system within the universe (usually on the back of prior experiences or observations), then formulate an experiment in which they can test the failure or success of their theories by further observation. In the early to middle history of science, these tests were usually somewhat conclusive, i.e. you could test the theories that gravity caused objects to fall at $9.8m/s^2$ toward the ground on earth simply by dropping an object from a pre-known height, measuring the time taken and dividing the distance fallen by the time taken squared. From this, you could derive the force gravity imparts on the object, and from that the gravitational constant. You could then test the reliability of your calculations via the usage of other objects with their own gravitational fields, and in effect reversing the testing process, starting with gravitational constant and verifying that the results you acquire match up again with the results of your initial experiments. This is how science progressed and developed models about how the universe functions.

In the modern world, testing tends to be less definitive and more akin to secondary observations, i.e. we use machinery and technology to "observe" (though not in the sense van Fraassen uses that term) how microscopic systems evolve under specific initial conditions. However, these practices taken to their macroscopic boundary conditions still operate similarly

to the experimentation of Newton and Galileo: Take a hypothesis, set up a real-world test to observe if your hypothesis is successful, and accept or deny your original hypothesis based off of that testing. This is the basic beginning point of why scientist do what they do.

A core aspect of this comes down to issues in the identification of one single "scientific method" via which we could judge the success or failure of science institutionally. This thesis will not make an explicit claim on one "scientific method" being better than any other suggested, though at this point it is worth introducing some of the discourse on the topic to demonstrate both a) the wide ranging nature of the discussion that can be different between scientists and b) the following need, which this thesis is demonstrating, for some lowest common denominator approach that allows us to talk coherently and cohesively about better and worse approaches when trying to discuss the interaction between the philosophy of science and the practice of science. Many philosophers of science and scientists themselves reject the concept, stated at the start of this section, that "the process of science is simply described". Wivagg & Allchin, in a 2002 article, make the case that seeing science as a simple process governed by one "Scientific Method" (*capitalisations theirs)* may be more useful on an explanatory basis, something especially useful within a classroom setting, but it re-enforces bad dogmas about the uniqueness and specialness of science, as if it, and its method, is the only path to finding truth (Wivagg & Allchin, 2002). This author and thesis sympathise with such concerns but sees simplicity as a method for *avoiding* dogmas as opposed to re-enforcing them. Properly construed, the simplest and most agreeable description of the aim of science and the simplest agreeable method for achieving that aim allows individuals with a wide-ranging set of dispositions or beliefs to agree a common starting point, from which progress towards a unified understanding can be reached. If we establish that science has a fluid, and necessarily vague, overarching 'method' based in empirical practice, we can identify historical beliefs about past science and use those to measure whether modern science is consistent with said past historical beliefs.

Richard Feynman, in the last of his 1964 Cornell Messenger Lectures entitled "Seeking New Laws", makes the case for simplicity. I quote him in full below:

"Now I'm going to discuss how we would look for a new law. In general, we look for a new law by the following process. First, we guess it… Then we compute the consequences of the guess, to see what, if this is right, if this law we guess is right, to see what it would imply and then we compare the computation results to nature or we say compare to experiment or experience, compare it directly with observations to see if it works.

If it disagrees with experiment, it's wrong. In that simple statement is the key to science. It doesn't make any difference how beautiful your guess is, it doesn't matter how smart you are who made the guess, or what his name is … If it disagrees with experiment, it's wrong. That's all there is to it." (Feynman, 1964)

For Feynman, the process of science is exceedingly simple, where hypotheses are subject to experimentation, and the process of science is to "guess" at how the world works and then test, experimentally, for it. This view of the scientific method, one driven by the work of scientists themselves, I believe allows for the greatest amount of implied wriggle room, despite Feynman's statements for simplicity. The process of guessing (testing, experimentation etc.) is left correctly vague, and only commits the scientist (and philosopher of science) to one concept: the belief in empirical evidence over rationalist thought, when the two contradict. This is the root distinction between a "science" and any other truth seeking and knowledge acquiring enterprise. Adopting a voluntarist framework about what constitutes the "scientific method" is a necessary consequence of already existent disagreement amongst scientists and philosophers.

If we take our understanding of the *aim* of science to come from this, we could formulate the statement that the aim of science is to "provide theories about the operation of the universe and systems within it and show these theories to be consistent with observation via the use of repeated independent testing of the theories". This description of the point of science is both limited and without any grander philosophical goals on purpose; the goal is to provide a statement on the aim of science without those goals to disconnect the work of an experimentalist physicist, our working realist, from as much philosophical thought as possible.

This point of science, from this understanding of the pure actions of scientists, now no longer references to the truth values of their theories. Being "right" or "wrong" about the structure of the universe is no longer relevant, as we can see that on this bare-bones formulation, all that matters is the consistency between the theories we can develop and observations we can make about the universe. Any question of "true" of "false" no becomes a secondary goal, one that we could add into the structure of science if we so chose but not one that is required for science to be successfully performed.

As an example of physicists in the real world operating on this basis, again consider the usage of Newtonian mechanics with the work of physicists. Within systems not approaching relativistic velocities, both Newtonian mechanics and general relativity both give equivalent answers, i.e. they are functionally the same, but with the benefit that Newtonian mechanics are

far simpler and faster to work with, both conceptually and mathematically. NASA still uses Newtonian mechanics to develop and design probes and rockets (NASA Jet Propulsion Laboratory, 2023) (as opposed to the more rigorous general relativity), as the time and resources saved by doing so outweighs any small, negligible inaccuracies. If the methodology of physics required the truth of the theories used as something irremovable from its practice, the physicists of NASA would be computing the launch trajectories of their rockets with general relativity, but because they are operating within an area where Newtonian mechanics and general relativity are empirically equivalent, they choose the easier option. The only situation where this would change is if those theories no longer functioned as empirical equals. Truth conditions, for scientists, are not required to perform good science.

This example also highlights the other selection of preferences physicists will use aside from truth when analysing the world via their theories. A physicist may choose one pathway to a solution over another if it is more "elegant", easier to work with mathematically or conceptually simpler than another, as long as both methods would produce the same empirical outcomes. The "truth", or at least a theory that would closely approach it, can be disregarded pragmatically for a whole host of reasons. The everyday actions of physicists, from those working to put rockets into space to those doing work in electrodynamics, where classical models are still used over the more rigorous quantum models, show that 'more correct' theories aren't in and of themselves preferable over a whole host of other, actor-driven preferences. The only fundamentally disqualifying quality a theory can have, making it useless for the physicist to entertain in any context, is if it doesn't match the observable reality around them. Aristotelean physics was superseded by Newtonian physics in a way that Newtonian physics wasn't by Einstein's general relativity on the basis that we can see, very simply with our own eyes, that a feather and a lump of lead will fall at the same speed through a vacuum (Galileo also predicted the very same thing, but did not produce a complete mathematical accounting of why this would be the case (Machamer & Miller, 2021)[32]). Newtonian physics being relevant in even the most cutting-edge physics practice is demonstrative that the average physicist doesn't operate entirely within the basis that a realist philosopher of science would want them to.

For these working physicists, they already operate as if they were constructive empiricists, which is van Fraassen's point. Rather than attempt to produce a philosophy of science that is built upon a philosopher's conception of the role of science, an understanding that will be

[32] Arguably his thought experiment is a conclusive a priori argument that they must fall at the same rate.

effected not only by the internal biases of a philosopher but one that could develop from an entirely distinctive way of asking questions such as “how does the universe work?”, van Fraassen wants to start from a blank slate. He asks “what do scientists do and what is their aim”, building the philosophy of science from that point onward. When you start the analysis from this origin point van Fraassen’s positions come conceptually easily, even if they do not give a great amount of hope for those who wish to buttress their pre-existing metaphysical beliefs via the work of scientists.

For our working realist, they may not even has realised it at this stage, but they operate in their daily physics work as if they were constructive empiricists, the realist intuitions they hold not coming from what their work in physics tells them *per se* but from the greater scaffolding of interpretation that surround their work. The constructive empiricist viewpoint doesn’t require them abandoning these realist intuitions, just that they recognise that physics cannot be the pathway via which they defend them, as it doesn’t have the tools to do so. The constructive empiricist viewpoint is one that may feel uncomfortable initially to the physicist, with its denial of the observability of the microscopic and its focus on pragmatic theory acceptance over any greater metaphysical content, but it is the theory that both best matches the everyday approach of scientists and provides a framework for them to understand their work without the problems that realist viewpoints can cause. Physicists, and scientists more generally, should be constructive empiricists because in most cases they already are.

In this Chapter, we have shown that realist understandings of physics are not just orthodoxy within the philosophy of science, but also permeate the language and practice of physics even when not explicitly adhered to. This bias, unconscious or otherwise, I maintain negatively affects the ability of physicists to do good work, and further to that may cause situations to develop where physicists operate outside of their epistemic reach, forming one aspect of why and how paradoxes like the Page-time paradox develop. Further to this, we have introduced a better approach; that of constructive empiricism. I have shown, and will continue to show over the subsequent chapters, that constructive empiricism holds the best chance for our working physicist to maintain their current working approach to doing physics whilst giving them the ability to either ignore and more comfortably solve problems and paradoxes in physics, by allowing them to be placed outside of metaphysical concerns and purely within pragmatic needs.

The next two chapters of this thesis will focus on applying the framework of constructive empiricism in greater detail to two different areas of physics, showing how van Fraassen’s

approach can help solve longstanding disputes and debates by either allowing the physicist to definitively see one approach as being more empirically adequate even at the cost of a more observer dependant view of the world (in the case of entropy in statistical mechanics), or that the debate is in and of itself not answerable solely by physics, as humanity reaches the limit of its epistemic grasp (in the case of general relativity and black holes). Both of these chapters will use the understanding of constructive empiricism laid out in this chapter, with Chapter 4 expanding on van Fraassen's view what forms an epistemic community.

Chapter 3 – Entropy; Information and Inevitability of Constructive Empiricism

Entropy is a concept that is deceptively complex, despite some attempts to simply describe it. Given its centrality to countless areas of physics, mathematics, and philosophy, as well as its venerable history, it would be easy to assume that questions such as "what is entropy?" have been fundamentally settled. This, however, isn't the case. Instead, there are multiple different forms of entropy or entropy-like properties, from Boltzmann's $S = k_B log\ W$ through to Shannon entropy's $H(X) = -\sum_{i=1}^{n} P(x_i) log P(x_i)$. Both are mathematically related, and produce consistent results across their usage, but the underlying axiomatic statements they're based on provide for different interpretations as to the properties of entropy, which in turn can produce vastly different philosophical implications. In this chapter, I will demonstrate that information theoretic understandings of entropy are preferable over kinetic views, that this demonstrates physics has the propensity to be uncomfortably subjective at least regarding entropy. I will also so that that entropy paradoxes, like Maxwell's Demon, are nigh on impossible to solve definitively by solely physicists within a realist framework. As previewed in Chapter 2, constructive empiricism allows us a pathway out of potential upset or confusion and gives us a greater contextualisation for the role and abilities of physics as a subject.

This chapter first addresses the differences between the informatic and kinetic versions of entropy[33]. Whilst one view, the more realist kinetic view, has become the pedagogical standard position within physics, it suffers numerous drawbacks in its attempt to produce a mind-independent concept of entropy that leaves it open to being challenged on grounds of paradox creation. The informatic view has the opposite tendencies: it is much harder to simply explain with a realist framework but does provide a more consistent description of entropy across all possible use cases. These differences will be analysed within two different fields, that of colloid physics and polymer physics. These two areas of physics have been chosen to demonstrate that the problems of the kinetic view, and the strengths of the informatic view, have direct empirical consequence. After this, an analysis of how entropy is used within black hole dynamics will occur, in which this thesis will show that black hole physics is already explicitly informatic.

[33] The "informatic view" and "kinetic view" is terminology taken from (Cates & Manoharan, 2015). I take it to be aligned with "information theoretic" views discussed later in this chapter but have kept the terminology aligned with the sources I am using, so as not to misrepresent the concepts.

After we establish the conclusion that the informatic view is more consistent and predictive than the kinetic view, the next aspect of this chapter will be considering the philosophical conclusions of holding the informatic view and seeing what they mean for problems in philosophy that include entropy. This will mainly focus on an analysis of work done over the past 50 years to exorcise Maxwell's Demon. Here, we will rely on work from Earman and Norton that, in essence, concludes that all past research work is wasted and there is not only no solution, but no obvious pathway to one. Agreeing with them, I further argue that constructive empiricism allows us to simply avoid this issue entirely via empirical adequacy; until someone produces an engine that lowers entropy for no work, or we see some system in the universe that does so, the theoretical questions remain entirely theoretical and beyond the scope of what singularly physicists can achieve.

3.1.1 Entropy – Questions from Colloid Physics

Entropy has always been a difficult concept to neatly define. We typically think of it as a property that tells us something explicit about the system's metaphysical content (i.e., a system's entropy is a real, objective property that grants us information about the state of the system and its place within the universe). This derives from the way it is taught at school and undergraduate level, e.g., this system has this amount of entropy S at time t, work is then done to this system, and it now has entropy $S - x$ at time $t + 1$. However, rather than thinking about entropy as a property that can be objectively observed to be a true property of the universe, an observable reality, we can instead think about entropy as if it were something observer dependant. On this reading, rather than entropy being a measure of a system's "chaos", set against a baseline grounded in our universe, entropy is a measure of how much chaos *an observer* can perceive, or, said differently, the lack of information the observer has about the system. These two readings have little pragmatic difference in terms of the mathematics for the small, straightforward problems presented as examples in undergraduate textbooks. On the other hand, these two readings have very distinct metaphysical consequences for the philosopher of science. They raise questions about how we can use physics to probe deep philosophical questions about the interaction between empirical study and the fundamental underpinnings of the universe. These two views are usually split into two camps: the kinetic view of entropy and the informatic view of entropy.

A brief introduction to both schools necessarily involves some description of the underlying physics at play, as well as motivating reasons for the existence of both. The informatic and kinetic view both come from two key issues presented by applying statistical mechanical and quantum theory to the world of complex systems. This chapter first analyses these two

opposing understandings of entropy in the context of colloid physics. A colloid is a mixture of microscopically dispersed insoluble particles that is dispersed through a fluid. The difference between this and a solution is one of scale, e.g., in a saltwater solution the sodium chloride particles are surrounded individually by water molecules, whereas in something like milk (a colloidal liquid), rather than individual molecules being separated, we have large particles of fat suspended within a water-based solution. Importantly for us here, we lose the ability to be able to track the molecular interactions of these large globules, as they are made up of vast collections of individual molecules, but they are still microscopic and numerous within a large enough sample. The beginning of the issues that separates these two schools is present here, in that it makes most sense to follow statistical mechanical practice by treating each of the globules in our milk as indistinguishable, i.e., each of our colloid globules is identical to any other. However, unlike individual molecules, which are in fact identical to one another, these globules are made up of huge numbers of molecules, which in turn contain huge numbers of atoms. Our globules should, in fact, be distinguishable from one another, if we so choose (Cates & Manoharan, 2015) (Frenkel, 2014).

Physicists who hold the informatic view square this circle and save statistical mechanics for the case of colloids by making the case that we can leave our individual colloids "undistinguished" for the purpose of statistical mechanics, i.e., say we could distinguish them, but there is in reality no need. They do this by reasoning that the quantum effects necessary when discussing individual molecules fall away when discussing colloids, which are large enough to be treated as if they were classical, meaning any difference between our "undistinguished" and "indistinguishable" particles becomes moot; for the purposes of classical physics, they are the same. In the informatic view, $S = k_b lnW$, where W is the number of distant and equiprobable undistinguished microstates that we choose to treat as indistinguishable when defining our macrostate (Cates & Manoharan, 2015, p. 6538).

For the physicist holding the kinetic view, this conflation of 'undistinguished' particles with 'indistinguishable' ones is unacceptable, as much for philosophical reasons as practical mathematical ones. In order to maintain some form of realism about our theories on colloids, we cannot allow these "undistinguished" globules to remain effectively the same as "indistinguished[34]" ones as that undermines the ability of classical statistical mechanics to do

[34] "indistinguished" here, as opposed to "indistinguishable", is terminology taken directly from the literature on this topic (c.f. (Cates & Manoharan, 2015)), and is meant to render the concept of particles that are *currently* not distinguished from each other, as opposed to having the property of being "indistinguishable". You can also think of it as, in effect, another term describing qualitative identity but not numerical identity.

anything more than simply produce models, not accurately describe reality. In short, in order to maintain that the entropy of our system is objectively real, we must view it as a dynamical quantity, not some epistemic "informatic" one that comes with choices we make about our macrostate. For someone holding the kinetic view, $S = k_b lnW$, where W is the measured volume of a phase-space that they system can possibly explore over time. This view holds weight as it tries to draw a deeper connection between the expected thermodynamical properties of the system and the "reality" demonstrated by our physics, as opposed to these properties being somehow dependant on the state of an observer, which would lead to a more subjective understanding of our universe.

These two views, for the philosophically minded, can be placed into two camps: the more "informatic" view (holding an empiricist stance) and the "kinetic" view (holding a metaphysical stance). This difference of opinion with physics is reminiscent of some of the arguments presented in Chapter 2, section 2.2.6, where Boucher and Forbes ask whether physicists should be anti-realists or realists. On one side, you have the epistemically grounded holders of the informatic view, who believe and maintain that they are pragmatically justified in considering colloids "indistinguishable", whilst metaphysically inclined holders of the kinetic view see that as an abrogation of the point of doing physics. In this section, we will demonstrate that the informatic view is fundamentally superior to the kinetic view when discussing classical statistical mechanics; that quantum derivations are not required to maintain indistinguishability between particles within the size range of classical physics; and from that, that information theoretic views of entropy are superior as they provide consistency across more areas of physics. First, we will discuss colloid physics, following that with polymer physics, before finally analysing the way entropy is presented in a third area of physics, black hole dynamics. Further to that, we will also demonstrate that how black hole dynamics conceptualises entropy already presupposes an informatic understanding.

The informatic view mainly has the strength of producing consistent mathematical descriptions of experimentally demonstrable physics in areas the kinetic view either can't or doesn't without introductions that basically readmit the informatic view. A good example of this is found by discussing the crystallisation process for colloid crystals. Taking a monodisperse (a colloid system with all particles within it of a uniform size) set of hard colloid spheres within a suspension at a volume fraction of $\phi = 56\%$, we would observe the state not only crystallising, but doing so rapidly and unambiguously. Equally, as the colloid spheres can swap positions easily in the liquid form but are unable to in the crystallised form, for indistinguishable particles (remembering that with our "informatic view", we can treat our "undistinguished" colloid

sphere as “indistinguished”) the entropy gain experienced as the colloid transitions from liquid to crystal is large. This makes sense, as in the crystal the individual spheres have more room to ‘wobble’ around, as their locations are set to a mean position on a lattice (Cates & Manoharan, 2015, pp. 6541-6542)

With the kinetic view, we cannot treat each of these colloid particles as indistinguishable. This means, rather than just tracking the particle swaps of individual crystals and their degrees of freedom as a measure of our entropy, we also have to track each possible permutations of each potential particle swap on our lattice. This additional entropy cost can be measured as $S_{perm} = k_b \ln(N!)$, where $N!$ Is the number of possible permutations of each colloid particle. This entropy cost outweighs the entropy gained transition for any large N as it is supra-extensive ($k_b N ln N$), meaning that our system will never crystalise. As we can account for a multitude of colloid crystals being existent, something must be wrong with the kinetic view at least mathematically, and certainly empirically. (Cates & Manoharan, 2015).

3.1.2 The Gibbs Paradox

Another way of addressing concerns with the kinetic view, or any view that attempts to address colloidal microscopic particles as distinguishable, can be shown by via the Gibbs paradox. The strength of the classical resolution to the Gibbs paradox is that it demonstrates that we don’t require quantum thinking in order to defend treating these particles as indistinguished; in fact, all that’s required is an understanding that the entropy of any given system depends on the choices an observer of said system makes with regard to what macrostates we are tracking.

The Gibbs paradox concerns the fact that, for distinguishable particles, we can derive a problem where the entropy is not extensive for the system, leading to a situation where the entropy of an entirely classical thermodynamic system decreases with no work done on the system (Jaynes, 1992, p. 3). The setup for the problem proceeds thus: Take two sets of coloured particles, yellow and blue, such that they are easily distinguishable between each set (yellow looks nothing like blue, vice versa) but such that the particles in each set are indistinguishable (all yellows are identical). The question then comes: what happens if all the particles merge smoothly and continuously into one single set of a green particles? The entropy for a set of unmerged particles within one system, i.e., the two sets of yellow and blue ones stated above, can be written as:

$$S_{IG}(N,V) = N ln(\frac{V}{N\lambda^2}), \quad (5)$$

$$S_{2Phase} = S_{IG}(N_a,V) + S_{IG}(N_b,V) \quad (6)$$

which is evidently different to the expression for the unmerged case, which is:

$$S_{1Phase} = (N_a + N_b, V). \quad (7)$$

The question, as Gibbs saw it, was at what point during the merging process do we move from using expression (5) to describe the entropy to expression (6)? This is the Gibbs paradox, finding the moment where the entropy jumps from being expressed by (5) to (6). One sensible solution is offered by considering quantum effects, and declaring that this merging can never take place, i.e. no possible transmutation can occur in the first place, identifying (5) as being the only possible description of the system's entropy that can be correctly identified with the original setup of the thought experiment (Cates & Manoharan, 2015, p. 6541). Gibbs, however, describes another path. Under this understanding, the entropy moves from one description to the other at the moment the person measuring it decides that the states are merged, i.e., when an observer makes decision to consider the entire set of particles indistinguishable. This resolution has the benefit of being more directly applicable in the case of classically determinable molecules, such as colloids, whose difference molecule to molecule could even be considered as being a spectrum of colours, exactly the same as our original setup. This points to the fact that effective indistinguishability of particles isn't just a consideration within quantum or semi-classical understandings of entropy, but basic to classical thermodynamic approaches. It seems, at least for colloid physics, whatever your prior choices, the epistemic position and decisions of the observer will matter hugely when it comes to entropy calculation, further strengthening the case of the informatic view, i.e., that the entropy belongs to the model and not the world.

3.1.3 The Subjective/Objective Distinction

At this stage, it would be sensible to ask why the kinetic view has any defenders. It necessarily requires complex approximates to resolve mathematical differences one runs into with its different interpretations of statistical mechanical entropy, as well as the large and seemingly intractable issue that it doesn't agree with observational testimony. The reasoning for this is less due to issues within physics itself, and more a question of individual priorities different physicists hold with regard to the "truth", "reality" or "objectivity" of their area of research. The kinetic view has a striking benefit for any strong realist: it directly connects an area of physical theory making, in this case work in statistical mechanics, to a foundational underpinning of the world and allows us access to it. Within the kinetic view, when we talk about the entropy of a colloid we are discussing some real, external property of the universe that we have the ability, though physics, to access.

The informatic view is more circumspect on any question of realism. Rather than taking that bold line, it maintains that entropy, as a measurable property of a system, is at least somewhat dependant on the choices of a person measuring it. The individual choice of what macrostate to measure and how fine-grained any approach should be added into the process, with these decisions having no wrong or right answer, allowing for situations that can have coherent but different measurements of a system's entropy at the same time. This situation is antithetical to the strong realist position on entropy. This problem also opens a small, but powerful, door. In having to accept that entropy is mind-dependant, we open the door to questions about the rest of physics, especially parts that are in and of themselves entropy-dependant. For a strong realist, this is an even more unacceptable position; for any anti-realist, this is simply a fact about the limits of science.

This distinction is evident, and in most defences or critiques of either the informatic or kinetic view directly addressed. A less considered aspect, however, is that the majority of physicists making a claim for or against allowing any level of subjectivity into work within physics don't fully grasp the consequences of the distinction, usually via sticking dogmatically and incorrectly to naïve realist conceptions about their work even at the cost of internal coherence.

In the next two sections, we will be addressing work within both a different field of statistical mechanics, polymer physics, as well as within black hole dynamics, in order to show two other areas where dogmatic approaches to a realist conception of entropy create problems that never need to exist, and again show the limits of realism. Once we have established these as existent, we will address how our new, improved view on entropy correlates to philosophical thought, doing so in order to reinforce correct accountings of the universe within the philosophy of science.

3.1.4 Polymer Physics – A Thought Experiment[35]

Another example of 'subjective' entropy naturally arising out of basic questions within statistical mechanics can be found in polymer physics. A polymer is a very large molecule composed of repeated subunits, examples including synthetic plastics (such as polystyrene or Nylon) to biological polymers, such as protein chains or DNA (Encyclopedia Britannica, 2021). Polymer physics is the study of the dynamics and mechanics of these substances. As these molecules are very large and made up of a large number of subunits, a statistical approach is

[35] This thought experiment was developed by both me and Prof. Tom McLeish over meetings in 2021. As such, I have no referenceable material for this section directly, beyond my own notes taken. For descriptions of polymer, I can recommend online resources from Yale (Yale University, 2024) or the University of Cincinnati (University of Cincinnati, 1998)

often required, leading to similar approaches as found in other areas of statistical physics when it comes to understanding properties such as entropy of polymers. As can therefore be expected, many of the same concepts utilised in sections 3.1.1 apply equally well within polymer physics.

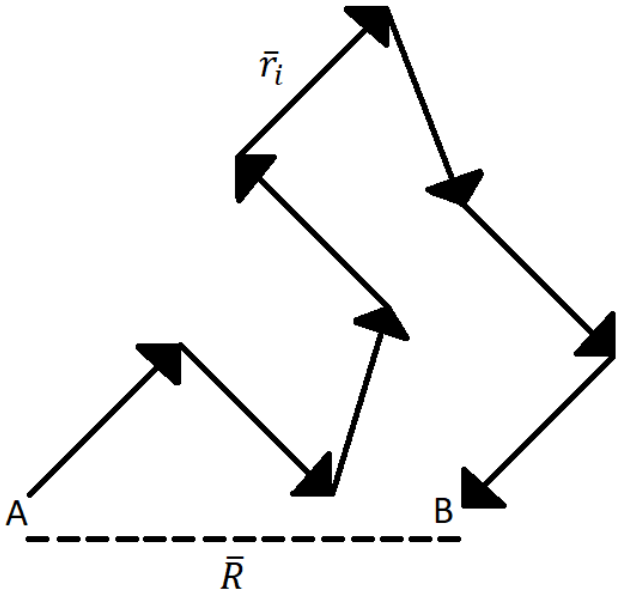

Figure 4 – The random walk of a polymer, of total displacement $\vec{R}$ made up of individual vectors $\vec{r}_i$.

The common way to model the dynamics of a polymer, and how a polymer will evolve kinetically over time, comes via the "random walk". If we image a polymer chain, being made of individual polymers, starting at point A and eventually arriving at point B, each individual section of that walk $\bar{r}_1, \bar{r}_2, \bar{r}_3, \dots, \bar{r}_i$ will "randomly" follow a path from A to B, with the vector $\bar{R}$ being used to describe overall direction and size of the polymer. This is shown in Figure 1. $\bar{R}$ here is distributed according to a Gaussian distribution, in three dimensions, meaning that we eventually arrive at $\langle \bar{R}^2 \rangle = Nb^2$, where N is the number of monomers within the polymer, and b is the length of each of these individual monomers. The end to end (point A to point B) vector of the chain is distributed according to the probability density function:

$$P(\bar{R}) = \left(\frac{3}{2\pi Nb^2}\right)^{\frac{3}{2}} e^{-\frac{3\bar{R}^2}{2Nb^2}} \tag{8}$$

Also written as

$$P(\bar{R}) = \mathcal{N} e^{-\frac{3\bar{R}^2}{2Nb^2}} \tag{9}$$

Given that each of the pathways our $\bar{r}_i$'s can take to get to $\bar{R}$ is modelled as a probability distribution, we can equally say that our $P(R) \propto \Omega(R)$, which simply counts the numbers of possible microstates within the system, we can turn Eqn. (8) into an entropy measuring the entropy within the polymer, which give us:

$$S = k_b \ln \Omega = k_b \ln P(R) \tag{10}$$

Which in turn give us:

$$S = k_b \ln \Omega - \frac{3k_b R^2}{2Nb^2} \tag{11}$$

Eqn. (10) is our equation for the entropy of a polymer.

3.1.5 Blobs – Problems of Graining

This result for the entropy of our polymer, shown in Eqn. (10), gives us something tangible to consider when talking about the entropy of our polymer. However, we are still far from a truly "objective" measure of entropy, something that can be highlighted by considering our polymer in "blobs" rather than individual steps r_i of length b. What then, are these "blobs"? Polymers can be viewed as a set of blobs, α, each of which contain within them a random walk and the both the start and end of each one of our blobs being at set intervals along our polymer (with all the blobs being the same size, g). It is important here to note that we have not altered the underlying polymer in any way; instead, all we have done is split up our original R into α composite sections of length g, as is shown in Figure 2.

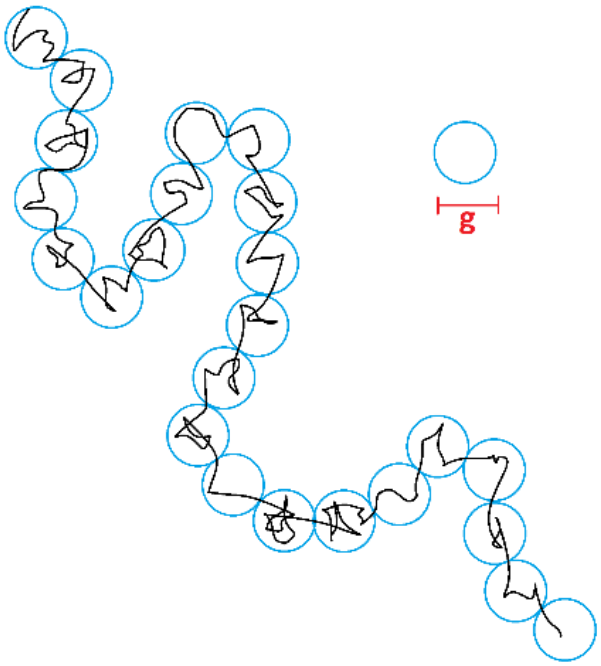

Figure 2 – A random walk of a polymer with blobs of size g dividing up the polymer. Each blob can be thought of as an individual polymer with length g replacing R, giving us a differently grained understanding of the composite polymer from the first example.

With this change, which is simply one of graining as opposed to anything independent that directly effects the polymer in question, we see a drop in entropy. How is this possible?

Consider setting the entropy from Eqn. (10) to the broad approximation:

$$S = \frac{-R^2}{N} \tag{12}$$

When we add our number of individual analysable sections, α, we see that $N \rightarrow \frac{N}{\alpha}$, $R \rightarrow \frac{R^2}{\alpha}$, $R \rightarrow R\sqrt{\alpha}$ and $r^2(n) \sim n$. When adding our new composite blobs together and substituting in our new relationships we get:

$$S_{blobs} = \alpha\left(\frac{-R^2/\alpha}{N/\alpha}\right) \tag{13}$$

Which, when cancelling, includes a $1/\alpha$ dividing factor to our original entropy. From this we can see that the amount of entropy seems to have dropped purely via choosing a different level of graining by which we analyse the polymer.

This is a very important result. A way to consider it conceptually could be to imagine our first set up as one long road going from Aberdeen to Berwick. Assuming this road follows a Gaussian random walk and the road follows doesn't affect its own pathing (i.e., it is a "self-avoiding walk"), it can take many possible routes from A to B, and we have no accessible information

about the route it actually takes, meaning the entropy of our system is high. Now imagine we chop up our road, and say it passes through Caithness, Dundee, and Edinburgh, giving us four road "blobs" going from Aberdeen – Caithness, Caithness – Dundee, Dundee – Edinburgh and Edinburgh – Berwick. Even still assuming that the path *within* each of these road blobs is a self-avoiding random walk, we now have information about the start and end points of four sections of the road, e.g., the A-C, C-D, D-E and E-B sections. This provides us with more information about the pathing of the road, cutting down the possible number of options our original random walk takes, meaning the entropy of our road drops by a factor of $^1/_4$ (4 being the α stated above). Bear in mind, our road need not have changed at all for this entropy drop to occur, it may have always passed through the towns and cities outlined in the road blobs setup. The only difference is the amount of information available to the observer and their decision to chop up the road into these specific blobs. In short, polymer physics demonstrates that the amount of information available to the observer and what macrostates an observer chooses to observe will affect the resultant observed entropy of a system, making the case that entropy is better represented by the informatic model as opposed to the kinetic approach.

We can further our understanding here by imagining stretching our polymer out, such that each blob has within it a random walk, but the blobs themselves align linearly, so that our blobs form a straight line.

In this instance, we do not have a polymer following a random walk in the manner described in the above two cases, instead having a random walk *within* the blobs but a straight line connecting them all, forming a stretched polymer, as shown in Figure 3. What is the entropy of this new stretched polymer?

In this setup, we recover $R \sim N$ as opposed to the $R \sim \frac{N}{\alpha}$ as the $\alpha's$ cancel out, producing an entropy equivalent to Eqn. (10) (when described fully with constants). Why does the entropy return to originally derived entropy for a randomly walking polymer, when there seems to be no random walk amongst the blobs? The answer lies in the fact that the location of the boundary between each blob now is built into the system, i.e., you can fully know the start/end of one blob directly from the first blob encountered (assuming all blobs are the same size and shape, which we do). Given this, we return to a situation where knowing what blob is where (as we do in the second example) doesn't provide us with any more information about the system than simply knowing where it starts and ends, returning the measured entropy to a higher figure. Using the road analogy, if we knew the road between Aberdeen and Berwick passed through towns that directly fall on the shortest possible path between them, knowing where it passes

through doesn't provide us with any more information than we already possessed by knowing it goes from Aberdeen to Berwick. Knowing the locations our road passes through doesn't provide us with any more knowledge about the road system than we already had, therefore providing us no change to the measured entropy of said road system than before we added in knowledge of its passing places. Again, we see that the connection between the mathematical description of entropy presupposing a connection to the information either added or taken-away by the chosen setup of the system, demonstrating again the observer-dependency of entropy calculations.

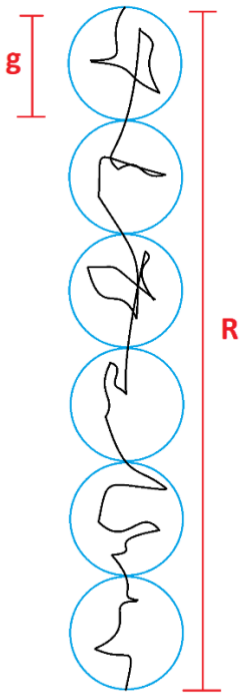

Figure 3 – In the stretched case, all our blobs are aligned such that they form a straight line, but the polymer inside these blobs still follows a random walk within the blobs. As we can see, $\alpha \cdot g = R$, meaning the information acquired via the blobs is no long relevant for the entropy calculation, returning us to the entropy found for the coarse-grained polymer example.

3.1.6 The Entropy of Black Holes

We have, by this point, shown the benefits of the informatic view, at least in the context of soft-matter physics and statistical mechanics. The more realist kinetic view, with its desire to place entropy firmly as *of* the universe, doesn't account for known, observable physics or the theoretical structures of statistical mechanics. If we place entropy as an important property within physics, i.e., one that cannot be explained away or modelled around, we must conclude that there are areas of physics the strong realist cannot account within a suitable ontological framework. To whit, if an observer's arbitrary choices can affect measurements of independent systems, entropy cannot be solely a measure of *something* in the universe, it must measure our observer-centric relationship to it. This should be deeply concerning for the realist. What is the approximate "truth" being pointed toward in our theories if the empirical results we can

measure can change *arbitrarily* depending on how course or fine grained an observer wishes to set up their experiment? Can we say objectively say *what* a system's entropy is sensibly? It would seem that a realist would like to be able to do so, and it especially seems like a physicist operates as if they can, mostly unconcerned with the potential philosophical consequences.

To this the realist may raise the point that we have only shown this in classical statistical mechanics. Potentially colloid and polymer physics have simply missed a deeper concept that can, again, allow us to sever the connection between an observer's arbitrary choices and measured entropy. In order to demonstrate the failure of this argument, we shall move away from core statistical mechanics into a discussion of black holes and black hole entropy, an area where thermodynamics, statistical mechanics, quantum theory and general relativity all meet. In doing so we can analyse what entropy is to someone who works in this field, and whether it is as subjective as the entropy of colloid and polymer physics (and, thusly, see if these conceptions of entropy are equivalent). We shall do this by discussing theories on black hole entropy, black holes as thermodynamic objects and the problem of black hole information paradoxes, where entropy plays a central part.

Let's return to the concept of a black hole, starting from first principles and re-outlining the thermodynamics of a black hole, much as we did in Chapter 1. A black hole is construed of an infinitely dense singularity, often caused by the collapse of a star much bigger than our own sun, where the gravitational force of the mass of said star is greater than the neutron degeneracy (limited by the upper bounds of the strong nuclear force), meaning the formation of an infinitely small, infinitely dense point in space-time. Outside of this, there is an event horizon, which is the point where the strength of the gravitational force generated by the singularity becomes too strong for even a particle moving at the speed of light to escape (Wald, 2010, pp. 299-300). For the purposes of this chapter, anything between the event horizon and the singularity will be described as the "black hole interior", and any region of space-time from the event horizon to infinity will be described as the "black hole exterior"[36].

It is also worth reminding ourselves at this stage that the black hole interior is not conceptual complex as a spatiotemporal region of the universe. Beyond an infinitesimal change in space-time topology from one side of the event horizon to the other, the black hole interior is theoretically no different to the black hole exterior. The Principle of Equivalence also enforces

[36] This understanding was also outlined in Chapter 1. We will also be using the concept of the "black hole interior" and "black hole exterior" in Chapter 4. When it returns, I will both state this definition again, as well as give the reasoning for it, to ensure clarity for the reader.

the important principle that no infalling observer should be able to tell that they have crossed the event horizon as they do so. For a realist, they must motivate and explain why this space-time region operates differently metaphysically for an external observer, because *prima facie* this region should operate metaphysically identically from a naturalist viewpoint. This will be important to bear in mind later into this chapter, as we make the case that the realist has questions to answer here that the constructive empiricist does not.

Entropy plays a central part to the paradoxes and complexities that we will discuss in this section, but first we must establish what the entropy of a black hole is, and again go over the grounding for why we can treat black holes as if they were thermodynamic objects. To treat the second point first, consider an object passing through the event horizon. As soon as this object enters the black hole interior, it becomes epistemically non-accessible to all those in the black hole exterior, as to acquire information (for example, seeing its current location in space-time) about the object would require photons moving at greater than light speed to break back past the event horizon and into the eyes of any observer, an impossibility. However, entropy associated with this object must still be present in the universe even if it is epistemically beyond our reach, to not break the Second Law of Thermodynamics. We therefore have an associated drop in the entropy of the wider universe (e.g., the black hole exterior) and a subsequent increase of the entropy in the black hole, demonstrating that for the Laws of Thermodynamics to hold, black holes must possess an entropy that is measurable. This in turn indicates that black holes themselves operate as thermodynamic objects (Wallace, 2018b, pp. 28-29).

As we have seen, the entropy of a black hole can therefore be given numerically. Bekenstein gives this entropy as being directly proportional to the surface area of the black hole, given by the size of the event horizon. One form of the equation is given in Chapter 1, but it can also be written:

$$S_{BH} = \frac{A}{4L_p^2} = \frac{c^3 A}{4G\hbar} \text{ (Bekenstein, 1973, p. 2338)} \quad (14)$$

where A is the black hole surface area, c is the speed of light, G is the gravitational constant, $\hbar$ is the Planck-Dirac constant and L_p is the Planck length. This entropy is dimensionless and only depends on the black hole's surface area A, which in turn only depends on its mass, meaning that a black hole's entropy is directly related to its size. Furthermore, we can deduce that the entropy of black hole must be very large, given the size of the numerator $c^3 A$ being considerably larger than the denominator $4G\hbar$. Applying this new entropy to Second Law and its

requirements, we must also develop a new Generalised Second Law of Thermodynamics, which states simply that the new entropy of the universe when our object passes into the black hole interior must either be the same or larger than the entropy of both the black hole exterior and interior. This can be written as:

$$\Delta S_O + \Delta S_{BH} \geq 0 \text{ (Bekenstein, 1973, p. 2339)} \quad (15)$$

Where ΔS_O and ΔS_{BH} are respectively the change in entropy of the wider universe and the black hole itself. With this, we have the framework for understanding the entropy of black holes that stays consistent with the Laws of Thermodynamics.

Another aspect of black hole entropy that must be understood in order to analyse whether its value is observer dependant or a real property of the structure of the universe is Hawking radiation. Hawking radiation is the process whereby particle creation at the event horizon causes a black hole, in effect, to "evaporate" away over time if placed in a closed system. It is caused by background quantum vacuum fluctuations producing particle creation, for example one proton and one anti-proton. If this occurs on or very close to the event horizon, one particle can in theory escape to infinity whilst the other is trapped within the black hole (dependant on their initial velocities). Because this radiation comes from a mixed quantum state, any outgoing radiation must be thermal, and therefore black holes are themselves a misnomer: they cannot be fully "black". In fact, they have a temperature, derived by Hawking as

$$k_b T_{Hawking} = \frac{\hbar}{2\pi\tau_\kappa} \text{ (Hawking, 1975, p. 199)} \quad (16)$$

Where τ_κ is the characteristic time of the particle, which is directly related to the mass of the black hole given by

$$\tau_\kappa = \frac{2r_s}{c} \quad (17)$$

With r_s being the Schwarzchild radius of the black hole. This black hole temperature is another key piece of evidence that black holes are in fact thermodynamic systems and plays an important role in the black hole paradoxes involving entropy that we shall discuss. Note here that the entropy of this radiation should increase monotonically as the black hole evaporates (due to the quantum entanglement implicit in particle creation) as the black hole radiates away.

3.1.7 The Page-Time Paradox – Entropy Causing Problems

What does all this mean for the question asked at the start of this chapter, which asked if which one of the informative view or kinetic view were superior conceptualisations of entropy? Should we take a view of entropy that is more epistemically driven, or a view that states measuring the

entropy of an object tells us something objective about the universe, removed from any subjective aspects? We have established in the previous section that black holes are thermodynamic objects in the same way that any other object within space-time has thermodynamic properties, and therefore black holes should be open to the same questions that we posed of colloids or polymers. Furthermore, in the case of black holes we have a region in space-time that is firmly split (for us here on Earth, with little chance of going to a black hole) between a region that is observable fully in principle (the black hole exterior) and a region forever hidden from us (the black hole interior). Black holes therein and offer a perfect case-study in which to ask questions about entropy, as we can clearly delineate epistemic concerns from metaphysical ones.

The first pathway toward answering our questions can be seen via the Page-time paradox. To give further context than is discussed in Chapter 1, the Page-time paradox is a continuation of previous work on the black hole information loss paradox developed by Hawking and others in the 1970s. In short, the information loss paradox questioned what happens to information bound up in the matter and energy that forms the black hole singularity. After Hawking radiation "evaporates" away all the mass content of the black hole, we should see a region of space-time containing no event horizon or singularity, but instead a number of charged particles within the area that once before contained the black hole. If we had two distinct initial states that collapsed to form said black hole, because all Hawking radiation is thermal and therefore of a mixed quantum state, we would be unable to distinguish which parts of the radiation came from which state, breaking unitarity. Solutions have been proposed to this problem, mostly invoking either a quantum encoding of information about the black hole interior within Hawking radiation, meaning Hawking radiation is not precisely thermal or that black holes in and of themselves have "soft-hair" particles that, over enough time, correct for the lost information.

Page, however, demonstrated that an information loss paradox can be seen considerably earlier than what is proposed in the original information loss paradox. This relies on the fact our original particle pair, the mechanism by which Hawking radiation occurs, must be entangled during particle creation. This, in turn, leads to increasing entropy of quantum entanglement as Hawking radiation continues to evaporate away the black hole. However, the maximum possible entropy value of this quantum entanglement has to be bound by the Bekenstein entropy of the black hole itself, shown in Eqn. (14). This leads us to a point, at around half-way through the lifespan of an evaporating black hole, where the macrocanonical (i.e., thermodynamic, Hawking-Bekenstein) entropy and von Neumann (i.e., quantum) entropy meet, with the latter then forever bounding the former. As our entropy of entanglement cannot

increase, we must see some large entanglement breaking effects occurring around this time, or else particles being entangled with multiple others, both temporally and spatially distinct (an impossibility, as this would break the monogamy of entanglement). This does not fit with the classical understandings of black hole information loss, meaning that if the "Page-Curve"[37] is followed, information must escape the black hole via some, as yet unknown, means (Wallace, 2017, pp. 12-14).

3.1.8 Confirmation of the Page-Curve – Wormholes and Emergent Physics

Can we therefore recover the Page curve via any physically reasonable model? The answer to this question is complex and currently up for debate, but there are some proposed solutions. The one we shall focus on was performed by Almheiri *et al.* in 2020 (Almheiri, et al., 2020a), where they demonstrated that after the Page-time is reached a so-called "entropy-island" forms within the black hole, bounded by a quantum extremal surface. This island is composed of radiation and particles which are no longer entangled with the outgoing radiation from before the Page-Time, and instead is a space both within the black hole but equally not part of it. As the black hole continues to radiate according to Hawking radiation, we see this surface become a larger and larger part of what was once the "black hole" (in effect the black hole is made smaller from the inside out). This in turn leads to modelling in which the Page curve is followed within an AdS/CFT (Anti-de-Sitter/Conformal Field Theory) universe structure, allowing the possibility that real black holes, when their entropy drops, provide information back into the universe (Almheiri, et al., 2020a, pp. 9-11).

Questions remain here. The mechanism for information leaving the black hole is still an unresolved question, as well more naively apparent questions such as "if the singularity is in the middle of the black hole, how can the entropy island contain that region?". To take the latter question first, the work of Almheiri *et al.* (Almheiri, et al., 2013) on black hole complementarity and firewalls provides a solution: the singularity, post the Page-Time, can be thought of as moving to the event horizon so as to prevent causality issues arising as to objects that fall into a black hole having spatial location. This new quantum extremal surface (the surface which bounds the entropy island) becomes, in effect the new singularity, or at the very least the location at which all the interesting high-energy quantum physics occurs. What this means for singularities this thesis won't address beyond noting the conceptual strangeness of the singularity being potentially macroscopic, it now occupying all of a region of space-time.

[37] The Page curveis the modelled rise and fall with entropy, maximising at the Page-Time. It can be seen in Figure 1, which comes from (Wallace, 2017).

The first question is more interesting. In order to apply the idealised version of events that highlighted the possible existence of these "entropy islands" to more conventional space-time setups, Almheiri *et al*. utilised both the gravitational path integral and the replica trick in order to consider the potential topologies of space-time play within a black hole approaching the Page-Time. They discovered that the saddle points of black holes that fit the model were placid topologies that could, in theory, connect spatial points in the universe in the manner of a wormhole. This provides us a topological structure for these potential entropy islands, but as yet no way to connect them directly to the Page-Curve.

In order to do that, we must calculate the entanglement (von Neumann) entropy of these saddle point regions, for which we require the density matrix of the entangled particles. This isn't accessible to us directly, but via the replica trick, we can recover the information required (Almheiri, et al., 2020a, pp. 12-14). This, "the replica trick", can be described as performing multiple simulations of "replicas" of an original set up and acquiring a composite result that describes the original setup. For example, take a coin. If you wanted to prove if it were fair, you could toss it an arbitrarily large number of times and see if there was a distribution of heads and tails that approaches 50/50. You could also take an arbitrarily large number of replica coins, flip them once, and perform the same analysis, showing you equally whether the original coin your replicas are copies of is fair. By running simulations of an arbitrarily large number of black holes, you can deduce the evolution of the entanglement entropy over time of an original black hole that your simulation is based off. Having acquired the density matrix in this fashion, we can discover what the evolution of the entanglement entropy is of these saddle-points during the lifetime of a black hole. By doing this, we find that as the Page curve is followed, the latter wormholes eventually better describe the dynamics of the black hole, the switch point coming at the Page-Time, as our entropy islands form. This provides us a pathway (through the wormholes) that information is preserved and re-enters the universe, and we see an associated drop in entropy as the Page-curve is followed. Whichever way you look at black hole entropy, the connection between it and information available to external observers seems profound.

3.1.9 Information, Entropy and Black Holes

Entropy related to black hole dynamics, on these considerations, seem to be connected to some denial or access to information either about the black hole at current time t or at some previous time $t - n$. On this initial reading, the informatic view is preferable, because the entropy of the black hole is directly related to our ability to observe matter that passes through to the black hole interior: something goes into a black hole, beyond our ability to view it anymore, and the entropy of the black hole rises. This is further strengthened by any

demonstration that the Page curve is an observed relationship within black hole dynamics. If black hole information being recovered is proportional to a drop in quantum entanglement entropy, we have a demonstration that both thermodynamic and quantum entropies are directly related to information being in principle accessible to an observer.

Further to this, what is described in section 3.1.8 is a theory in which the entropy of a black hole drops, proving the Page-curve is followed, solving the Page-time paradox (by introducing new physics), and recovering the information 'lost' to a black hole. Section 4.2.1 in the next chapter discusses the consequences of this solution to the Page-time paradox for the constructive empiricist who is seeking to resolve black hole paradoxes, but for the purposes of this section there is a different takeaway; the creation, and solution, to black hole information paradoxes are found via an understanding of entropy that is implicitly centred in an external observer's observation of the black hole system.

This point, whilst naïve, demonstrates again one of the core concepts of this chapter: that entropy is directly related to the information either available to the observer, or the choices an observer makes informationally about the system. In black hole dynamics, we can see this understanding presupposed not just in our solution to information loss paradoxes such as the Page-time paradox, but in the construction itself of black hole entropy. The Bekenstein entropy fundamentally summarises down to $S_{BH} \sim A$, where the larger a black hole's surface area is, the larger the entropy of the black hole. A black hole's surface area, the area bounded by the event horizon, can be easily conceptualised as the sphere within which we, external observers, can no longer see, but otherwise it is functionally normal, non-problematic space-time. The generalized Second Law given by Bekenstein in Eqn. (15), demonstrates explicitly that the entropy of the universe is connected unambiguously to how much space-time is bounded by the event horizon (i.e. the area we would call the "black hole interior"), outside the epistemic reach of an observer outside that interior (e.g. on Earth). This also speaks to the greater point this thesis makes throughout the piece: our ability to understand theories of physics that depend on entropy rely decisively on our epistemic abilities as humans. When our theories of physics go beyond that epistemic range (which, in Chapter 4, will be demonstrated), the empirical study of physics cannot be truth-providing, and if you choose to allow physics to do so you cannot easily maintain realism.

Overall, Black holes offer us another key reason to believe entropy is directly connected to "information", which can only be rationally understood as what an observer has access to.

"Information", being a directly epistemic quasi-substance[38], has no meaning without reference to some external observer or decision maker, and thus any property that is directly connected to "information" (like entropy) will also be directly epistemic. We can firmly conclude by this point that entropy is a property not *of* the world, but *in* it, dependent on the choices and individual makes about their setup, or the area of the world they have epistemic access to.

3.1.10 Philosophical Implications

The undercurrent throughout this Chapter so far has focused on two different understandings of entropy within statistical mechanics and an analysis of how entropy is used and functions in black hole dynamics. This is done intentionally, so as to set the groundwork within physics for understanding any greater philosophical consequences for both how physicists should view their own subject, but also to help philosophers understand the paradigms they are actually dealing with when discussing the point or aim of physics from a third-party perspective. Any philosophy of science here should not simply be internally consistent with the beliefs and thoughts of the philosopher creating it but align with the practical work done by physicists and the choices they make when "performing" physics.

To that end, let us return to our working physicist from Chapter 2. Upon seeing this discourse, they should prefer the informatic view over the kinetic view when the problem is considered holistically. Outside of philosophical concerns surrounding the subjectivity/objectivity, or how we foundationally consider entropy functioning as a property, the informatic view simply works more consistently, across a greater number of fields, with a greater mathematical consistency. We can see in Section 3.1.5 an example that the entropy of a system is dependent on the arbitrary choices made when setting up the system, i.e. directly connected to the amount of *information* we preassign any observer knowing about the system when creating our initial setup. That this changes the underpinning mathematics within our statistical mechanics analyses is demonstrative that the difference between the kinetic and informatic views isn't simply two different ways of viewing entropy. In fact, it shows there is a *better and worse* way of understanding entropy, at least from a purely empirical perspective. In Section 3.1.9, we can see the case made that the way black hole physics understands and uses entropy can only be correctly conceptualised and discussed cogently by connection of the entropy and information available to the observer. Again, we see another area of physics that deals with the property

[38] Any terminology used to describe what "information" is as some stand-alone thing will feel insufficient. We have a standard unit of information within computing (the bit) representing the logical arrangement of electronic charges, but this also isn't fundamental or appropriate in capturing the qualia of "information" required here. I have gone with quasi-substance as an unfortunate but shorthand way of conceptualising "information external to an observer" as some form of entity/property/thing.

"entropy" in a way that is only coherent by accepting, at least in broad terms, the informatic view.

To our working physicist, this is where they can stop the analysis, if they so choose. They have a superior way of understanding entropy, which works more consistently in producing empirically accurate results. No further thinking required. However, I think this demonstrates something fundamental and potentially scary to any realist philosopher of science that speaks to the first paragraph of this section, where I discuss the need for a good philosophy of science to align with the practices and "performance" of physics. If we take the informatic view to be better than the kinetic view, it necessarily suggests that arbitrary differences of information we have as observers and non-passive actors in the world affect the properties of external systems in a way that is non-objective. That we, the people attempting to analyse the world through our epistemic abilities and chains of reasoning in fact change what are, *prima facie*, properties entirely external to ourselves through the sheer act of experiment setup. This centrality of the observer over some abstract and external *the way the world is* implies that there is no universal, can be agreed upon, metaphysical single reality underpinning the universe. Instead, as we can see from our polymer physics example, we can have mutually inconsistent, but both entirely accurate, answers to a question "what is the entropy of X system?"[39]. This inconsistency is entirely derived from the information that two independent observers may have, the system is the same beyond our choices about what we choose to observe in the system.

At this point, our philosopher and curious working physicist may both share the same question: "what does this mean for physics as a truth generating enterprise?". I will attempt to show throughout the rest of this Chapter that physics cannot be truth generating in the way a realist would want, that physicists and philosophers have already wrestled with this entropy-centric problem in other areas of statistical mechanics, and that our best path is to follow van Fraassen's lead and either adopt, or recognise one has already functionally adopted, constructive empiricism.

[39] N.B., this mutual inconsistency is found if asking simply "what is the entropy of X system?". If we normalize it to any given observer, such as "what do I measure the entropy of the system as (given my circumstances)?", we can remove this inconsistency. However, asking questions via said normalization simply isn't the way these questions are formed in physics (which deals with properties like entropy naturally as of-the-world) or, as is maintained throughout this Chapter, doing so either forces you to accept observer relative positions that a realist would be uncomfortable with. Therefore, I am happy maintaining that this is a genuine mutual inconsistency iff you wish to be a realist.

3.2.1 Maxwell's Demon – Szilard's Engine

Szilard's Engine is a development on the concepts raised by Maxwell's demon and other statistical mechanical thought experiments that, at least seem, undermine the laws of thermodynamics. In this section I will introduce some of these thought experiments; the problems they cause for the laws of thermodynamics; proposed solutions and resolutions to these prima facie paradoxical systems; and demonstrate what they can tell a philosopher of science about the underlying concepts within this area of physics. Further to that, I will seek to show that informatic theory solutions, similar conceptually to the ideas proposed earlier in this chapter, turn us away from metaphysically committed forms of philosophy of science, and instead demonstrate that van Fraassen's understandings of science lead to the construction of better philosophy, namely constructive empiricism.

Maxwell's Demon is a well-known paradox within physics and lays the groundwork for Szilard's engine and future concepts of information-as-entropy, so for the benefit of the reader I will outline its core concepts here.

In 1967, James Clerk Maxwell, in a letter to Peter Guthrie Tait, first introduces the problem that will eventually be called "Maxwell's demon" (although, in the original letter, he doesn't use the word "demon"). In it, he asks us to imagine a box split into two sections, A and B, which are both in full thermodynamic equilibrium with each other. Into that split, add a frictionless shutter than can open and close at the will of an outside party (in later descriptions of this problem, a demon). Inside each of these sections are two gases made up of particles travelling at different velocities (some fast, some slow) but which, as a mean average over all the particles in each section, remain the same average velocity across either section (thus, remaining in thermal equilibrium). If the demon were quick and smart, they could open the shutter just as one of the fast particles was approaching the door from A and let it into B and shut it again before any other particles could travel through. Further to that, the demon could also stop the slow particles already in B leaving by keeping the shutter closed. Over time, the fast particles will accumulate in A and the slow ones in B, such that the mean velocity of the particles in section B would rise and in section A fall. This, in the kinetic model of thermodynamics, would mean a higher temperature in section A than section B, and through a conventional heat engine we could extract work from our setup. As the shutter is frictionless and the collisions with shutter are elastic (i.e., don't lose energy as they bounce off it), our external actor has performed no work in this operation, but their operation has allowed work to be extracted out of a previously in-equilibrium system. This violates the Second Law of Thermodynamics (Maxwell, 1931).

The immediate questions for physicists are two-fold: why does such a system not naturally generate in nature and, even if not, can such a perceptive being actually exist? Given the nigh-on supernatural abilities our demon must have, we can safely discount such a being at least being currently present to us. Even if they were to exist, their ability to manipulate the world in ways that break our fundamental physics would seem to be a feature, rather than a bug, of their existence. The paradox is only truly worth worrying about if we could create a physical system that operated within the proposed constraints that showed a consistent violation of the Second Law. In 1914, Smoluchowski (Smoluchowski, 1914, pp. 89-121) (Leff & Rex, 1990, p. 11) proposed a physically reasonable modification of Maxwell's thought experiment involving a trapdoor on a spring rather than a frictionless shutter. In this setup, the trapdoor can only open when a particle of sufficient energy pushes through it, meaning faster particles on one side of the door, slower particles on the other sides. However, oscillations within the springs of the mechanism would build up over time, leading to the door randomly opening and closing the longer the system was left to evolve, leading to an equal chance that fast and slow particles transit the gap, returning us to the safety of a working Second Law. In Smoluchowski's understanding, any physically reasonable system designed to model Maxwell's will, over time, always return to obeying the Second Law, i.e. no system can, consistently and reliably, produce reductions in entropy for no work done to the system.

Szilard's engine considers the second of our two-fold questions: can we use intelligence to reduce the entropy of a system despite performing no work on it? Szilard's thought experiment considers this question, the idea of taking a measurement on a system being the way we both acquire knowledge of a system and the point at which we have an associated entropy cost that allows us to remain in compliance with the Second Law.

Consider a box, in thermal contact with a heat bath, that has a single molecule in it. This molecule bounces around the heat bath elastically, transferring energy to and from the heat bath randomly due to the thermal contact. Into this box, we can insert a piston that can frictionlessly move from one side of it to the other. If we insert this piston directly into the middle of the box, there is equal probability that the molecule will be on either side of the piston. The molecule can bang against this piston, moving it further to one side or the other, the direction being at this point random due to us not knowing the position of the molecule inside the box when we inserted the piston. To this setup we add a pulley that, if the piston moves in the direction of pressure of the molecule, raises a weight, and *vice versa* if the molecule applies pressure in the other direction.

The maximum amount of work that can be extracted from this system by raising the weight would be $kTln2$, following the ideal gas law $PV = NkT$, where $N = 1$ (only 1 molecule in the system). As the molecule remains in thermal contact with the heat bath at all points, we can assume that the kinetic energy of the molecule remains the same throughout, and eventually the molecule will via its energy push the piston to one side of the box. The piston can then be removed, with heat extracted from heat bath being converted into work (raising the pulley). This process can then be repeated infinitely by simply putting the piston back in at the point we know it will push the piston in the desired direction again. This setup seemingly breaks even Smoluchowski's modified Second Law, as our system never ends up in a higher entropy state over time (Szilard, 1964, pp. 302-303).

Necessary for this break of the modified Second Law, however, is the impact of human knowledge. If we were to simply put the piston back in arbitrarily, the random bouncing of the molecule means that it could push the piston against the direction we need to extract work, potentially leading the system back to its initial state. What is required here is for the observer placing the piston back into the box is a *measurement* of the position of the molecule, which can allow us to correctly reinsert the piston *ad infinitum* and extract work from the system with no entropy change. Szilard argued that this is simply a development of Maxwell's demon, in that for the Second Law to be saved what we need is a corresponding entropy cost attached to the acquisition of this knowledge about the system (Szilard, 1964, p. 304).

In the next section, we will discuss Landauer's principle, and the concept of logical reversible and irreversible process leading to heat dissipation and entropy change, and how that effects our understanding of "subjective entropy" and Szilard's engine. This will then lead into discussion of Earman and Norton's "sound vs profound" dilemma, investigating two separate branches of solutions to scour Maxwell's demon out of Szilard's engine: whether we assume the second law to be true of the whole system (demon plus engine) or whether we think new physical postulates are required to save the second law, usually via information theoretic approaches. What I will attempt to show is that what is seemingly a mature field of research, investigating the consequences of Maxwell's demon and applications of thermodynamics/statistical mechanics/quantum theory to help exorcise it is, in the words of Earman and Norton (1999) "[the] literature is far from the stable sciences it emulates" (Earman & Norton, 1999, p. 24), again echoing the sentiment of Curiel from Chapter One, when discussing problems in another research field. This similar problem is, in turn, the consequence of over-realist, over-deterministic and over-ambitious attempts to wrestle problems outside of the realm of physics into physics, and another reason to adopt both clear

definitions of the limits of physics and corresponding reasonable philosophies of science, such as constructive empiricism.

3.2.2 Landauer, Earman and Norton – Sound vs Profound

Landauer and Szilard had different concepts as to how the demon was exorcised from the various engines it was present in. For Szilard, the entropy cost that must occur for the Second Law to be maintain was present in the first, measurement stage of the process. For example, for Szilard the entropy price the entire system must pay during the thought experiment that underpins Szilard's engine comes when the demon takes note of what side the particle is on, before the direct heat to work energy transfer. The system must have an entropy cost greater or equal to the work gained by the system in distinguishing between n possible equiprobable states, in units of $k\log n$. For Landauer, the entropy cost came upon the irreversible "erasure" of the information taken by the demon in this measurement, say when resetting the system back to its initial state, which again must be greater or equal to that same $k\log n$ work extracted.

Landauer's Principle, in the context of Szilard's engine and Maxwell's demon, is nicely summarised in Earman and Norton's 1999 work thus:

> *"[Landauer's Principle] sees an entropy cost in the erasure of the memory devices that stored that information. In erasing information that can discern n states, we dissipate at minimum entropy $klogn$."* (Earman & Norton, 1999, p. 5)

As is implied in this, for physical systems people who adhere to Landauer's principle conceive of the demon having to store information about the particle in some physical memory storage unit, which we could think of as being akin to the hard drive in the computer. The erasure of this information limits the demon's ability to break the Second Law in the way. For example, Smoluchowski modified second law works in a different way to that envisaged by Landauer's Principle, by enforcing an entropy rise at the end of any complete cycle of the engine. To start the whole process over anew and continue to extract work from the system, new information on the particle must be acquired that would overwrite the previous information, erasing it.

Both schemes have flaws present in them. Before we investigate them directly, it is also important to be aware of Earman and Norton's "sound vs profound" distinction between different solution methodologies for Maxwell's demon, as arguments using Szilard's and Landauer's Principles us both the "sound" and "profound" broad categories when seeking to show that the demon cannot exist.

The ”sound” programme states that both the demon and the system (for example, Szilard’s engine) are part of the same canonical thermal system and assumes outright the success of the Second Law. Therefore, no new physical principles are required when exorcising the demon, and the necessary entropy rise that the demon should cause must be present somewhere within the system already. This is why physical system such as these don’t miraculously spring into being and demonstrate the failure of the Second Law: it works as is and doesn’t require further modification.

The “profound” programme doesn’t believe the demon and the system must be within the same canonical thermal system. Instead, new physical postulates are required to connect the information the demon requires to function ‘correctly’ and break the Second Law with the entropy rise, or other new, potentially esoteric understandings of entropy. These new postulates can then be used to create a modified Second Law, in which the demon can be exorcised. Landauer’s principle can be seen as style of this, in which temporary decreases in entropy whilst work is extracted from the system can occur *up until* the point at which information about the system must be erased and reset.

Both programmes have their issues. The “sound” programme runs into the problem of potentially not solving any problem. That the Second Law is broken and therefore potentially an incorrect general Law is in fact the problem under question here; simply asserting that it is correct may not be a pathway to any good, general solution. The “profound” programme similarly has issues of not being able to potentially justify its claims to new physical postulates, where new information theoretic concepts may not fully succeed in backing up stated claims.

Rather than address all the possible concerns within this large literature (a thesis in and of itself), to make the point this section will continue to use Earman and Nortons 1999 work, “EXORCIST XIV: The Wrath of Maxwell’s Demon. Part II. From Szilard to Landauer and Beyond” to present a summary of two exemplar critiques of both “sound” and “profound” approaches. In doing so we will show why Szilard’s and Landauer’s Principles can’t be used to successfully exorcise Maxwell’s demon and comment at length about their closing argument: this field of research is not a mature and stable science and contains within it potentially no productive work beyond that which Maxwell did. The paper ends (excluding appendices) with “The Demon lives.” (Earman & Norton, 1999, p. 25), which I take to be another example, as we saw in Chapter 2 and will see again in Chapter 4, that these such physical paradoxes, rather than being an opportunity to discover something new about the universe, instead present physicists with the limits of physics and make the realist case hard to maintain. After discussion of some

of the physics literature related to Maxwell's demon, I will address that point again and conclude that a simple empirical analysis of the world around us demonstrates the success of the Second Law, and that thought experiments such as Szilard's demon are only useful within physics if used carefully, contextually and cautiously. In doing so, we will show again the utility of constructive empiricism, a philosophy of science that allows us to appropriate consider problems like Maxwell's Demon outside of metaphysical concerns and focus on whether it provides useful and good physics.

3.2.3 Can Physics Kill the Demon?

We shall begin here by introducing two separate approaches when exorcising Maxwell's demon. One focuses on the "sound" paradigm, the other focusing on "profound" paradigm. Both are taken from Earman and Norton's 1999 work summarising past literature. For the "sound" principle, let us focus on Brillouin's simple description of Szilard's principle and following issues presented in it.

Let us begin by taking a generalized form of the Boltzmann Entropy as:

$$S = k\log W \tag{18}$$

Where W is the number of microstates in the system. As the system reduces from W_0 to W_1 equiprobable microstates during a process that transforms the system from state 0 to state 1, we can define information (I) as:

$$I = k\log\frac{W_0}{W_1} \tag{19}$$

If we directly connect the change from state W_0 to state W_1 to the information available to us, we can combine (13) and (14) to see any entropy reduction in the system is directly connected to a reduction in the entropy of the system (15). So far, so in keeping with informatic theory concepts, and operating equivalently to earlier sections in this chapter.

$$S_1 - S_0 = k\log\frac{W_1}{W_0} = -k\log\frac{W_1}{W_0} = -I \tag{20}$$

This can be understood as information (I) being equivalent to "negentropy", or a property proposed to be the simple inverse of entropy. i.e., the more information you have about the equiprobable microstates of the system, the lower the entropy of the system.

This doesn't exorcise Maxwell's demon alone. To do so, we need to find a way to make the entropy *rise* for the system as the demon acquires information about the system. To do this, Brillouin simply asserts what is in essence a statistical form of the second, i.e. that there *must*

be an associated entropy cost for acquiring this information in the first place. All the system is, for Brillouin, is a machine that converts entropy into information and then back again, with a greater or equal amount of entropy within the system (Earman & Norton, 1999, pp. 6-8).

This is a patently poor attempt at exorcising the demon. The demon is only "exorcised" in this solution via a rephrasing of the thermodynamic Second Law, which is the very question under consideration. He is very clearly operating under the "sound" approach and has attempted to simply force the demon into necessarily being in the same system as the engine, said any information gained by the system has to cause a corresponding entropy rise at some other point of the system, and called the problem solved. This example from Brillouin, rather than being chosen for being the weakest available (and therefore easiest to defuse), is chosen as its simplicity is its strength. Being the simplest approach is both beneficial for this Chapter's purpose in being written with philosophers in mind, but because it allows us to better understand the failures of all "sound" approaches: they end up having to simply assert the law that is under question to remove the demon[40]. Other approaches within this paradigm often invoke specific thought experiments. Turning away from Earman and Norton temporarily (although mentioned in that paper), consider another thought experiment from Brillouin in the Smoluchowski mould, which attempts a physical refutation of Maxwell's demon within Szilard's engine (though this time, directly addressing rising entropy as opposed to Smoluchowski's which became physically inoperable over time).

Consider how the demon finds the area, A or B, in Szilard's engine in which the particle is located. To do that, he should be able to "see" the particle. In seeing the particle, he must be able to pick them out from the general blackbody radiation within the system, which is kept in thermal equilibrium and therefore is the same throughout the box. The entire box would be black (or red, or green, or blue, depending on the intensity of the blackbody radiation) and therefore the particle unseeable. However, if we gave the demon a torch, he could acquire information about the location of the particle. This, however, is a physical process: a quanta of radiation is released from the torch, hits the particle, bounces back, and is perceived by the demon iff it's sufficiently energetic to be seen above the black body radiation. Brillouin shows that the entropy rise caused by this process must exceed any entropy drop the demon can now cause via their newly acquired information. Thus, the Second Law is saved, and Maxwell's demon is gone (Brillouin, 1951, pp. 334-336).

[40] This includes Maxwell and his original construction of the argument. He conceived of his demon to *criticise* the second law and ask if it was correct, not seek to defend it!

The issue here is that this example is not generic enough to hold weight. Just because we can solve for one specific physical mechanism by which the demon operates, this doesn't mean all physical causes have been undone. And again, further to that, we have simply here attached the demon to the system in a canonically thermal way. What if the demon's measurement system isn't canonically thermal? Earman and Norton (1999)'s Appendix 2 introduces just such a machine, which isn't discussed here in length as its mathematics is outside the scope of this thesis[41].

The "sound" approaches are either circular or too specific to provide the generalised refutation of Maxwell's demon that we require. Next, we shall discuss a "profound" approach and its subsequent failures, finally finishing with a brief example of the failure of "erasure" schemes, such as Landauer's principle, which can lead to runaway entropy reductions simply by carefully considering the memory storage process and algorithmic cycle of our clever demon.

There are two main ways in which "profound" solutions are started, either by people attempting to develop a new generalised Second Law outright using whatever mechanics they think will work, and those who seek to invoke something akin to Landauer's principle, without justifying its inclusion or foundation, as a sort of general panacea to the issues presented by Maxwell's demon. Roth (1964) is a good example of the former. Following on from Brillouin's analysing on entropy-negentropy and its connection to information, he defines an information related quantity, "uncertainty", which he defines as:

$$U = k\log W \tag{21}$$

Such that information is merely the reduction of uncertainty between two separate states:

$$I = U_0 - U_1 \tag{22}$$

[41] For those interested, here is their brief description of the machine: "The system we will consider is a pressure vessel maintained at constant temperature by a heat sink... The vessel is filled with a kinetic gas. Dividing the vessel in half is a thin membrane that will turn out to pass the gas molecules more easily from left to right than right to left. The membrane itself is just a region in which one of Zhang and Zhang's fields prevails. The result will be a pressure differential between the two halves. After an equilibrium pressure differential has been achieved, that differential can be tapped to produce work by means of a device such as a ... frictionless piston that expands reversibly under the higher pressure and recompresses reversibly to the lower pressure. The rate of flow through the device is kept sufficiently low so as not to disturb the equilibrium pressure differential materially. Through the device's action, energy is drawn from the gas which cools. As long as the temperature of the gas is maintained by heat supplied by the heat sink, the device will continuously convert heat energy from the heat sink into work energy, in violation of the Second Law of thermodynamics". Further details can be found pp 31-38.

When adding in entropy to information equivalences, Roth arrives at his new generalised Second Law:

$$\Delta(S - U) \geq 0 \tag{23}$$

This is unfortunately simply wrong. As both entropy and uncertainty are state functions, the introduction of an inequality is groundless: they should simply be equivalent to one another (if information is the change in uncertainty between two states, and entropy is change in information between to states, you have a double negative occurring for any given state, and therefore they're equivalent), therefore producing $S = U$ or $\Delta(S - U) = 0$. (Earman & Norton, 1999, pp. 11-12) Roth's analysis is mathematically flawed but still remains a good example of physicists desiring simple, "profound" new informatic solutions to the problem of Maxwell's demon. The motivations and complexities are showing another example of why philosophers should be deeply weary of the claims of physicists, and yet more demonstrations of incoherence from physicists makes the case stronger that this problem, as we will see in Chapter 4, may be beyond the scope of simple physics resolutions.

Landauer's principle investigates the necessary entropy costs involved in logically reversible and irreversible processes, in short stating that any logically irreversible operations (for example, the erasing of information on a hard drive and returning all bits to 1 or 0) necessarily involves an entropy cost of at minimum $k\log W$. Its usage in Maxwell's demon problems follows from this point. If our demon is able to get information about the location of a particle, it must be stored somewhere to complete the cycle. Once the cycle is completed, this information must be erased to perform a new cycle. If any memory erasure is a logically irreversible operation, it must have an entropy cost of at least $k\log W$ associated with it, which is greater or equal to the work say Szliard's single particle engine can produce. Therefore, the Second Law is saved. Beyond this, those such as Bennett urge that Landauer's principle succeeds where Szilard's principle fails, for much the same reasons that we have stated above: if we can have non-canonically thermal measurement devices, the inductions made by Brillouin and others simply fail. Their work is the necessary fix to previous understanding, and proponents of Landauer's principle state it can finally allow us to exorcise Maxwell's demon.

Or so it seems. Earman and Norton again produce another thought experiment to undermine just such an analysis. Take a computerised demon with two binary states in its memory, L and R. Switching between these states isn't an erasure as it's fully logically reversible, if in L it can only have come from R, and if in R it can only have come from L. As being in one state is fully traceable to being back in another state, logical irreversibility never arises. When the engine is

turned on and the demon activated, it can select one programme in its memory register depending on if it finds the particle to the left of the partition or the right of the partition, the L-programme or the R-programme. Again, for proponents of Landauer's principle, the entropy cost does not come upon the demon acquiring the information, so no entropy cost need be associated with the system at this stage.

The system defaults to the L state and L-programme. If the particle is on the left-hand side of the partition, the programme runs fully through, extracting work via the pistons set up for left hand side energy extraction and cycles back. No memory erasure or reset has occurred here, leading to an entropy drop. If the particle is on the right, the demon switches memory states from L to R (not an erasure) and runs the R-programme. At the end of the R-programme cycle, an extra final step is added: switching back to the L memory state. Again, like the earlier switch, this is not an erasure, or not one that Landauer would consider one. Here again, the R-programme has run, extracting $k\log W$ work from the system via pistons set up for right hand side energy extraction. In either scenario, left or right of the partition, we can extract work from heat with no entropy rise. Landauer's principle has failed to save the Second Law (Earman & Norton, 1999, pp. 16-17).

The important takeaway from this for philosophers is correctly and succinctly summarized by Earman and Norton themselves:

> *"It is our perception that this research programme has been a disappointment if not an outright failure and this leads us to urge that the dilemma be taken seriously as a dilemma rather than an opportunity".* (Earman & Norton, 1999, p. 5)

In the next section of this Chapter, I will concur wholeheartedly with Earman and Norton's analysis of this research field and present the case that the cause of this isn't simply the vagaries and complexities of one single research field, but a fault with underlying foundational approaches from physics and physicists to thought experiments and hypotheticals beyond our epistemic reach. Further to that, I will lay some blame at the feet of realist philosophers of science who take metaphysical and ontological positions based on positions in science (in this case physics) that are not grounded or justified within their own fields, leading to reconfirmation both of the errors within the hard science but also a doubling down that physics alone understand coherently these problems. I will then reintroduce van Fraassen's constructive empiricism as the toolkit by which we can begin to solve these problems, including given a constructive empiricist account as to how we can scourge Maxwell's demon and save the Second Law, something that the realist has failed thus far to help do.

In using constructive empiricism to help frame proposed exorcisms of Maxwell's Demon, we can also step back and see the grand success of the overall project of this thesis. This chapter's central conception, both that entropy is this informatic property, that this informatic nature makes it not solely of the world, and that the resulting failure to be realist about it or the theories depend on it, can be shown to not only be demonstrated, but also resolved. Constructive empiricist approaches don't just form the way philosophers *should* view physics, but also better model the actual pragmatic approaches of physicists. The reason we generate paradoxes like Maxwell's Demon, which produce "disappointing research fields" (paraphrasing the Earman & Norton quote above), isn't itself a failure of research, but a failure of both physicists and philosophers to correctly frame the research. Where they see entirely physics-centred consequences and research methodologies, they should see mathematical argumentation and quasi-philosophy, things that are interesting to physics (the research field) but not necessarily relevant to it.

3.2.4 Possible Solutions and the Benefits of Pragmatism

As we've seen above, scourging Maxwell's demon isn't a simple or easy task to achieve, in fact Earman and Norton believe that it is far beyond our abilities as physicists and may be beyond our reach for as long as we wish to hold to other axioms in physics as resolutely as we currently do. Central to this issue is entropy, the objective-cum-subjective property *de rigour* in this work, which seemingly has a consistent ability to confound, perplex and unsettle physics who seek to formalise its qualia without producing further complex issues to solve. This entire chapter introduces concepts that should unsettle any physicist or philosopher who holds any form of realist stance; is it better for our working statistical mechanics physicist to adopt a new, non-realist, approach?

This section contends that yes, there is. These next sections will cover my proposed set of solutions, some of which are necessary choices and follow from one to the other. I will first begin by outlining a selection of immediate and necessary biases/stances one must take to the philosophy of science, and physics, in order to allow us the ability to exorcise paradoxes like Maxwell's Demon. These in and of themselves do not solve the problem for any ardent realist, in fact these stances may stand in fundamental, unfalsifiable contradiction to their worldview. However, they remain necessary if we wish to seek for a generalist solution to the problems introduced by Maxwell's Demon, and beyond that further entropy paradoxes that are similar to it.

To achieve this, let us return to Section 2.3.1, in which we introduced *stances* and *pragmatic justification*. Boucher and Forbes' description of the pragmatic scientific realism debate can be used as a framework within which we can understand the stance required to adopt a constructive empiricist paradigm, how that differs from the realist stance, and how our "empiricist" stance is superior.

Van Fraassen introduces the concept of the "epistemic stance"[42]. An epistemic stance is not a belief or disbelief about the functioning of the world, nor can it be true or false. It is more readily understood as a policy taken to address a particular question, a value-derived understanding that isn't in and of itself falsifiable. In this way, its success or failure isn't contingent, and people choose whether to have one stance or another via their value-system or innate understanding of the world. Boucher and Forbes counterpose the "empirical stance" with the "metaphysical stance" in the scientific realism debate, the former concerned merely with the functioning of physics, the latter believing that physics must allow us some form of access to "truth" about the universe (Boucher & Forbes, 2024, pp. 8-9).

Stances operate alongside another central theme of this pragmatic turn, pragmatic justification. Pragmatic justification is the belief that we should assess the philosophical success or failure of both realism and anti-realist worldviews based upon the successes they can provide within the context of the individuals goals, policies, and values. Rather than seeking objectivity regarding the logical success or failure of any philosophical belief, we should (and do frequently) assess the merits of various school based upon what they provide *us* as frameworks and structures. A realist is not just a realist because they believe logically that The Miracle Argument is without flaw; they are a realist because realism fits neatly into their value-system, biases and judgement calls about the way the world functions, and they can most neatly justify their beliefs via logical arguments that realism provides (Boucher & Forbes, 2024, p. 7).

For the rest of this chapter, I will be adopting an explicitly empirical stance and will be pragmatically assessing the question of how we scourge Maxwell's demon within this framework. I choose this approach not because I seek to show an indefeasible argument against realism, but because we do not need to so. All that is required for our working statistical mechanics physicist is a pathway that leads them to doing better, more productive physics. If Earman and Norton contend this research field is fundamentally broken, it seems immediately

[42] We have dealt with this concept thus far in the thesis entirely within the structure of Forbes and Boucher's paper, but van Fraassen's understanding forms an entire worldview outside of the applications we are applying it to now. His book, *The Empirical Stance* (van Fraassen, 2004), covers and defends his empiricist stances and values at much greater length than I do here.

appropriate to offer suggestions that are first and foremost pragmatic. In doing so, I will be utilising, as an unspoken starting position of any argument, that the point of physics is to produce theories and models that can allows us to predict the future dynamics of a given system, *and no more*. This is not a proposition or belief, but a value-led stance on the purpose and function of science.

Now we have the groundwork established of stances, let us directly address the process within which the constructive empiricist or pragmatic philosopher of science can approach the problem of Maxwell's Demon and "solve" it. This isn't, and can't be, a "solution" in the manner that a pure physicist would want. Instead of producing a mathematical proof for why Maxwell's Demon can never arise, along the lines of the work of Brilliouin, I will show that, for now and I suspect infinitely, the issue remains entirely theoretical. In doing so, I will show that no true "paradox" exists here and, further to that, demonstrate that not only were Earman and Norton correct about work on Maxwell's Demon being "an outright failure", but that we can identify why it is and see where other potential research failures could occur.

To begin, let us look at the core physics of Maxwell's Demon via a constructive empiricist lens. The Second Law of Thermodynamics, which states that entropy must always rise in a closed system, a cornerstone of our understanding not only of entropy, but almost all areas of physics. The 'you can't get something for nothing' relationship it formalises *vis a vis* energy is both intuitively correct but also fundamentally unimpeachable. It is an accepted axiom of thermodynamics across all fields it could apply to, from quantum-scale physics through to the macroscopic physics of massive stars. The reason Maxwell's Demon is so thorny an issue directly generates from attacking the solidity of the Second Law, which has been tested empirically more times than one could count.

However, we have at our disposal, as constructive empiricists, a tool that allows us to directly test whether we should or should not "accept" a theory: empirical adequacy. Rather than consider the truth or falsity of a theory directly, we can simply take a theory and subject it to the empirical world around us to measure its success or failure. How does the Second Law match up here? Is it empirically adequate?

I would argue that the Second Law is empirically adequate, and that this is beyond current dispute. It is essential and central to numerous theories that in and of themselves are empirically adequate and describes the dynamics of all thermodynamic systems well. It also has demonstrative predictive power, in the sense that you can use the Second Law to

accurately model the future development of a dynamic closed system and verify its accuracy experimentally after later time.

It would be folly to simply list out every area in which the Second Law is empirically adequate, as such a task would extend far beyond the limits of this thesis. However, if we were to be fully rigorous as to the empirical adequacy of the Second Law, we should consider any potential paradoxes that bring the Second Law's empirical adequacy into question, and then analyse if they have explicit experimental tests. One obvious example is found in Maxwell's demon itself, but, up to this date, no experimental setup has been created to test it. We lack the ability as humans to build a system such that we can control for all the variables required for, as example, Szilard's engine, and obtain accurate and reliable measurements. The thought experiment is simply too precise, the external controls too challenging to achieve. Naturally, experimentation is merely one process we can use to test the empirical adequacy of the Second Law: can we observe a naturally developing system in the universe that defeats the Second Law. One proposed, but yet unobserved, area of potential inquiry could be black-holes which have evolved past Page-Time. More discussion on such black holes will arrive in Chapter 4, but for now it merely needs stating that for the Page-time to be reached, approximately half the mass of the black hole must have been evaporated away, which we can easily conceive of currently via thought experimentation, but we have yet to see generate on a macroscopic scale in the universe. Astronomical black holes have the propensity to naturally absorb, via gravity, the matter around them, meaning the physical ones in our universe are more likely to consistently accrete matter than lose it. For now, that pathway of epistemic inquiry is also lost to us.

This attitude to the Second Law, that its centrality and solidity means our priors should be very strong on its future reliability, extends deep into the history of physics. Arthur Eddington, the British astrophysicist talking in the late 1920's, stated:

> *"If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations—then so much the worse for Maxwell's equations. If it is found to be contradicted by observation—well, these experimentalists do bungle things sometimes. But if your theory is found to be against the second law of thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation."* (Eddington, 1948, p. 37)

Our inability to resolves the problem of Maxwell's Demon sufficiently, our inability to create perpetual motion machines (which is what Maxwell's Demon in a sense amounts to) and our

inability to find examples in the natural world of work-doing, non-entropy increasing systems, speaks to the solidity of the axiom I will take forward further in this chapter:

a) The Second Law of Thermodynamics is empirically adequate.

If we accept a) to be true, then what follows is simple: Maxwell's Demon becomes merely a set of ungrounded conjectures that we cannot, or at the very least can't yet, test empirically. Such a problem is still of interest and may prove useful for knowledge gathering in as much as it allows us the ability to see where potential future problems arise, but it is not in and of itself a question of *current, practical physics*. This may seem evident. It is, after all, a problem addressed nearly entirely within *theoretical* physics. But the dogmas of realism still effect the way we view even theoretical problem, which leads back to the problems Earman and Norton outline when discussing the core issue with the entire field. They urge physicists to recognise "that the dilemma be taken seriously as a dilemma rather than an opportunity" (Earman & Norton, 1999, p. 5). We cannot guarantee that new work in this field will solve this issue or open up new physics as yet not understood by humanity, nor can we be sure that this paradox doesn't undermine our current physics in some, as of yet to be discovered way. What we can be certain and of, as of this moment, is that the Second Law of Thermodynamics holds true in all empirical contexts. A perpetual motion machine does not power the computer I am currently writing on, and no-one serious is suggesting we build one to do so. The Second Law holds in all the ways that are important, so thought experiments suggesting otherwise need not be worried about.

There are many potential responses to such a claim. Before addressing them, I will provide a bullet-pointed outline of the argument so far, for ease of future reference.

a) The Second Law of Thermodynamics is empirically adequate.
b) Maxwell's Demon only generates as a paradox empirically if we can find or build systems that are function analogously to it, e.g. perpetual motion machines.
c) If Maxwell's Demon cannot be realised in the physical world, the subject of physics does not need to concern itself with the "consequences" of Maxwell's Demon or the inconsistencies in physical theories it supposes.
d) Research into these inconsistencies will be fruitless, as we attempt to apply the empirical standards of physics to an area of research they cannot be guaranteed to hold, like unfalsifiable thought experiments, e.g. Maxwell's Demon.

By ‘pushing the problem’ further down the line of empirical discovery, constructive empiricism allows us to safely ignore paradoxes such as Maxwell’s Demon. Critics may argue, however, that this is counter to the goals of physics and science in general: if we can simply ignore problems, how do we generate the motivation for new discovery? It might further be argued that if we had simply ignored past “paradoxes” and been contented with our current lot, no further progress would have been made in the progress of science. I argue that this is a misguided view of the motivations of science.

As stated above, rather than entirely breaking the cycle of physics progression, all the constructive empiricist is doing here is demonstrating the only necessary methodological progression for new theory development within physics and science. Physics doesn’t *require* new theories if current theory functions as intended, instead only needing them describe and make useful new empirical observations. As such, the motivation to progress physics comes not from new theoretical work done from the library, but from laboratory or field environments, which generate the conditions for the theoreticians work in. Rather than a pathway of “old theory → new theory → experimental evidence for one or the other” we have “old theory → experimental evidence showing a lack of empirical adequacy in old theory → new theory → experimental evidence shows empirical adequacy in new theory”.

Point b) and c) may also provide some cause for concern for our realist-inclined working physicist. Whatever the theoretical state of Maxwell’s Demon, its subject matter is physical in nature, and it is asking questions internally about physical laws and theories. Surely this should be of relevance to how we think about our physical theories? Surely it would be folly to discount its warnings?

This can also be a concern for the constructive empiricist. As is mentioned in Chapter 2, the constructive empiricist doesn’t simply want to dismiss all “as yet to be observed” phenomena as irrelevant to the empirical adequacy tenets of our theories. In fact, they are fundamentally opposed to such a consideration on the basis that doing so undermines the motivation for new and better discoveries. How can we be unconcerned about the warnings Maxwell’s Demon gives us if the mutually incompatible concepts under discussion *are* in some way, shape, or form theoretically observable, even if we have yet to observe them?

More discussion related to this point follows in Chapter 4 regarding black holes, where an observable/unobservable distinction is found for the internal structure of black holes that help solve other paradoxes related to entropy. Here, I will make the case that the entities involved in Maxwell’s Demon are unobservable and that point c) holds.

Let us address the observability of "entropy", the property of a system. As we've established, entropy is in part a subjective property of the world, dependent on the observer choices. Because we can have equivalently accurate quantitative description of the entropy of a system via informational differences or observer choice, it follows that the 'empirical' outcomes of our theories where information theoretic understandings of entropy exist can also be different. As is shown in section 3.1.5, two empirically accurate understandings of the universe can be generated by one theory dependant on observer choice. Observability, for van Fraassen and the constructive empiricist is a veridical aspect of the world, where "X is present to us, we observe X" and this should hold true across multiple observers of an observable entity. Whilst this heuristic isn't a fully precise definition, entropy as a property cannot fulfil this criterion: our theories can generate many forms of X under circumstances Y which can have empirically different results for entropy despite identical underlying systems. Ergo, a system's entropy isn't "observable", and if that is the case, paradoxes involving measuring the entropy of a system can also be argued to not be "observable". If we are constructive empiricists, we don't have to concern ourselves with being correct about the unobservables, which means that the paradox doesn't generate in the first place. The strength again of the constructive empiricist argument can be found in providing a simple but rigorous approach in which we can frame why many paradoxes that feel unresolvable are in fact questions physics doesn't or can't answer: the dynamics causing the problem's generation are actually unobservable.

Another approach to take in defending points b) and c) is the overarching success of the Second Law of Thermodynamics. Pragmatically, it makes more sense as a working physicist with finite time to work on problems that present extant, solvable issues, or at the very least work with complex areas productively. The words of Eddington, Earman & Norton and many others make the case that the Second Law is uniquely unchallengeable and successful, and that working to undermine this is the act of folly. Our failure to build a perpetual motion machine isn't because our physics has yet to find some loophole to the Second Law, there simply is no loophole. To note here, this is not some argument for realism via the backdoor. We do not have to be metaphysically committed to the "truth" of the Second Law to recognise the predictive and descriptive power it has. The strength of the constructive empiricist case over that of the realist can still be found in fact that we can develop alternative, rigorous models for the universe around us whilst remaining epistemically modest and removing metaphysical conceptions. The constructive empiricist doesn't have to commit themselves to beliefs about the world to argue the success of, for example, the second law. They can base that entire worldview on the

empirical adequacy it has and ignore theoretical thought experiments because they aren't empirically grounded.

Broadly put, the Second Law can have no descriptive or explanatory power with respect to the structure of the universe and still be empirically adequate, and it makes little sense for the constructive empiricist, committed to the "rationality of science" (Monton & van Fraassen, 2003, pp. 407-408), to be un-swayed by the orthodoxy within physics that the Second Law in inviolable in ways other physics theories are not.

Connected to this, another argument for b) and c) can be found in analysing what makes a theory "physics-based" or "physical" (i.e., can only be necessarily understood through the methodology and practice of physics). We have established necessary and sufficient conditions for physics already for the constructive empiricist through empirical adequacy, and we do not have to be bound by further intuitions on whether a question, however formulated, is physical or "science-y" in nature. Maxwell's Demon, being as we've already established a truly theoretical paradox (there are no instantiations of perpetual motion machines or the like within our observable universe), can equally be formulated as a purely philosophical or mathematical question, depending on the language used. The 'physics theories' used as axioms to define and describe the paradox are physics concepts and theories that throughout their paper Earman & Norton demonstrate are either not fully grounded within physics or are fundamentally incorrect understandings on the underlying physics (Earman & Norton, 1999). How we view and understand the paradox is implicitly related to the way it is presented, i.e. by 'physicists', in 'physics journals', using the language and styles expected of physics. All of this, however, does not make it 'physics[43]', at least within the constraints of how a constructive empiricist would view the subject. The constructive empiricist knows firmly what defines physics, the empirical study of the universe around us, where the theories it produces are either accepted or not accepted based on their empirical adequacy. The Second Law is empirically adequate precisely because we haven't observed any scenario in which it fails, which can lead to the conclusion that we will never be able to observe anything that operates like Maxwell's Demon does. Worrying about Maxwell's Demon is therefore folly and a waste of research energy, as the

[43] 'Physics' here is an exceptionally broad concept, but by 'physics' I am simply referring to the empirical study of the world around us as described in Chapter 1. "Physicists" can both be correct that the work they are doing is 'physics' in the sense that they are doing it and society can apply a "purpose of a system is what it does" approach, and incorrect when 'physics' is viewed as delineated from other areas of knowledge acquisition via its empirical nature. Realists, I maintain throughout, require the second understanding of 'physics' to be the meaning of the term for their understandings to hold, as any naturalist position requires solid empirical foundation to back up future metaphysical claims.

question itself hasn't generated. Interesting philosophically, maybe, but not interesting physics.

Finally, we can observe a parallel between the work of Earman and Norton and point d). Whilst Earman and Norton don't identify reasons for their beliefs about the failure of the research project associated with Maxwell's Demon, I contend that the failure is due to the presence of the aforementioned issue at hand: the problem, involving the non-objective property entropy, is beyond the necessary scope of empirical inquiry, and therefore cannot be solved satisfactorily via purely empirical means. To ask a physicist to simply use the tools of physics to resolve this "paradox" is not only a waste, but unfair as a request, as until we can produce or find a system with the same or similar dynamics present in, for example, Szilard's engine, we will be unable to test empirically the result of said theory. Any theory-crafting up until that stage necessarily requires deductive reason not necessarily grounded in the empirical observation of how a system functions, which in this case leads to a 'failed research project' from which physicists are only left with dilemmas.

3.3.1 Where Do We Go from Here?

Two pathways then behove the working physicist who considers the question that titles this section and is willing to accept the constructive empiricist case over the realist case. They can either take an 'avoidant' route, wherein they ignore the consequences of Maxwell's Demon and the holes it proposes to show within Statistical Mechanics, thereby avoiding the issue; or an 'adaptive' route, wherein they structure their thoughts and ideals as not merely "physics", but as an interdisciplinary effort involving explicitly deductive reasoning based on stated stances. These stances can be metaphysical in nature if so chosen, e.g. one could say "Szilard's engine could exist in the real world and be created via human abilities", but these stances go beyond what physics can supply, and potentially what any other source of investigation could, depending on your wider philosophical stances.

Further to this, because of this chapters strongly defended case that entropy is subjective, many issues that cause the realist concern never faze the constructive empiricist. Rather than seeking to view properties like entropy as of the universe and connect our ontologies to the success of them, we can view them as pragmatically as part of a toolkit and be unconcerned with future metaphysical implications, as long as the theory is empirically adequate. Entropy is some odd and confusing thing that depends on the choices or epistemic access of an observer? Fine. There is this theoretical paradox that suggests that empirically adequate physics theory is wrong? It does not matter; these systems are not empirically observed in the

world, depend on dynamics that are unobservable, and any theory crafting seems to make physics less progressive. We can use constructive empiricism to correctly dispose of these problems and for our working physicist, it allows them to distinguish things they *can* solve via empirical methodologies alone and those things they *can't* solve without extending themselves to potentially entirely theoretical concepts. When Brillouin proposed solutions to Maxwell's Demon, I maintain that he had progressed from forming arguments akin to Newton's work on gravity to asking "how many angels can dance on the head of a pin?"

Now that we have identified a case study example for the application of constructive empiricist thought and where it can provide benefits for physicists via understanding the limits of physics methodology and the benefits of epistemic modesty about what we can observe, we can now extend this out to other areas of inquiry. In the next Chapter, we will discuss Black Hole Paradoxes, their origin within physics and the benefits of constructive empiricism in another area of ongoing physics research. Also within the next chapter, we will see areas where constructive empiricism heuristic on observability seems to fail in a context it otherwise shouldn't, but, via a holistic analysis van Fraassen's work and a rigorous defence from both philosophy and physics, we can produce a new heuristic that I maintain is key to adequately applying constructive empiricism to our modern scientific theories. This new extension of both epistemic communities and van Fraassen's heuristic on observability is key in allowing the constructive empiricist to offer a more attractive account in the context of black hole paradoxes.

Much like the work in this Chapter, Chapter 4 will show that these problems centre around entropy, both as a property and conceptual tool for predicting the dynamics of a system. Whilst avoiding the information theoretic concepts introduced within this Chapter, Chapter 4 will show that entropy's connection to information and the epistemological consequences it causes are essential to grasp not just for physicists but also for philosophers of science who wish to create overarching structures for modelling science. The view one takes on entropy leads to direct philosophical consequence in a way unlike any other property within physics, meaning it should be handled with caution for fear of causing more problems than you solve.

Chapter 4 – A Constructive Empiricist Approach to Black Holes

And on to black holes. As will be, and has been, discussed at length in this thesis, black holes offer new, complex problems for both the physicist and the philosopher of science. Our working black hole physicist operates at the intersection of numerous fields of physics and mathematics, and the research field itself is still developing and growing with each passing year. Over the course of this writing this thesis, new work has been produced by Almheiri et al. and others that, whilst not changing the conclusion of this thesis that constructive empiricism is more attractive than the realism for the working physicist, demonstrates the fluid and active nature of the research ground below our feet.

This chapter has three main goals:

a) To show that black holes are epistemically complex entities that defy easy and quick understanding. In doing this, this chapter will show that problems exist *prima facie* for constructive empiricism when considering whether black holes are observable. This problem appears to generate the conclusion that black hole interiors are observable, despite the constructive empiricist having good reasons not to want this.
b) To defend the conclusions of a) for the constructive empiricist. I do this by revisiting van Fraassen's understanding of what an epistemic community is, offering a novel friendly amendment to his account that gives us a rigorous way of showing that black hole exteriors are observable, and black hole interiors are unobservable. This is attractive for both the physicist and philosopher as it grants the constructive empiricist the ability to resolve problems like the Page-time paradox in ways the realist cannot. This understanding of black hole observability, I will also maintain, is hinted at by our physical theories on black hole dynamics.
c) Revisit the Page-time paradox and using the pathways for our working physicist outlined in section 2.2.5, analyse if we can directly either endorse a solution or exorcise it entirely using our new, novelly extended, constructive empiricism, and how this approach further demonstrates the inferiority of realism compared to constructive empiricism.

In achieving these goals, this most important thing this chapter does is expand an aspect of constructive empiricism that I maintain hasn't been fully expanded by van Fraassen, epistemic

communities. Rather than just the collection of all humanity van Fraassen takes it to be, I believe that a view that takes into account not just epistemic ability but also requires the ability to *communicate* between different parties, gives us both a more holistically useful understanding of what an epistemic community is, but also correctly accounts for the "weirdness" of some black hole physics: if it is deeply unintuitive and beyond conceptualisation, you may be talking about things outside our epistemic community, which physics can't adequately discuss.

All of this combines to the central conclusion of this chapter; black holes operate at the intersection of what is possible in physics and what isn't. Once we step outside the bounds of what *can* be purely empirical, we can no longer rely on the rigorous methodologies of physics to provide us answers that are comfortable for a realist. This thesis gives a rigorous underpinning to such an assertion within an extended constructive empiricist framework, and without this extension I maintain it is harder for the constructive empiricist to do so. This application of a subject that can only be empirical (physics) to areas that humanity can't access empirically (the black hole interior) is the central cause of black hole paradoxes, and any solutions to the Page-time paradox will only be found via finding ways to be empirical again, or via ignoring the issues on pragmatic grounds. These thoughts are combined in section 4.2.1 to outline why constructive empiricism is more attractive than realism, which centres empiricism first and foremost in its analysis of science, as opposed to realism, which I maintain must explain how it can maintain metaphysical commitments about scientific theories so obviously beyond the scope of what we can know.

In terms of the structure of this chapter, firstly we will consider what it is to observe a black hole and how the esoteric structure and dynamics of black holes affect our assumptions about their observability in a constructive empiricist context. Secondly, we will introduce a dilemma for the constructive empiricist, showing how considering the entirety of a black hole either "observable" or "unobservable" can lead to problems. Thirdly, we will reintroduce epistemic communities, giving a novel description of them, and use them give an answer to the question "are black holes observable?" that the constructive empiricist can accept. This leads into the next section, where we will demonstrate that current physics has already anticipated this solution. Finally, this chapter will return to the Page-time paradox, showing that constructive empiricist approaches allow our working black hole physicist to coherently dispose of the paradox, using with our new understanding of what an epistemic community is being key to those resolutions. Overall, I will show that whilst the realist case is required to grapple with these paradoxes to provide a consistent theory of what is going on inside black holes, the

constructive empiricist can be satisfied with a physical theory that gets things right about black hole exteriors, even if what it has to say about black hole interiors looks paradoxical. Further consideration is also given to the most recent developments in black hole physics, such as those covered in Chapter 3, section 3.1.8, where I conclude that the realist case is still inferior to the constructive empiricist case. I make this case on the grounds that realism is still required to be epistemically immodest about observables, even if our theories can now provide a theoretically empirical basis on which we can test and validate said theories on black hole dynamics.

4.1.1 How Would We Observe a Black Hole?

We have in this thesis already outlined what constructive empiricism is. In this section we shall briefly define it again, focusing on the observable/unobservable distinction. This is key not just to chapter directly and highlight some critiques of the constructive empiricist viewpoint not covered in Chapter 2. This is all done with the goal of establishing whether black holes are observable for the constructive empiricist, because the initial reading does not seem cut and dry.

Constructive empiricism is a school of thought that seeks to bypass tricky metaphysical questions about truth and existence in approaching science and instead replace them with a simple, consistent heuristic: one can accept a theory if it is empirically adequate (that is, roughly, if it is correct in what it says about observables). Such a school hopes to develop an internally consistent understanding of science that avoids the problems that affect realist philosophies whilst modelling the actual practice of scientists as close as possible. In doing so, the constructive empiricist must define (or at least give a good heuristic for) what is observable in the universe and outline the epistemic skills that scientists need to observe something.

This framework is easy to outline for conventional, everyday objects. A tree is observable, as someone can see a tree with their own eyes. Therefore, theories about the operation of trees need to be consistent with all the processes we can observe a tree display, e.g. its leaves falling off in autumn and reappearing come spring. If a theory manages to adequately describe all these observable processes, we can accept our new theory on trees.

For the constructive empiricist, all the entities posited by scientific theory can be sorted into two camps, those that are "observable" and those that are "unobservable". The constructive empiricist states that the aim of science, its core purpose, is to provide empirically adequate theories. Empirical adequacy is the threshold where a theory’s propositions about reality correctly model observable aspects of reality. If a theory fails to accurately model such a thing,

for example, "the precession of the orbit of Mercury", then the theory would not be empirically adequate and should be replaced by an empirically adequate one. A constructive empiricist limits the "observable" entities of the world to merely the things we could perceive around us unaided. For example, a theory should be able to predict the movement of a magnet to an iron bar but has no requirements to get things right about the unobservable objects it posits while generating such predictions – e.g., the electromagnetic force and its charge-carrying particles. In general, van Fraassen and the constructive empiricist want to say that all things in the macroscopic world, the trees, plants, and animals around us, are observable, and the theoretical entities of physics, such as the electron or photon, are unobservable. This distinction is essential as van Fraassen views the aim of science as describing only the observables correctly, everything else being less critical (van Fraassen, 2005, p. 113). Van Fraassen presents this heuristic for helping deduce whether an entity is observable:

X is observable if there are circumstances which are such that, if X is present to us under those circumstances, then we observe it (van Fraassen, 1980, p. 16).

The usage of the word "present" contains some ambiguity within it. However, van Fraassen goes on to support the notion that things that are "present" to us are the things we have direct epistemic access to, i.e., for an object to have the property of being “present to us”, it must be in some way accessible to us in the world directly. A tree is “present” if circumstances are right then we would be able to observe it (Monton & van Fraassen, 2003, pp. 415-416).

As we can see from these heuristic van Fraassen presents, what we consider “observable” isn’t something that can be rigidly defined, i.e. we cannot simply provide an overarching split between the observable and the unobservable. Instead, we are best left to answer this question on a case-by-case basis as we consider different objects. Many philosophers take this as a fundamental weakness of constructive empiricism. They contend that van Fraassen has no fundamentally principled reason for drawing the lines where he does, in doing so offering little useful in answering the question “what is observable?”. Churchland and Teller both make arguments to this effect. Churchland maintains that van Fraassen’s arguments about the very small should also apply to the very far away, both in time and distance. If the constructive empiricist was forced to remove these from observability, too, we are left with very few “observable” objects (Churchland, 1985, pp. 37-40). Teller more directly attacks the concept that the microscopic is inherently a realm of unobservable entities, stating that any entity that is “viewed” through an optical microscope doesn’t require the interpretative steps that van Fraassen maintains are present when using any microscopy technique, as in the case of optical

microscopes we are simply enlarging the image of an object using the same physics that our eyes use to perceive the world around us. For Teller, no distinction can be drawn between viewing a macroscopic image of a tree through a mirror and viewing a bacterium under a microscope (Teller, 2001, pp. 128-130).

Both critiques miss a critical concept: observing an object isn't solely the activity of one observer. Instead, for van Fraassen, observation requires a non-interpretive aspect, whereby what we are viewing can be directly viewed via the eyes of observers without necessary recourse to some piece of external machinery. In the case of microscopic objects, such as viruses or bacteria, we will never be able to view them without some recourse to an external tool to produce an image of them which is large enough for us to see. On the other hand, whilst seeing a tree through a mirror also uses external tools to produce an image we can see, we could also simply turn around and look directly and perceive the tree, no external tools required, as it is macroscopic. In fact, observation through a microscope can never occur, as we never see the object itself, merely an image of it enhanced and enlarged using technology and optics. What we see is simply an image of the object rather than the object itself. With the macroscopic world, we can cut out the middle man and view these objects directly, rather than rely purely on images brought through external instrumentation (van Fraassen, 2001, pp. 156-158). The same applies to objects very far from us, both temporally and spatially: observation is an agreed-upon collection, an aggregate, of the things we humans all have direct access to. This is both the cause and saviour of van Fraassen's vagueness when defining what observability is; it is fundamentally contingent on the ambiguity present in humanity to describe its own abilities (van Fraassen, 1985, p. 160).

Modern physics contains many examples of non-conventional entities. Black holes present problems for the constructive empiricist, arising from the fact that they, being large and gravitationally strong, affect the processes by which humans interact and perceive the world, bending energy and matter back on themselves in a manner that makes them completely opaque (or, *black*) to external observers. How do we categorise black holes, therefore? Are they observable or unobservable?

To understand whether a black hole is observable or not, we should begin by again outlining the main properties and dynamics of a macroscopic black hole, focusing on those relevant to the concept of observing black holes. A macroscopic black hole is a large region of space-time containing at its centre a singularity typically caused by the gravitational collapse of some supermassive star (e.g. Cygnus X-1 in the constellation Cygnus). In this region, gravity is so

strong as to prevent even the escape of massless photons, giving it a 'black' appearance to all external observers. The event horizon is a closed surface that contains the singularity, defined by the set of "points of no return" at which photon escape is impossible. The event horizon functions as the boundary of the black hole. It is important to note that whilst the event horizon is something that is visible to an external observer, it is not a physical boundary. Due to Einstein's equivalence principle, any observer passing through should note no physical changes between either side of the event horizon. The event horizon is not some 'wall-like' concept; instead of being more akin to a border between nations inside the Schengen zone (i.e. a boundary which isn't immediately noticeable as you move through it, but which does delineate two separately functioning systems on either side of it) (Rindler, 1956, p. 663)

We have already established in this thesis that, because of the event horizon's specific properties, we can separate the space either side of the black hole into the "black hole interior" and the "black hole exterior". The black hole interior covers the region of a black hole within the event horizon, i.e. the part of the black hole from the singularity to the event horizon. The black hole exterior is the region outside of the event horizon where black hole physics can still be observed[44] to, in theory, infinity. The motivation for splitting these two sections is to delineate the region of a black hole we can naively see with our eyes and the part of a black hole shielded from us by relativistic physics. As will become evident later in this chapter, this epistemic difference is not to be treated lightly and forms the centre of the problems the constructive empiricist can run into when considering the observability status of black holes *in the round*.

Where does all of this leave black holes? These are macroscopic objects that are very far away and yet in principle, it is possible for them to be viewed up close and personal by a human observer. Whether this thinking can be extended naturally to the conclusion that black holes are observable for constructive empiricist is the question of the next section. I maintain that when such an extension is performed, and we consider a black hole (the whole system, interior and exterior) observable, dilemmas present themselves for the constructive empiricist, especially regarding the black hole interior. In the next section, I will outline these dilemmas.

4.1.2 The Dilemma for the Constructive Empiricist

To answer the question of black hole observability, we need to address the central aspect of van Fraassen's heuristic, i.e. if we were to get close enough to a black hole, can we perceive it

[44] This can range from microscopic processes such as Hawking radiation, the process by which black holes "evaporate" away over time (Hawking, 1975, pp. 202-203), to the sheer application of gravitational force on macroscopic objects, e.g. a star orbiting a black hole and X-ray emitting gas-infall.

unaided, using the faculties available to us solely as human beings. Black holes are huge, macroscopic objects with large areas of effect that contain macroscopic objects. Suppose a star was to be caught in the gravity of a black hole and begin orbiting it outside the black hole's event horizon. In that case, we could easily imagine a scenario where an observer can perceive the orbiting star with purely their eyes and thereby directly perceive the physical consequences of a system with a black hole and singularity at its centre. We can also observe the event horizon because we can "see" a black area of space that blocks the view of energy and matter we otherwise could perceive. Further to this, it should be possible to detect the Hawking radiation and evaporation of a black hole either directly (via readings of the radiation leaving the black hole) or indirectly (via the "black" region bounded by the event horizon shrinking). For the constructive empiricist, this region of the black hole, the "black hole exterior", should be considered observable, and equally non-problematic.

The "black hole interior", however, is another question entirely. Due to the equivalence principle, if we remain outside the event horizon, we cannot possibly view objects that pass through the event horizon into the "black hole interior". This is not a nuanced position; it is absolute. Black hole interiors are, by definition, "black" and closed off to observers who are positioned outside the event horizon. Once something passes from a region in which we could observe it (say a space probe that our intrepid observer releases on a path designed to fall into the black hole) into the black hole interior we cannot know even of its existence beyond a certain point, never mind being able to see its physical state with our own eyes. How can we reasonably claim the black hole interior is "observable" to those of us positioned outside of the event horizon when it is the region that contains the black hole's core dynamics and mechanics and is therefore observationally shielded from us? They may well be macroscopic objects within this region, but how is that relevant when the core tenets of general relativity demonstrate we can never access them?

Does this mean that, for the constructive empiricist, black hole interiors are unobservable? As an argument against this position, it is still perfectly plausible to construct a thought experiment where a human observer passes through the event horizon both unscathed and unknowingly, being able now to perceive objects that have previously passed through the event horizon but are further away from the singularity (such that photons can travel from the newly perceived object to the observer's eyes in sub-light speed times). In this thought experiment, objects that once had to be discarded as "unobservable" due to the region they had traversed into would suddenly become observable again, merely due to the observer's change in spatial location. This thought experiment recalls a similar argument between Churchland and van

Fraassen. Churchland maintained that the constructive empiricist account made no explicit distinction between the very small and the very far away. Van Fraassen's response was that very far away objects can themselves be considered "observable" as they are macroscopic and therefore the arguments that hold for the things we see before us must also apply modally to those objects not yet seen, but theoretically visible to purely our eyes. This response, however, seems on shaky ground within the context of black holes. Once our observer perceives the newly rediscovered space probe within the black hole interior, they cannot send information back to an external observer due to the super-light speeds required to transmit that information through the event horizon. No direct description, which human observers who are located outside of the black hole can partake in, can be achieved. Here, we have a macroscopic object (space probe) that has been perceived by a valid observer in a correct spatial location that operates under the same physics as the rest of the universe. Van Fraassen would be forced, by his own heuristic, to consider the black hole interior observable, even though we would be unable to ever perceive the space-probe from any location outside of the event horizon, which is where all of humanity and all physics work is located.

In effect, the "black hole unobservability" position requires there to be some break either in the operation of physics in black hole interiors that would explain why what is occurring between the space probe and the observer isn't perception or a fundamental flaw in van Fraassen's heuristic quoted in section 4.1.1. Neither is acceptable for the constructive empiricist, as the consequences of accepting either position are too significant. However, the converse, "black hole observability", still runs into the same issues outlined above in this section. Van Fraassen cannot consider a black hole interior to be "observable" due to the implicit desire for epistemic modesty present within his conception of constructive empiricism. To explain this, consider that:

A) The constructive empiricist advocates belief that to accept a theory as empirically adequate it only must get things right the observables, because they think it is immodest to think that we could get things right about unobservable matters.
B) If we consider black holes observable (in this case viewing both the black hole interior and exterior as one single entity) then this implies that our theories about black holes must be empirically adequate for them to be accepted, which further implies that these theories can be correct about things that go on within the black hole interior.
C) This is turn requires that we must be able to *know* things about the black hole interior, which when we are outside the black hole event horizon we cannot, as relativistic

physics prevents information transfer from the black hole interior to the black hole exterior.

D) Therefore, black holes should not be held to be observable, as we must either soften our requirements for empirical adequacy to save our epistemic modesty, or the converse. This is something the constructive empiricist will never be comfortable with.

Epistemic modesty is the entire underlying motivation for van Fraassen's work in observation, the point of his heuristic, and the primary way he seeks to solve problems that plague realist philosophies of science. On the other hand, a literal reading of the constructive empiricist account of observation would force us to consider black hole interiors observable. We could view them if the circumstances were present to us to do so. Despite this, it is still the case that if we could leave Earth, enter a black hole, and report on our surroundings, there would be no way to transmit that information back to other humans on Earth. If someone observes something but has no way of making that observation known to anyone else, can it indeed be said to be observable to anyone other than the initial observer? Black hole interiors demonstrate a case wherein one must choose between van Fraassen's desired epistemic modesty and his own heuristic on what is observable, an untenable position for the constructive empiricist.

Two pathways are open for fixing this issue:

1) There is some, yet undiscovered, esoteric physics at work in black holes that recovers the ability for objects within the event horizon to be observed directly, i.e. there exists some mechanism by which information is retrieved from the black hole[45],
2) Our understanding of a publicly available, direct description is flawed. There are potential splits within humanity's ability to perceive objects and inform others of their perception that allow us to consider black holes in sections, one observable and one unobservable, depending on the region an individual observer is in at that time. This could be further achieved by relativising the concept of observability to a group of

[45] This may not be a helpful pathway for the constructive empiricist: if the information that is retrieved is entirely invisible to the human eyes (e.g. all information recovered is similar in character to Hawking radiation), then it would still be "unobservable" to van Fraassen, not allowing for information to be recovered in a manner he would accept. However, as we have *no* concept of what said information recovery would be or by what process it would proceed, it is still possible that we could recover objects in exactly the way they entered, by arcane and as yet unknown quantum physical processes. For this reason, I have left this pathway open as a possible solution to this dilemma for the constructive empiricist.

observers, conceding the concept that "observability" as a property cannot be established by a single observer alone.

Route 1) is mainly discussed by physicists, with recent work by Malcedena *et al.* (Almheiri, et al., 2020b) and Akers *et al.* (Akers, et al., 2020) demonstrating that black holes, over sufficient time scales, must return information to the universe via modelling of the black hole analogues within AdS/CDT (anti de-Sitter/conformal theory correspondence) universe structures, such that as their macrocanonical entropy drops via Hawking radiation, their entropy of entanglement must also fall, meaning information returns to the universe during black hole evaporation *somehow*. The process by which this would occur is still unknown, and this work is in its very early stages[46], so this chapter doesn't analyse this work in depth in favour of a philosophical solution to the problem of black hole observability along the lines of route 2), one which van Fraassen uses as a defence of the constructive empiricist position in different contexts, epistemic communities. The recent work is returned to in the final sections of this chapter as foundational to the understanding that a constructive empiricist approach is superior, and further outline that our black hole interior/exterior distinction is not just found in philosophical analysis but hinted at by the most recent research work.

4.1.3 Epistemic Communities

In section 2.3.2, we introduced van Fraassen's understanding of what an "epistemic community" is. Section 4.1.3 and 4.1.4 will make the case that van Fraassen's understanding of what an epistemic community constitutes can in fact be extended via the introduction of an "ability to communicate with each other" necessary condition. I believe that this extension is both in keeping with constructive empiricism as a philosophy of science and in fact help the constructive empiricist answer the question "are black holes observable". To make this case, I will revisit van Fraassen's understanding of what an epistemic community is and restate it usage in respond to the critiques offered by Churchland discussed in Section 2.3.2. From here, I will introduce my novel approach to epistemic communities, demonstrate its utility for the constructive empiricist and make the case that physics already points us in the direction of adopting it.

As a reminder, an epistemic community consists of a grouping of humans that share fundamentally the same epistemic abilities, in which members can all equally trust the

[46] And is potentially not of great importance to the constructive empiricist, depending on what the "information" constitutes in a practical sense. This form of solution is not central to our analysis on the basis of how new these approaches and conclusions are, but it may be fruitful to return to them some years down the line.

reported epistemic experiences of each other and thus form a collective agreement on the basics of what is and is not, epistemically accessible. Such an understanding is essential for the good practice of science. Without it, if we try to keep as strict as empiricism as possible, we end up with solipsistic epistemologies. Epistemic communities can be construed differently in different epistemological and societal contexts. The largest community we can have is the set of all human beings that have/have had higher-level consciousness along with perception and cognition abilities. This epistemic community consists of all human beings. It is the baseline van Fraassen takes when invoking them in response to another of Churchland's arguments against constructive empiricism (van Fraassen, 1985, p. 254).

In this argument, Churchland asks us to consider an electron-microscope-eyed humanoid that is functionally identical to an average human but has an electron microscope for a left eye. For this humanoid, microscopic entities such as viruses, bacteria and crystal structures would have to be considered observable, as they could "see" them in the same way any other humanoid could see a macroscopic object. Importantly for Churchland, however, there is no difference between the humanoid and an average human simply looking through an electron microscope. It is up to the constructive empiricist to explain why these two functionally identical entities (our electron-microscope-eyed humanoid and our human looking through an electron microscope) have such deeply divergent understanding of what is observable (Churchland, 1985, pp. 44-45).

Van Fraassen responds to this argument (van Fraassen, 1985, pp. 256-257) by saying that we only have to concede this as a challenge if we can say humans share in the experience of having seen bacteria via the testimony of our humanoid counterparts. Given both parties' vastly different perceptual ranges, such a declaration is not trivial. We can only comprehend the world based on a shared set of references underpinned by our epistemic experience of the world around, references the humans and electron-microscoped-eyed humanoids may not share. Humans and our humanoids are within different epistemic communities with differing epistemological limits. As our science only occurs within the epistemic community that contains all humans (and contains no electron-microscope-eyed humanoids), this thought experiment does not challenge van Fraassen regarding what counts as observable, as our understanding of observability only comes from within the human epistemic community. Given that it seems trivial that both our electron-microscope-eyed humanoids and humans have access to differing levels of entities, making the case that they are part of different epistemic communities falls out trivially for van Fraassen. For example, even if it were the case that these humanoids could learn English and express to us what microscopic objects look like, we could

never have the sort of direct experience that allows our epistemic community to have confidence in such a description. For van Fraassen, observation, perception and, from them, science are all grounded within an agreed-upon epistemic community that shares a language and way of viewing the world based upon a shared set of references, further grounded in a shared set of epistemic abilities. One can only talk reasonably about questions of observation if one first identifies that all observers share a singular, unique, epistemic community.

Having now established van Fraassen's description of epistemic communities, this chapter will show that his usage be developed and strengthened to provide an improved understanding applicable to problematic questions about black hole observability. The proposal reinforces van Fraassen's original description of epistemic communities and consists of adding an additional requirement beyond shared epistemic abilities, such that if two people share an epistemic community, they must:

1) Are part of a community with the same theoretical range of epistemic abilities such that, all going well, they can see the same things, hear the same things, touch the same things, etc., and
2) Be able to communicate with each other, in every direction, even if over long distances and times,

Thus, any grouping of people that fulfil both these requirements would be in one unique epistemic community. For example, one can group all human beings living on Planet Earth as part of one epistemic community. We all have the same (theoretical) limits on what we can perceive and communicate that information to any other human being on this planet in finite time. This epistemic community would also hold for all human beings at any theoretical distance, assuming non-pathological space-time geometries. Two human beings, not separated by a space-time singularity 70 light-years apart, could be part of the same epistemic community (although, at a minimum, it would take 70 years for information to go from one to the other).

The conditions above are the conditions for two individuals sharing the exact same epistemic community but, in the context of black holes, more nuanced positions can develop. In said situations, due to the fact we can theoretically send information in one direction but not the other, we don't have two separate or one single epistemic community, but epistemic communities that can be though to stack on top of one another. In this setup, an observer within the black hole interior can receive all the information an observer outside can, plus some more within the region they have now accessed within the black hole interior, meaning they are

in differing epistemic communities with large amounts of overlap. This potential non-reciprocal relationship between epistemic communities is where the power of our new understanding derives, with our two observers' differing epistemic positions being both non-trivial and powerful when it comes to considering questions about black hole dynamics.

Consider the consequences of this non-reciprocal relationship between epistemic communities with this black hole driven example. Two identical twins (Alice and Brenda), with identical abilities to perceive the world, can be said to be in two separate epistemic communities if Alice can contact Brenda but Brenda cannot contact Alice, meaning any observation Brenda makes cannot be known by Alice. However, such a relationship does not have to be an equivalence one. Alice can be part of Brenda's epistemic community and not the reverse if Brenda has epistemic access to all that Alice does plus some extra area (Brenda, however, would not be part of Alice's epistemic community)[47]. Therefore, how many epistemic communities there are in a particular situation depends on one's epistemic judgment and position. This can be thought of by imagining a tree-topology, where in the example above Brenda's epistemic community sits at the top, with access to the greatest area of space-time, with Alice's epistemic community being non-reciprocally related to Brenda's, where she can pass Brenda information but not the inverse. This however is not the only structure two different epistemic communities can take. If we imagine Alice and Brenda both being within the black hole interior of two separate black holes, they both have access to their surroundings and the environment of the black hole exterior, but not the interior of the other's black hole. In this situation, although their epistemic communities overlap, one is not the subset of the other and they have permanently diverged. Being in a specific epistemic community is a changeable thing that can alter through time and location, and relationships between differing epistemic communities can only be evaluated via knowledge and understanding of what the epistemic communities contributing observers can observe.

We can consider this effect through time via the case of long-dead physicists. Isaac Newton is an entity with no epistemic abilities at present, as he is dead, a fact that also prevents present-day humans from communicating with him. However, present-day physicists still want to be able to refer to his observations and written works as they are both relevant and have genuine insight into how the physical world operates, so it would be helpful to be able to consider

[47] While these epistemic communities can "differ", they may not be "different". This relationship can be thought of as non-symmetric, in that A is in B but B is not in A, meaning while one person might have epistemic access to some region another doesn't, the person with the wider epistemic reach still possess all the epistemic reach of the person with the smaller community.

Newton as being within our epistemic community since his observations are epistemically available to us via testimony. Using the non-symmetric relationship described above, we can now say that Isaac Newton is part of our epistemic community, as we have access to said body of works that can be tested and retested by ourselves. However, we are not part of Isaac Newton's epistemic community because he does not have personal epistemic access to the modern world, modern physics, or modern ideas. This distinction allows us to maintain his physics as relevant to modern physics. It still holds the ability to enable reliable insight into current physical questions while avoiding tying someone dead to present epistemic actors as if they shared equivalent epistemic positions in the current year.

The motivation for this view of epistemic communities follows naturally from empiricist views of the world. What does it mean to observe, perceive, and experience in the solipsistic abstract? Without an ability to share, validate, and verify one's perceptions about the world, we cannot distinguish between seeing some event occurring in the world and a complex hallucination merely present in our heads. Even from a less sceptical perspective, this new understanding is useful. Science cannot be based solely on the testimony of one individual; it is by necessity a team exercise. We need rules that tell us what to listen to and what we can safely ignore to ensure we only get good testimony. This understanding of epistemic communities grants us these rules; if we follow them, we can trust that reported experiences are of the same kind as ours and that any information within our epistemic community is transferred only via causal links. Fundamentally, our epistemic abilities are not just related to what we can perceive but are regulated by those abilities present in our community. It, therefore, seems natural to say that if there is an unbridgeable communication gap between two parties, we must conclude that two different epistemic communities are present, even if the starting epistemic skills of both parties are identical.

How is this relevant to the question of black hole observability? Using this new understanding of epistemic communities and how they can differ through time and space, take an observer Alice to be outside a black hole event horizon, and Brenda is inside one. Alice would have epistemic access to the world area outside the black hole but no current ability to see what is inside the black hole due to the Principle of Equivalence. On the other hand, Brenda has access to information now lost to Alice, such as her mere existence inside the black hole event horizon and the areas Alice has access to outside the black hole. Alice is in Brenda's epistemic community as Brenda can receive testimony from Alice, but Brenda is no longer in Alice's epistemic community as Alice cannot receive information from Brenda. This means that for Brenda, the interior of the black hole is observable as within Brenda’s epistemic community, we

can perceive a whole new region of space-time that we could, if possible, communicate to any third party seeking to join us inside the event horizon.

For Alice, nothing has changed. What was unobservable before remains unobservable now, with the only change being Brenda no longer being present within Alice's epistemic community. The black hole remains "unobservable", but only due to Alice's epistemic community being unable to access the black hole interior. What Alice can be aware of, however, due to her strong understanding of constructivist epistemologies, is that the only reason she cannot "observe" the black hole interior is due to her spatiotemporal location; and that if she were to join Brenda inside the event horizon her epistemic community, and thus her ability to observe the black hole, would change. The epistemic community an observer is in is fully relativised to the spatiotemporal location in which one is present. Alice can be in the same epistemic community as Carla, 50 light-years away, but not in the epistemic community of Brenda, even if only meters separate them from one another. This demonstrates how powerful and important the event horizon is when considering one's epistemological position, as well as the fact that black holes present complex edge-cases for the philosopher of science.

As discussed earlier in this section, one can only talk reasonably about questions of observation if one identifies the epistemic communities of all those concerned at an early stage. In the case of black holes, our original problems have fallen away as the question "are black holes are observable?" is flawed. There is no unifying set of things that are "observable" and set of things that are "unobservable" for all of humanity. Instead, all that the epistemologist is left with is a set of rules that allow us to highlight what is potentially observable to one set of humans, at one time, in one place[48].

This is not a position that is based on philosophical considerations alone. Many examples with physics demonstrate that general relativity, quantum mechanics, and statistical mechanics point in this direction. In the section, we shall discuss these and highlight several examples where physicists have presupposed this conclusion and offered their own unwitting defences of the view that the ability to observe something is not just dependent on one's own abilities but also on one's position within the universe.

[48] This doesn't have to be things that we can only perceive with our eyes. The constructive empiricist can still consider things "observable" if they fit the criteria originally set out in constructive empiricism and are within an observer's epistemic community even if they haven't been observed before. A white dwarf star 70 light-years away from Earth across non-pathological space-time would still be "observable" under the framework as it is macroscopic and within the scientific community on Earth's epistemic community, as someone could visit it, see it with their own eyes and send testimony about it back to Earth about it.

4.1.4 A Defence from Physics

The concept of an epistemic community, though arising in van Fraassen's philosophical work, can be seen to be implicitly present in both the practice and methodology of physics. In fact, it is something built deeply into both the practice and methodology of modern physics. As an empirical science, physics relies on what an observer of a phenomenon can or cannot know about the system they are observing. This seems trivially obvious when considering older physics, e.g. Newton watching an apple falling from a tree and deriving the mathematical relationship that dictates the dynamics of its fall. Black hole physics, with the existence of the event horizon and the principle of equivalence, operates no differently, except that we are presented with a scenario containing *more than one* epistemic community (or equivalently a structure containing some number of non-reciprocal epistemic communities). In such a situation, our understanding of the system's dynamics becomes vastly less intuitive, as we cannot access one set of epistemic reference points by which to judge the success or failure of any given community. This is not to say good physics cannot be done. Instead, modern physics has adapted well to such a confusing landscape, pre-empting the challenges currently faced by the philosopher who wishes to study black hole physics.

In this section, we will discuss three different examples of modern black hole physics suggesting the existence of differing epistemic communities when discussing processes that occur universally across the black hole's interior and exterior or involve an observer moving from one to the other. These are offered outside of previously discussed work, such as black hole complementarity, to further strengthen the case that these hints come from across many different perspectives in black hole physics. The first involves a simple thought experiment using a Penrose diagram; the second is a more mathematical approach using Bekenstein's entropy, and the third an argument by Di Nunno and Matzer, showing how concepts such as a black holes interior's "volume" is contingent on the geometry one selects, again demonstrating that once again that when one enters a black hole, our black-hole-exterior viewpoint can lead to wildly non-intuitive and confusing results.

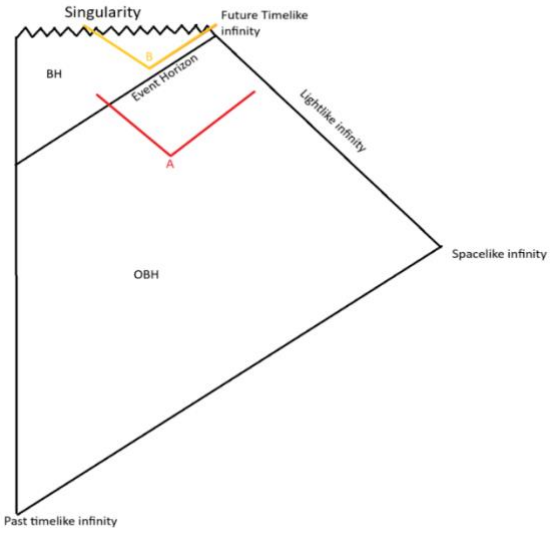

Figure 1 – Penrose diagram of the worldlines of two observers at points inside and outside of an event horizon

To address the first example, we can state that as someone inside an event horizon (and therefore within the region of the black hole interior) cannot return through the event horizon, they must have a world-line that that contains only spatiotemporal locations within the area of space-time bounded by the event horizon.

Therefore, their worldline has permanently diverged from all space-times external to the black holes. Figure 1 shows a Penrose diagram representing what happens to an observer's worldline before and after entering a black hole. For observer A, outside the black hole, we can see that observer A can move freely either to our lightlike infinity or enter the black hole, as would be expected. For observer B, the only accessible regions in their worldline fall within the black hole, there remaining no ability to re-exit the black hole and all possible futures ending at the singularity.

Observer A can move to be in both space-time regions represented here, the black hole region (BH) and the outside black hole region (OBH). Observer B only has access to the BH region. B's worldline has permanently diverged from all worldlines in the OBH region, and Observer B has access to information in the BH region that.

Observer A cannot access unless they join B inside the event horizon, at which point A would be restricted in the same way B currently is. All observers inside the event horizon have permanently diverged from worldlines outside the event horizon, lending credence to the notion that there could be something different about our ability to experience and report internal black hole space-times in an epistemic sense. This, again, helps lead us back to our concept of non-

reciprocal structure to generalised epistemic communities being present between an observer both inside and outside the event horizon.

This problem can also be seen via the differing physical quantities our two observers could experience (and potentially report on) due to their causal discontinuity, as shown via the mathematical descriptions we use to understand the properties of black holes. The entropy of black holes, concerns about which were once the source of seemingly paradoxical breaches of the Second Law of Thermodynamics, was given a generalized form by Bekenstein and Hawking with both equations for the entropy of black holes themselves (1) (Bekenstein, 1973, p. 2338) and a generalized form of the Second Law of Thermodynamics (2) (Bekenstein, 1973, p. 2339), given as:

$$S_{bh} = \left(\frac{1}{2}\frac{ln2}{4\pi}\right) kc^3\hbar^{-1}G^{-1}A, \tag{24}$$

and,

$$S_{bh} + \Delta S_c = (\Delta S_{bh} + \Delta S_c) > 0, \tag{25}$$

respectively (where S_{bh} is the entropy of a black hole, ΔS_{bh} is the entropy change of a black hole over time, ΔS_c is the entropy change of the region outside of the black hole; k, c, $\hbar$, G are constants and A is the surface area of the black hole, measured via the size of the event horizon). From these results, it is trivially obvious that in (1) $S_{bh} \propto A$, so that the larger the observed event horizon of a black hole is, the larger its reported entropy will be. Equally, we can say that (2) implies the monotonic increase of universal entropy as assessed by any observer, keeping us in line with the Second Law. However, we are still left wondering what it means to measure the entropy of a black hole from *inside* the event horizon. The underlying assumption made when using the surface area is that information within said area is inaccessible, and given entropy is in some ways a measure of information loss regarding an individual observer, the connection between the inaccessible area within a black hole and the entropy an external observer would measure for it is evident. For an individual travelling into a supermassive black hole, there can be a non-negligible time between passing through the event horizon and being killed, meaning this observer can suddenly see, observe, and experience things otherwise thought inaccessible to human beings outside the black hole. Given that our prior assumptions about the immediate death of the infalling observer do not now apply and they should be able to perform some form of observation and calculation before their death, how can we mathematically express the entropy reported by the infalling observer? A plausible approach would be to reduce our surface area A by the area within the light cone we would now have

access to having moved a distance into the black hole, which we can defend mathematically via the implicit degrees of freedom hidden to an external observer behind the black hole that the Bekenstein entropy implies with $S_{bh} \propto A$.

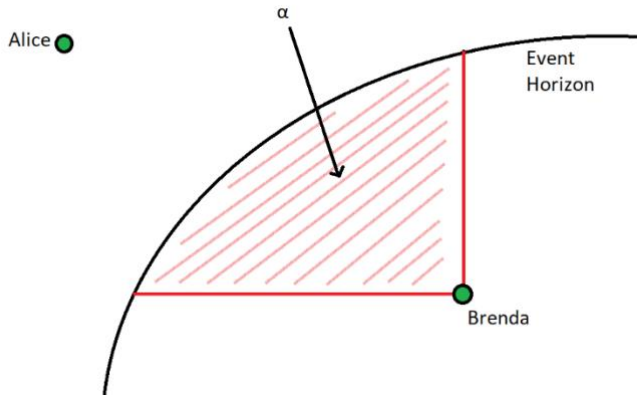

Figure 2 - A 2D representation of a black hole showing the position of an external observer (Alice), and internal observer (Brenda). The hatched area represents in a 2D plane the new area, α, available to Brenda once inside the black hole that is not available to Alice. This area is bounded by Brenda's worldline as she falls into the black hole.

This quantity, given by a, gives us the area of the black hole that we could theoretically access, which this chapter will define as the area bounded by the past worldline of an infalling observer, as opposed to the entire black hole interior itself (which may not be accessible to us given the nature of lightlike paths further into a black hole, which should not be able to transfer information outward), which is shown in Figure 2. This holds some *prima facie* merit; as if we consider entropy (as measured by an observer) negatively proportional to the information that observer has access to, having access to more information means lower entropy. The A, which represents the black hole's surface area, can also represent the area that an observer outside the black hole cannot observe[49]. It therefore follows that if an observer can now observe some part of an area that he previously could not (area A), we should adjust that area to take this into account, allowing for the expected entropy adjustment. Altering (24), we now have:

$$S_{bh} = \left(\frac{1}{2}\frac{ln2}{4\pi}\right) kc^3\hbar^{-1}G^{-1}(A-a), \tag{26}$$

Where a is the area of the black hole interior we can now access. If taken at face value that (3) is a correct description of black hole internal physics[50] (say via using a model akin to the

[49] Bekenstein makes this connection in his work on black hole entropy, noting it as interpretation of entropy that allows us to recover the thermodynamic properties of black holes. Most literature refers to A as simply the surface area of the black hole as in most situations the descriptions are equivalent however, as this example clarifies, this is not always the case. In fact, recent work by *Malcedena et al.* explicitly uses the concept of an " entropy island" inside the black hole to solve the entropy related Page-time paradox, making areas of the black hole, in effect, not part of the black hole when it comes to entropy calculations.

[50] A note of caution: there is no good reason to suppose this is correct, beyond priors about entropy being conceptually related to information available to the observer and black hole external physics. There is no good way of proving (26) is correct or wrong, it is unfalsifiable given current experimental techniques.

Holographic Principle, where microscopic degrees of freedom are hidden in the black hole event horizon to external observers, but available to an infalling one), the entropy of the black hole for the observer falling through it should be decreasing, with a corollary decrease for the entropy of the entire universe, for no work done by the observer. This, evidently, would break the Second Law, so all things being equal demonstrates an issue with applying external black hole physical theory to the black hole interior.

It is further worth exploring the issues with (26). While we have a complete conceptual reason for equation (26), this being centred around how an infalling observer would perceive the entropy of the black hole, it is still just a mere conjecture regarding what forms black hole entropy. We can reduce (26) down to $S_{bh} \propto A$, holding all the constants equal across scenarios both inside and outside the event horizon, but on a grander level, we are questioning the very validity of physics for black hole exteriors when applied to black hole interiors. An observer inside the event horizon might find a region with completely different physics, which stays within the bounds of the Second Law even if our conceptual case is correct. Then, the point of (26) is less to show some new physics in and of itself, but to demonstrate that our current physics can have apparent conceptual inconsistencies when applied to realms epistemically cut adrift from us. The issue highlighted here, that epistemic issues lie at the core of black hole paradoxes has also been the basis for many other "paradoxes" of black hole physics, such as the Page-time paradox (Wallace, 2017, pp. 16-17)

Further to this, equation (26) does highlight a necessary issue when conceptualising the realm of space-time inside the event horizon. There are potentially no good models or equations, despite the introduction of equation (26) within this piece, because we cannot be sure our counterfactual assumptions that hold for the rest of physics still hold true. As shown earlier, we have no accounts of what there *is* on the other side for firmly established epistemic reasons. The effective agnosticism we must take forward regarding the realism of our physics in these areas hamstring our attempts to describe, conceptualise or elucidate what is going on cleanly. However, these areas are still parts of our universe, with space-times that *should* obey the laws of physics and obvious causal relationships to areas outside them (things outside the event horizon get pulled into the event horizon by the singularity inside the black hole, a clear causal relationship).

How do we square this circle? By using our previously developed concept that the epistemic community of an observer can change depending on that observer's ability to transmit information to other observers, we can resolve this problem. Our infalling observer's differing

understanding of the black hole's entropy isn't some break within the physics or some inconsistency in the work of Bekenstein and Hawking; it is simply a changing in the epistemic status of the observer, which cannot be accounted for by any possible physics[51]. Rather than being some interesting paradox that should encourage future work, it is an effect of humanity's inability to be a super-observer. If humans remain epistemically limited, physics will never have the full ability to explain these changes in an observer's epistemic community, as trying to do so goes beyond the scope of physics as a subject. This can be seen in the ultimately non-consequential "paradoxes" that are present within modern black hole physics. The key benefit of this new understanding of constructive empiricist observation and the centralising of epistemic communities is that it allows van Fraassen and other constructive empiricists the latitude to be agnostic about questions that concern the black hole interior whilst maintaining that their philosophy of science is complete. The constructive empiricist could simply state when presented with such a paradox "We cannot know, within our epistemic community, anything about the black hole interior, as physics and the philosophy of science is limited fundamentally to what we can perceive. But neither are we, from outside of black holes, required to produce theories that get things right about black hole interiors, since the aim of science is to get things right about what is observable to us, and for an epistemic community located outside of a black hole, this excludes black hole interiors".

A third example within pre-existing physics literature where we can see non-intuitive results is expressed by Di Nunno and Mazter (Nunno & Matzer, 2009) in their paper "The Volume Inside a Black Hole", making the case that the volume an infalling observer will experience is dependent on the coordinate system used, and can even be zero. They take their definition of volume to be "the three-dimensional spatial content [of a black hole] at one instant of time", which when compared across different co-ordinate geometries (Kruskal-Szekeres, Schwarzschild, Kerr) produces non-equivalent mappings for what the inside of a black hole is spatially. For example, inside a black hole under the Schwarzschild metric, they find "*There is zero volume inside the black hole in any Schwarzschild time slice of a Schwarzschild black hole space-time*. This surprising result arises from the specific definition of $t = constant$ in the Schwarzschild coordinate system" (Nunno & Matzer, 2009, p. 1) [italics theirs] (this result potentially highlights the lack of usefulness of the Schwarzschild metric when discussing black hole internal geometry).

[51] Black hole complementarity, the "solution" to the Page-time paradox introduced in Chapter 1, gets close to this understanding by divorcing the perspective of external and internal observers of a black hole system entirely, but it doesn't quite extend fully to it.

For a more appropriate coordinate system, such as Kruskal-Szekeres, we find a result that changes as the observer falls inward, with a discontinuity at $v = 1$ as their spatial slice interacts with the singularity, destroying the infalling observer. While their conclusions are limited by the fact the geometries themselves are under discussion, as opposed to the explicit physical structure of the black hole, again we have another instance where, past the event horizon we discover a situation that is both non-intuitive and seemingly at odds with our everyday understanding of the world. Rather than this just being an effect of more complex physics, be that relativistic or quantum, the breakdown of very basic concepts such as "volume" speak to a greater shift in paradigm. When an observer enters a new epistemic community, we should expect all of our pre-supposed physics understandings of reality to be under threat, as our entire understanding of the universe has, thus far, been derived by one epistemic community operating on earth within the same framework of epistemic skills. This Di Nunno and Matzer paper demonstrates that when an observer does enter a different epistemic community, even basic concepts such as volume cannot be understood for them by an external observer. This is because nobody writing, modelling, or experimenting on black holes here today on Earth can have any epistemic experience of a black hole interior. Black hole interiors are a deeply complex area of space-time that potentially physics alone cannot address fully, with even the productive areas of black hole physics struggling with paradoxes concerning concepts where information re-emerges from the black hole interior. From our epistemic standpoint, it is hard to see how we can know anything about a black hole's interior.

With physics repeatedly demonstrating the need for both a greater focus on epistemic concerns within modern philosophy of science and the usefulness of epistemic communities as a starting point for such a focus, where does our constructive empiricist now stand? It seems clear that the black hole exterior is observable to those of us on earth, both in a naïve sense of it being possible to see the orbit of stars around the black holes, as well as the ability to maintain a single epistemic community that can see all such macroscopic events. The black hole interior, on the other hand, must be unobservable, as whilst it is possible for a human to leave earth on a spaceship, enter a black hole and "see" the region inside of the event horizon; by doing so, they permanently join a different epistemic community to those who have remained outside the black hole. As a result, the constructive empiricist does not have to maintain empirical adequacy about theories focused solely on objects inside the event horizon and can instead treat these potentially macroscopic objects as one would protons or bacteria. This also would mean that van Fraassen's heuristic on observation, whilst successful for non-relativistic questions, is not entirely adequate. I, instead, propose this as a better heuristic: X is

observable to us if, it's possible for X to be present to a member of our epistemic community, and were X present to such an individual, then that individual would observe X.

What this means for our working black hole physicist is evident. Our paradoxes are constructed within the framework of some "physics" that is deeply unintuitive and abstract as to point us already in the direction that something isn't right when attempting to view it from a realist standpoint. constructive empiricism allows for is a consistent understanding of *why* that occurs (the practice of physics has progressed beyond what the realist can be comfortable with) and *how* to solve for it (by accepting there are some areas even of our universe, namely black hole interiors, that are beyond our epistemic reach and therefore outside the grasp of solely physics). Our working black hole physicist has both the problem and solution outlined for them: adopt a constructive empiricism as a philosophy of science and the empiricist stance and reap the rewards of being able to spend their finite research time doing work with higher chances of theoretical and practical success. The next section, ***The Page-time paradox: Revisited***, tests this belief against the Page-time paradox and shows why the constructive empiricist case is superior to the realist one.

4.2.1 The Page-Time Paradox – Revisited

The Page-time paradox has been outlined on multiple occasions throughout this thesis therefore, to save the readers time, we will not retread old ground in this section[52]. Instead, the aim of this section is to take what we've established about some paradoxes in black hole physics and achieve two things. The first is to take constructive empiricism and the developments of this chapter to help show how to exorcise these paradoxes by showing their consequences don't generate. The second is to argue firmly that, in the context of black hole physics, constructive empiricism is the most attractive framework for the working physicist. Constructive empiricism either offers solutions to issues generated by the Page-time paradox that realism cannot or constructive empiricism better accounts for the epistemic modesty required to accept cutting edge theories, even if the realist could *prima facie* view them as compatible with their metaphysical desires.

As we do this, we can keep in mind our pathways highlighted in Chapter 2, section 2.2.5. Our working black hole paradox physicist, when faced with the Page-time paradox, has in my mind three different pathways to go down:

[52] If the reader wants a refresh on it, Chapter 1 has a description of the paradox.

1) Wallace is wrong, this isn't a true paradox and one of either statistical mechanics or QFT don't have "enormous plausibility", one will be shown to be evidently superior empirically.
2) Empiricism cannot solve this paradox, and our entire framework for physics is wrong. This is an extension of 1), with more dire consequences, as it means abandoning entirely the substructure that underpins physics. Entirely new frameworks are required for studying black holes.
3) Empiricism is good, but our epistemic community can no longer do good physics here, at least whilst being realists. We on Earth have surpassed our abilities to do empirical work along the way whilst researching black holes, and these problems can no longer be solved by physics and physics alone.

Now that we have both our constructive empiricist framework and my novel extension of epistemic communities, can we find pathways to solving the Page-time paradox? In this section, I offer a set of possible thoughts a working physicist could adopt in solving or ignoring the Page-time paradox which fit within the 3rd pathway, the 1st being unlikely to be correct given the sheer weight of evidence showing that Wallace is right and the 2nd being deeply uncomfortable for any physicist, who wishes to believe the theoretical ground beneath their feet is solid and not want to overwrite theories that have empirical backing because of one, empirically as yet unobserved, paradox[53].

In and offering these thoughts, this section will be focusing on one core aspect: our epistemic access to the theoretical systems causing this paradox. Black holes at the Page-time are much like Maxwell's demon; theoretically described and the consistent outcome of pre-existing theoretically supposition, but not an entity in the universe that we have yet had access to experimentally. Secondly, given the nature of these black hole systems being inherently related to the dynamics of the black hole interior, we have good reason to believe that they are beyond the understanding or even the scope of modern physics. Taken together, they form a convincing case that we can either simply ignore the Page-time paradox as working physics *as things currently stand* or accept our currently proposed solutions without concerns about metaphysical implications.

Let us start again with a brief description of the black hole interior and how the region should operate. The black hole interior is the region of space-time bounded by the event horizon, which has at its centre the singularity. The event horizon is a spherical "boundary" where the infalling

gravitational pull of the black hole exceeds the possible acceleration of an outgoing photon, meaning it becomes the "point of no return" for matter. This black hole interior is absolutely a region of space-time. Via the Principle of Equivalence and other aspects of General Relativity, an infalling observer should have no ability to discern when they have passed it from their frame of reference, and the area contained within the event horizon (e.g. black hole interior) can change with the mass of the black hole. Areas inside can become outside and *vice versa* with no way to discern that is the case from an observer passing through the event horizon's reference frame. This region of space-time has no reason to be distinct from other regions *a priori*, therefore any set that includes "the universe" should include the black hole interior, potentially up to and including the singularity. The epistemic gap produced by the black hole interior is not a fundamental property of the space-time in which the black hole interior is present, but an effect of the physics that creates the black hole interior *in that space-time region*. A realist should have no reason to believe that physics can't, in theory, provide a fully ontological description of the black hole interior, as it theoretically can the Earth, the Moon or the Sun, if we can generate theories in physics about the region.

In attempting to find pathways to resolve the paradoxical consequences of the Page-curve using a constructive empiricist framework, I will attack the problem from two distinct directions. Firstly, we will consider the reasonableness of observing a post-Page-time black hole within the heuristic of observability provided to us via van Fraassen and make a pragmatic case that our working physicist could feel reasonable in extending beyond van Fraassen. Secondly, I will look again at black hole complementarity and firewalls, two competing theories that offer very different accounts about what happens after the Page-time is reached, showing that the constructive empiricist doesn't have to worry about the mutual inconsistency between the competing theories in a way a realist would. Thirdly, we will talk about very recent developments in black hole physics that, whilst giving us giving solutions to the Page-time paradox that are compatible with a realist understanding of science, hint again at the core argument of this Chapter: black hole interiors are unobservable, meaning any physics-based solution to any information loss paradox will only be found in the dynamics of the black hole exterior. This case is natural for the constructive empiricist but is problematic for the realist. Busso *et al*'s (Busso & Penington, 2023) work should be seen as very exciting for our working physicist, but it again highlights the problems the realist has in trying to have their cake and eat it to. The current evidence in physics, and the progression of research, demonstrates that if you want to solve black hole information loss paradoxes, you must eschew a realist mindset at fear of being internally inconsistent or arbitrarily dismissive of the clues given to us by physics.

To begin, let us first note the fact that post-Page Time black holes would seem to be rare beings. The Page-time is only hit by a black hole when half of its mass is evaporated away by Hawking Radiation[54], and involve the 'evaporation' of matter with whilst none enters the black hole in return, which would further delay the process of hitting the Page-Time. Such an object would seem only the preserve of modelling in and of itself, and the lack of literature covering the core concept of "does an astronomical/macroscopic black hole that has evolved past the Page-time actually exist?" is remarkably lacking. Therefore, I present an argument here that the reader can judge the reasonableness of themselves, which is similar in construction to the arguments made the physical reasonableness of Maxwell's Demon in section 3.2.5.

If we take our analysis of physics to be predicated primarily in empirical adequacy of the observables, we must want black holes to be analysed to the same standard we would wish to analyse other macroscopic entities that effect the observable world around us. Black holes, if they operate under the same physics, must be as analysable as the Sun, the Moon, or the other stars present within our arm of the Galaxy. All of these objects operate in the real world (layperson's understanding and language usage), i.e. they are affected by the systems around them. The Earth orbits the Sun, which orbits the Milky Way, which is moving relative to other galaxies within the universe. Astronomical black holes operate under the same conditions. The images we have of black holes are not simply black regions, but are existent in the context of the universe around them that allows them to be visible, such as gravitational lensing giving them a corona (Event Horizon Telescope, 2022)[55]. I contend that it is a feature of the universe that we have yet to observe a post Page-time black hole visually: there is too much matter in the universe that is too likely to fall into any one we could "see", and we on Earth are too far away from them to actually send a spacecraft up to find and observe one. Whilst the constructive empiricist wishes to offer and account of how physics is, rather than how it should be, and therefore drawing a distinction between the observed and observable "fails to capture our idea of what it is to do good science" (Monton & van Fraassen, 2003, p. 407), a heuristic understanding of what is potentially observable within human life-times but as yet unobserved

[54] Other models put this as 1/3rd of the mass, c.f. (Gautason, et al., 2020), but for the sake of this argument the difference is fundamentally arbitrary.

[55] To be clear, the colouring of the corona is not a property of the corona itself, but a rendering choice made by the physicist when representing the corona as a photograph. Questions of direct or indirect observation are relevant here, as whether this should even count as observation. This especially true for a constructive empiricist, but modern physics considers this a photo and uses this terminology, so I will concede solely for the following argument that this is some form of "observation", as I maintain even if it is, it doesn't override the following point.

versus an object that is so far beyond the practically observable or even practically existent[56]. If we can be so doubtful as to the relevant existence to physics of an object that is currently only the preserve of toy modelling that cannot hope to account for all variables, is it worth worrying about any potential physical paradoxes that can generate?

This argument is not meant to be firmly conclusive. As stated, it is an argument that relies on the reasonableness of the assumptions I present; that post-Page-time black holes are so rare and unlikely to form in our universe as to be only worthy of merit as interesting models, not as physical objects. This even goes beyond van Fraassen's understanding of what is observable, but here I maintain that there are so many factors that could stop the generation of the observation, even if the object is technically observable but not yet observed, that saying "this object cannot be observed", whilst technically wrong, is a reasonable conclusion. Note also here that this applies only to post-Page-time black holes, for which the setup involves extremely long time-scales and extreme isolation, not any other form of conventional black hole. Further to this point, our concerns about the observability of a post-Page-time black hole only generate when we consider their black hole exteriors. As we have firmly established in this chapter, black hole interiors (including post-Page-time black hole interiors) must be unobservable due to the epistemic consequences of the event horizon.

Even if we were to find and observe a post-Page-time black hole, metaphysical issues generate for the realist that never trouble the constructive empiricist. Let's assume for the sake of the argument that black hole complementarity and firewalls are both equally plausible consequences of the Page-time paradox, specifically what happens to the relationship between entropy and information as we hit the maxima of the Page-curve. In terms of what an external observer would see, there is no distinction between them. As we see in section 1.1.5 and 1.1.6, on both accountings the infalling individual would be thermalised at the event horizon from an external observer's reference frame. Despite being mutually incompatible theories, they both predict that an external observer would perceive the same thing. For the constructive empiricist, this quirk isn't a problem: they can simply ignore their different predictions about the operation of the black hole interior as it isn't observable and our theories don't have to be empirically adequate about them, accept both theories agree about the observables, and ignore the paradox they create. The realist cannot do this. They are forced by

[56] It behoves us, however, not to make this case too strongly: the universe is a very large place and has existed over a very long time, a post-Page-time black hole could really be existent in the universe, as physics describe it, without us knowing where it is currently. The following paragraphs will deal with this eventuality, but I note it here for the reader.

metaphysical necessity to choose which one is at least approximately "true", as both cannot be at the same time, and both propose radically opposed reasons for what we'd observe if we enter the black hole exterior (a region the realist would maintain is observable). Extending this further, as both theories are empirically adequate to the external observer, who is in the same epistemic community as us[57], for the constructive empiricist the paradox never generates in the first place. Only the realist must deal with the consequences of any paradox. In doing this, the constructive empiricist can resolve the Page-time paradox in a way the realist cannot, and may never be able to, demonstrating the superiority of the constructive empiricist case.

Even if you don't accept this argument as one the constructive empiricist could make without begging the question, as anyone with a voluntarist epistemology might, you can accept it *pragmatically*. The debate between two theories "truth" which, for the constructive empiricist, may amount to nothing about the observables, has taken many hours of work and countless words to elucidate and describe. Whilst belief in either black hole complementarity or firewalls could be *rational*, it may not be *useful,* as they are directly hindering the process of physics merely because they exist and are mutually incompatible. This truly highlights the attractiveness of the constructive empiricist approach over that of the realist: they don't have to care anymore about the paradox. In section 2.3.1 we discuss whether scientists should be anti-realists or realists on the nature of what benefits either school can offer to science. I maintain that constructive empiricism can offer pragmatic benefits to physicists in this area, helping them work more progressively, in a way that realism can't.

To further show the strength of constructive empiricism over realism in black hole physics, we can provide the best possible circumstances for the realist. To do so, we can reintroduce the work of (Almheiri, et al., 2020a) mentioned in Chapter 3, and introduce the work of (Busso & Penington, 2023), both of whom are interested in "entropy islands"[58]. I will show that even by doing this, we are still left in a position where constructive empiricism is the superior philosophy of science, albeit because the work of physics presupposes concepts that align with the constructive empiricist framework. Recent work revolves around "entropy islands", pockets of entanglement information created on the particle-antiparticle process central to

[57] But not the same epistemic community as the infalling observer after they pass the event horizon.
[58] I introduced these very briefly in Chapter 3, and this thesis has considered them as potentially not even relevant to the constructive empiricist due to their very recent introduction to the literature as a concept and their silence on *what* form the potential information recovery takes. To be relevant to the constructive empiricist, further research must be done on this topic. We have returned to them here anyway to further strengthen the case that the constructive empiricist viewpoint is *always* superior to the realist viewpoint regarding entropy paradoxes, as well as offer a scenario that should be tailored made to help the realist: a solution we can theoretically test experimentally.

Hawking Radiation. Relevant papers include (Almheiri, et al., 2020a), (2020b), and (Busso & Penington, 2023) (pre-print) where these islands are claimed to either extend beyond the event horizon at sub-Planck length scales or, intriguingly, in recent work by Busso, up to an atoms length outside the event horizon. These arguments are interesting, and in effect allow for a solution of the Page-time paradox in the only way that both the realist and the constructive empiricist may be happy with; by information being available and recoverable outside of the event horizon. The dynamics of these modelled and mathematical approaches are still limited by the issue that affects black hole complementarity or firewall approaches: they still have not been shown empirically to be correct. However, where they differ is that they offer us pathways to do experimentation in a way that could actively give us answers about which of the QFT or statistical-mechanical approaches are better. Busso et al.'s paper is of particular interest, as it implies the ability to create a physically realisable experiment that could test which of the firewall or complementarity approaches are more likely to be right. To quote him directly from UC Berkeley's article on the paper:

"We show these islands actually protrude beyond the horizon of the black hole far enough that, in principle, there is no obstruction to probing them and coming back out… That's actually pretty dramatic because it means that there's some extremely surprising and radical new physics that is no longer hidden behind black hole horizons or hits you when you try to jump into a black hole, but which is, in principle, accessible to us." (Wilkins, 2024)

Here we have a pathway that should bring joy to the realist. Instead of having to wrestle with questions about how we access information about the black hole interior, physics has progressed and found an answer. If we can acquire empirical evidence for this theory, we don't have to accept constructive empiricist position as inevitable or even pragmatic: we can do physics about black hole interiors, we have solved information loss paradoxes like the Page-time paradox, we can rest content that realism is our best operating philosophy of science.

I maintain this is not the correct read. Beyond pragmatic experimental concerns still going unanswered (where are the post-Page-time black holes? Will we ever advance technologically to the point Busso's hoped experimentation can occur?), entropy islands extending outside the event horizon concedes hugely important ground to the constructive empiricist's position: that the *region* of the black hole interior is inaccessible. The realist still must square the circle of what is so pernicious about the spatiotemporal region inside a black hole, which the constructive empiricist doesn't. Given the constructive empiricist argument for epistemic modesty still holds and they can accept this progress in physics as well as the realist can (if it's

empirically adequate about observables), it remains more attractive to be a constructive empiricist, and our working physicists should still become them. Physics keeps hinting that the constructive empiricist understanding of what is observable is correct, why should we abandon the epistemic modesty that motivates it for metaphysical headaches?

Of all the approaches I outlined for solving the Page-time paradox by the working physicist in section 2.2.5, pathway 3) still seems the most likely to be correct. All solutions to the Page-time paradox, and all related information loss paradoxes, revolve at least around the dynamics of the black hole interior. If that interior is not accessible to us, any solutions that require its dynamics to be right about the black hole interior will inevitably suffer from our inability to breach the epistemic gap that must form at the black hole event horizon and actually test their validity. Even modern work showing that these islands can extend to regions outside the black hole, potentially giving us access again to the lost information in the black hole interior, are predicated on a physically coherent description of the black hole interior. In short, our lack of ability to *test* the black hole interior will always play a role in what we can or cannot accept as pure physics. This chapter demonstrates that the black hole interior spatiotemporally is beyond the reach of physical experimentation, and we have no reason to believe otherwise due to the empirical adequacy of the theories that underpin that[59]. Even assuming the best possible scenario, that we've found a past-the-Page-time black hole and successfully entered some observer via technology that doesn't disturb the Page curve (which we may do by adding mass back into the black hole that must be radiated away) or who can survive in one until the Page-time is once again reached and who's measurements about the region we can eventually acquire via exotic physics, questions about the black hole interior still remain.

Leaving aside the developments from Almheiri *etc al.* and Busso *et al.,* all possible solutions to information loss paradoxes such as the Page-time paradox suffer the same problem: the epistemic gap at the black hole horizon is still there. Two epistemic communities now exist, one within the other, due to the non-reciprocal way those on Earth and those within the black hole can communicate. If we are to say that the study of physics is dependent on everyone being members of a single epistemic community, it follows that we can no longer say we have just one physics. Instead, two distinct epistemic communities now exist (even if one is the subset of another). For those on Earth, who will always remain on Earth and outside of any given black holes, the epistemic community split that necessarily forms. To say that the physics of the

[59] For concerns of circularity here, where we are using physics theory to outline what methodological understanding of physics is appropriate, see section 2.3.3.

black hole exterior is the same as the black hole interior seems at best a hopeful proposition, at worst an unknowable conjecture.

Taken all together, I see two only two possible reasonable conclusions, which are either optimistic or pessimistic about the most recent developments. Both are reasonable, and within both I maintain it is superior for the working physicist to adopt constructive empiricism over realism.

Pathways for the Working Physicist:

a) We remain pessimistic about solutions to information loss paradoxes such as the Page-time paradox and adopt the tenets of pathway 3). The entire scope of the problem and their theories have extended far beyond what we can now reasonably call physics. The systems and objects within them are either too idealised to reflect the universe as it is constructed or too dependent on things outside of our epistemic abilities. We are better off simply not trying to solve these problems, as these combined factors will simply lead us to create more. We cannot hope to be realists about any of our solutions as "physics" can no longer do the work required of it and must therefore become constructive empiricists. Either black hole complementarity or firewalls being "correct" doesn't matter, as long as both achieve empirical adequacy about the black hole exterior. Even if we allow for the rational of choosing either black hole complementarity or firewalls, we are better placed as physicists to work productively by accepting the gift constructive empiricism offers us: wilful ignorance.
b) The modern solutions to information loss paradoxes are good and we should hope in their success. The paper from Busso et al. presents a pathway to forming a testable experiment (even if it's not possible to do such an experiment currently) where we can show through physics which, if any, theory about a post-Page-time black hole is empirically accurate. Although this solution is compatible with realism, I have argued that epistemic modesty should lead us to be wary of inferring the truth of a theory concerning black hole interiors even if one can be found that is compatible with our current best physics. The strength of the constructive empiricist case isn't found in stating that realism is incompatible with the most modern work. Instead, constructive empiricism doesn't have to concern itself with any future ontological and metaphysical concerns that being realist must because they are epistemically immodest. Even for

Busso, the *region*[60] of the black hole interior is still inaccessible empirically to humanity, despite offering a pathway by which we can recover information about it. They can accept Busso's solution to the information loss paradox and remain realists, but at the cost of conceding to a view that comes naturally to the constructive empiricist that they must explain within the framework of realism: black hole interiors *are* unobservable and different to other macroscopic regions of the universe. Thus, it is superior to simply be a constructive empiricist, as it provides more stable future ground that is appropriately epistemically modest.

The argument for defending **Pathway for the Working Physicist:** a) can be presented this way: Even on the most optimistic case, our infalling observer is outside the realm of our epistemic community, which a realist shouldn't have an issue with accepting. Theoretical work about the black hole interior, as it can't be tested within our epistemic community, is no longer "physics" alone as understanding what "physics" is requires understanding that it based on our, common, epistemic community. Instead, the work can be conceived of as being rationalist extensions of concepts of physics outside the realm they can be directly applied, something that is more akin to the work of analytical philosophy. This is no bad thing in and of itself, but as the work has strayed beyond the scope of what is empirical, it can no longer be considered solely *physics*. On the other hand, our constructive empiricist is perfectly comfortable with all of this, as they don't require metaphysical content to explain the nature of what science is. For the constructive empiricist, good physics requires only empirical adequacy, which we have achieved for black holes. Thus, the paradox never generates in the first place. We have the option of not worrying if one set of equations is mutually incompatible with another at esoteric boundary condition. It functionally doesn't matter if our theories agree about what is observable. If our current theories agree in what they say about black hole exteriors (the part of a black hole that is observable), and if our aim is to have theories that are empirically adequate, we do not need to worry about their disagreements. If these disagreements solely concern what's happening with the unobservable (to our epistemic community) interior of a black hole, as constructive empiricists we can happily state there is no issue. Thus, the paradox never generates.

[60] Important to note here we are talking about the space-time region itself of the black hole interior, not the information present in the black hole interior. Entropy islands are offered as a way of information being recovered from the black hole interior, but we, as humans, still don't have access to the black hole interior in the manner we'd want to, for example, demonstrate that black hole interiors and exteriors are part of one singular epistemic community.

Beyond this argument, even more strength can be found via Boucher and Forbes' understanding of the pragmatic scientific realism debate (Boucher & Forbes, 2024). Even if there is no irrationality in the act of accepting one of black hole complementarity or firewalls over the other, the very act of doing so is less *pragmatic* than not caring which is the correct theory. Any act of theory acceptance about black hole interiors leads one to epistemically immodest conclusions, which themselves demand defences, which can't be provided for because the region that is being discussed is outside our epistemic community. It is vastly more attractive to simply assert "an infalling observer gets thermalised at the event horizon" and not worry about which model for how that occurs is correct. Humanity will never be able to know, so why bother looking?

The framework for understanding our **Pathway for the Working Physicist:** b) can be thought of thus: If we take physics to be a necessarily empirical subject, as we do, whilst we may be able to say any and all of the black hole information loss paradoxes are "resolved" by the introduction of thoughts such as entropy islands or replica wormholes, we still cannot do it in a way that can make the realist fully happy. Similarly to pathway a), we still generate the epistemological consequences of committing ourselves ontologically to the region of the black hole interior. For the realist, we are required be talking at least approximately truthfully about all aspects of our theories for us to want to accept them. Scientific realism is most fundamentally the claim that our theories are approximately true about the universe. Even if we put a willing realist into space and sent them through the event horizon, whilst we may be able to recover the information they extract, that won't grant us, humanity's epistemic community, physical access to that space[61]. In the very best-case scenario that we can produce for the realist, they are still left with headaches and problems to solve. For the constructive empiricist, however, we can either fall back to the pragmatic conclusion offered in **Pathway for the Working Physicist**: a) or, even accepting information returning to the black hole exterior via entropy island (something we don't have to do given potential worries about its success/newness), be happy that physics has again reconfirmed the tenets of the constructive empiricist approach: that the region of the black hole interior remains inaccessible to our epistemic community,

[61] Busso's work does indicate that a pathway for recovering information *may* be achievable outside the event horizon on long enough timescales in idealised enough conditions, but as is stated when talking about these forms of solutions in Chapter 3 and at the start of this section, any information recovered may still not be considered valid for observation for the constructive empiricist. For example, the breaking of continuity implied by going into the black hole interior in the first place may add 'chain and of custody' issues amongst others. Until more work is done on the validity and physicality of "entropy islands", conclusions are hard to draw on this front, so I have left such analyses outside the scope of this thesis.

which consists of those on Earth who can communicate. The realist must give reasons why the region of a black hole interior operates differently epistemically to other regions of space-time, for example that which includes the Sun. The constructive empiricist does not.

This argument can also be presented this way: what is being argued, especially by Busso *et al.*, is that the atoms-length extension of entropy islands beyond the event horizon give us a way to investigate the black hole interior in a way we previously didn't have. The constructive empiricist could ask reasonably of the realist, "does this *actually* teach us anything about the dynamics of the black hole interior, or does it simply avoid the black hole interior entirely by moving the dynamics to the black hole exterior?". If the answer to this is concede the latter path[62], which I think it does, I maintain the realist is still left holding the bag of an immodest ontology. Let us not forget that the black hole interior is a *moveable*, *non-pernicious* area of space-time topologically[63]. It is not a singularity; there are no problematic infinities at play. The realist should want to be able to ontologically commit themselves to theories about its dynamics in the same way they would, for example, the Sun. Any observability of black holes interiors, given we still have no reason to think we could go through the event horizon and come out again, as humans, must come at a remove from the direct empiricism of our unaided observations. I maintain it this is unsustainable. When CERN smashes sub-atomic particles together in the LHC, the "unobservable" (for the constructive empiricist) is at least conceptually connected to the world we have epistemic access to through our senses, from sub-atomic particles, to atoms, to molecules and then to the macroscopic structure of our existence. Can we make the same case for our observations of the black hole interior via entropy islands, which directly reconfirms the inaccessibility of the black hole interior *to us*? I believe the realist can't make this case, and that epistemic immodesty is at the root of their headaches. The constructive empiricist doesn't have this ongoing question to answer, so even under this scenario, where one could maintain realism, constructive empiricism is still more attractive as it doesn't commit oneself to epistemic immodesty.

Both approaches rely on the case being successfully made that black hole interiors are unobservable as we currently perceive them. As this observability is in and of itself a theory-laden notion, it can change as our physics theories do. I personally sway closer to pathway b)

[62] This is itself potentially a difficult to answer questions: the physical instantiation of an entropy island has yet to be found or conceptualised. A realist could just shrug their shoulders and say "I don't know", but this doesn't offer much philosophically to the debate.

[63] As best theory tells us. Almost all these conceptions are in and of themselves theory-laden. Again, for questions on circularity and the problems it does or doesn't present to the constructive empiricist, see section 2.3.3.

on this basis. I remain optimistic about the ability of physics to correctly describe the accessible universe around us, believe theoretical evidence that we do follow the Page-curve is substantial, and firmly adopt the constructive empiricist commitment to the "rationality of science". If Busso's work can help us redefine black hole observability, I eagerly await what we discover next. However, at present, black hole interiors seem to be clearly epistemically inaccessible to humanity on Earth, despite being a part of our universe that a realist would want to be able to describe ontologically. There is nothing especially pernicious about the space-time inside a black hole event horizon, as we currently view it. There is no obvious reason that a realist would be happy to believe that it is outside the obvious jurisdiction of science. This isn't a problem for the constructive empiricist who, via an understanding of the observable/unobservable distinction and the knowledge that all physics requires for success is the empirical adequacy of its theories, can. Constructive empiricism is either outright superior in terms of removing these paradoxes or is more elegant when it comes to aligning them with its understanding of what physics does. In any case, it remains more attractive as a framework for the physicist to adopt pragmatically, as I contend it will lead to better theory development and less time spent arguing nigh-on philosophically about things we can't empirically test.

We will conclude this chapter by outlining the work achieved in it, before discussing the consequences of these conclusions for our working physicist, who I maintain should firmly adopt a constructive empiricist outlook. This is not solely because it prevents realist concerns about the "truth" of a region we will never access from generating, but also because it provides a superior framework for understanding and talking cogently about modern physics.

4.2.2 Why the Black Hole Physicist Should be a Constructive Empiricist

This Chapter began by discussing black hole observability for the constructive empiricist. It becomes clear that when we talk whether the black hole interior is observable or unobservable in some universal sense, problems arise in the initial constructive empiricist accounting. In seeking to consider the black hole interior and exterior as one, we end up at a position where the constructive empiricist must either abandon van Fraassen's epistemic modesty or the accuracy of his heuristic, both being deeply unappealing. To solve this, we have shown that the limits of what we can know about black holes, including whether we can observe aspects of them, can be defined by a novel approach to epistemic communities. Whether one can observe a black hole is dependent on your epistemic community, which we now understand as being relativised via one's spatiotemporal location to other beings with the same epistemic skills as them. The limits of science are derived from the boundaries of what a community of scientists can perceive, experience, and communicate to one another. Therefore, it should be no wonder

that the same limitations bound our constructive empiricist understanding of whether we can observe a black hole.

After we have established this new framework, we demonstrate that work in physics already presupposes these ideas conceptually. Black hole exteriors are, to generalise, much more intuitively graspable than black hole interiors. That they conceptually operate in this way I don't take to be a function of arbitrary physics, but a hint about their relative observability and the fact we have never observed any region that is currently epistemically removed from us. This again speaks to the constructive empiricist case and makes it superior to the realist one.

Finally, we turn back to the Page-time paradox. Returning to Section 2.2.5, we outline the pathways a physicist could take to the Page-time paradox and offer both a pessimistic and optimistic reading. On both accountings, the constructive empiricist case is more attractive than the realist case.

What then, should our working black hole paradox physicist do? I will start by highlighting my belief that constructive empiricism is the best and right approach, and that with either a positive or negative outlook to breaking research work, it is superior to realism. A realist framing of the paradox creates headaches and problems that will both waste valuable research time and lead physicists in the direction of not providing *useful* progress. This is no slight at anyone with realist leanings, whose motivations for said realism are noble and meaningfully important. However, constructive empiricism a simply more pragmatic approach to use in the context of these issues and adopting it will lead to the better progression of research o. Physicists should adopt a constructive empiricist view on, for example, the observability of black hole interiors, base that within the constructive empiricist framework regarding what science is and can do, and not concern themselves with larger metaphysical questions. All that is required is reframing their desires from "I want to find out truth about the universe" to "I want to best model and practically use all the things I can observe about the universe".

If we take a positive outlook to current work on information loss paradoxes, the only cogent way to understand modern solutions to information loss paradoxes like the Page-time paradox is to view them through a constructive empiricist lens. To do so through a realist lens leaves us fundamentally adrift in metaphysical questions about a region that physics and physicists cannot hope to epistemically access. If we wish to be a realist, we could be committed to the truth of theories that, in the case of firewalls as an example, contend that the singularity simply *is* the black hole interior after a black hole has passed the Page-time (as Susskind makes clear is a logical consequence of abandoning a complementarian view (Susskind, 2012, p. 11)). On

the other hand, our only other option is to deny the truth of otherwise empirically adequate theories, such as QFT or Stat Mech, which looks either a painful prospect or something we are currently not able to do reasonably[64]. "Entropy islands" and the theoretical framework that underpins them may provide some relief from this dire of choices for the realist, but it does not help the realist with concerns about epistemic immodesty or make their framework more pragmatic than a constructive empiricist approach.

Taking a negative view of the modern progress, the constructive empiricist can point to the newness of work like Busso et al.'s and the open questions about its relevancy to constructive empiricism. Our ability to do physics is limited by our abilities as humans, and thus even if we can now theoretically recover information from a black hole, we cannot test if this actually does recover the information either from our perspective here on Earth (there are no post-Page-time black holes to be found on this planet, currently) or remaining within our current epistemic community. If we reject such work from our accounting and think only of, for example, black hole complementarity and firewalls, our working physicist is still better being a constructive empiricist. Both theories agree about the black hole exterior and predict the same things, therefore the paradox in choosing between them never generates, meaning we don't need to spend time thinking about them.

Other pragmatic arguments also exist. We could make the case that as we have not yet seen a post-Page-time black hole or done any practical investigations as to what happens when you send a probe through the event horizon, it is still an open question if even the very starting point of the paradox generates for the constructive empiricist. We have no empirical evidence for a post-Page-time black hole currently; they may just not be present to us in the universe. Finally, by being a constructive empiricist we can understand how the work in this field is no longer simply "physics", and whilst it may be illuminating and knowledge-baring, it is not knowledge-baring in the same way "physics" is, as it is seeking to apply an inappropriate set of beliefs and desires to a region of the universe that can't sustain them. We have not yet empirically tested many of the theories outlined in this Chapter. We may never do so. How can we be sure we are still doing physics under such circumstances, given physics is primarily an empirical subject?

The novel friendly amendment offered in this chapter to van Fraassen's understanding of epistemic communities is what allows the constructive empiricist to do most of these things.

[64] Is it reasonable to abandon decades of empirical evidence for things like the monogamy of entanglement based on one paradox in a single physics niche which we currently have no way of currently empirically testing? I think not.

Without it, they are left with headaches as the realist is as they are forced into accepting the black hole interior as observable, which given best current theory it cannot be. Without this novel understanding of epistemic communities and new heuristic for observability, the constructive empiricist would be left in the same place as the physicist: Requiring theories about the dynamics of the black hole interior to be empirically adequate.

Whichever approach to the Page-time paradox you choose, the consequences remain similar. Realist understandings of the universe set us up for failure, either in identifying issues in physics we can fix or in giving us unreasonable and unsatisfactory consequences being forced to align our metaphysics with science. Our working black hole paradox physicists are better off being constructive empiricists in every scenario.

Conclusion – The Way Forward

"To be a scientist is to be naïve. We are so focused on our search for truth; we fail to consider how few actually want us to find it. But it is always there, whether we see it or not, whether we choose to or not" – Valery Legasov (as portrayed by Jared Harris) (Chernobyl, 2018, ep. 5)

Science, and physics more immediately for this thesis, is framed as being the central "truth" finding practice. This framing comes not only from philosophical schools, but also from the interaction between science, scientists and the society that surrounds and creates them. To consider physics and science as monolithic, isolated subjects, removed from the context of the people doing the subject, fails to correctly account for the influences that can lead to intuitions and unknown philosophical work.

The quote that leads this conclusion comes from the hit HBO mini-series, Chernobyl, which follows the actions of Valery Legasov, a historical physicist, in the wake of the Chernobyl disaster. He is presented to the watcher as being a solid and fundamentally 'correct' scientist, a man fighting against apparatchiks and bureaucrats in the cause of "truth", as that is the thing that will save lives and stop a disaster turning into Armageddon. I choose this piece of media and present it here to help demonstrate the point that our societal understanding of what it is to do physics and how connected the "search for truth" and "doing physics" are, is deeply realist. If we accept that individuals are products of their environments and develop unconscious and conscious understandings of the world via their interactions with it, we must accept that a narrative they see form around them across media, education and politics will take hold.

Valery Legasov, as written by Craig Mazin, states that being a scientist is fundamentally connected to a supposed "search for truth". Millions of people will have watched that line on their TV screens and rather than react with shock will immediately recognise it as orthodox and normal. That is why the line is written the way it is, screenwriters of big budget TV shows not seeking to fundamentally undermine the core conceptions of their audiences but tell a story within the societal narrative that surrounds them, that doesn't confuse them. This orthodoxy will stay with the audience, whether they go on to interact with scientific thought in no way at all or interact with it daily. The central argument of this thesis is this presentation of what science is leads to a pervasive realist orthodoxy which hampers the progress of physics. From here, I endorse constructive empiricism as the school for working physicists to adopt if they wish to recover that progress or *usefulness*.

Have we established this case? And what was the pathway we took to get here? We begin in Chapter 1 with a discussion and explanation of one of the central problems of this thesis Page-time paradox. I claim that this paradox is both generated and kept alive by realist intuitions in physics. In the Chapter, I introduce some underpinning physics theory to help explain the paradox and highlight interesting epistemic aspects of black hole dynamics. After this I introduce the paradox and two theories connected to it: black hole complementarity and firewalls. Both theories have interesting conceptual and philosophical consequences, and further hint that the work being done in this field is removed from the empiricism that underpins the study and practice of physics, which in turn I contend leads to the problems for physics in terms of progress and usefulness.

Chapter 2 follows by presenting the terms of engagement for our argument philosophically, establishing a bare description of "what physics is", a discussion on how terminology will be used and providing a protagonist against which we can judge the practical reasonableness of our future conclusions: the working physicist. From here I show that these realist orthodoxies are existent within the *milieu* surrounding science, giving pedagogic and linguistic defences to that position. After this I show the problems that this can cause the working physicists, that by focusing too deeply on metaphysical desires we lose track of what the function of physics is, and limits in what it can do. To frame this argument, I then reintroduce the Page-time paradox, show pathways our working physicist can take, before outlining what I maintain to be a solution already present to us within the philosophy of science: constructive empiricism. Chapter 2 then makes the case for constructive empiricism and defends the idea that working physicists should either become constructive empiricists or accept that they are constructive empiricists anyway given the way they practice physics. I endorse this school of anti-realism as I believe it offers the least objectional account of what science is for those with pre-existing realists leanings, as it begins by providing a model of what is necessary for science that is grounded in the behaviours and actions of actual scientists.

Chapter 3 begins by investigating another core reason to believe in constructive empiricism over realism: that physics presupposes it via the arguments it has and the way it views properties and model. I make this within the context of entropy, looking at problems and thought experiments in colloid, polymer, and black hole physics. The second grand narrative tackled in Chapter 3 seeks to show that even famous and long-standing paradoxes in statistical mechanics, namely Maxwell's Demon, can be more easily exorcised or solved if we endorse the constructive empiricist case, as the realist case's metaphysical requirements place responsibilities on physics it is not well placed to handle.

Chapter 4 starts by talking about black hole physics and asks an important question we must respond to if we wish to make the case for constructive empiricism over realism: are black holes observable? At first, the naïve constructive empiricist answer seems problematic, with the constructive empiricist seemingly committed to black hole interiors being observable, them being permanently beyond the range of what humanity on Earth has epistemic access to. This is solved via a novel introduction of my own creation to the constructive empiricist framework: extended epistemic communities. By viewing our epistemic community as not just those with whom we share the same broad epistemic skills, but also those with whom we can communicate, we can identify that black hole exteriors are observable, black hole interiors aren't observable, and that the event horizon is definitively an epistemic barrier our epistemic community here on Earth can't cross. I follow this up by demonstrating how physics already presupposes such a conclusion via thought experiment and example, before using this new and improved version of constructive empiricism to finally re-approach the typical example given in Chapter 1, the Page-time paradox, and see if pathways for solutions are now available which aren't available to the realist.

In short, I make the case that adopting constructive empiricism if you are a working physicist, especially a working physicist working on the problems outlined in this paper, is always superior to maintain realist orthodoxies and intuitions. This case is argued in several different ways, which allows us to adopt it comfortably.

In adopting constructive empiricism, we must accept some compromise with our intuitions, though no compromise that is already outside the realms of what is reasonable. By adopting the constructive empiricist case, we must accept there exist limitations to the practice of physics that are, in some sense, discomforting to those who have certain realist intuitions. Chapter 4 also wrestles with this idea, that there may be some areas of physics that, whilst physically present within what we would previous conceive as "our universe", are beyond the scope of physics inquiry. The reason for this is well founded: there necessarily is an epistemic gap present at the black hole event horizon[65], which fundamentally inhibits our ability to observe or make tangible this region of the universe with respect to our current epistemic community. Showing a macroscopic, theoretically enterable, region of our universe is somehow something physics can't research cogently *is* discomforting. It demonstrates a

[65] A consequence of the Einsteinian General Relativity and the unbreachable top speed of light. The constructive empiricist accepts this as it is an empirically adequate theory, and a good *note bene* is to remember that constructive empiricism is fine with accepting theory-laden defences of empirical adequacy, as it accepts that the framework itself is theory-laden.

disconnect between the supposed “truth” finding desires of physicists and the realities of empirical inquiry.

This discomfort, however, is not only necessary but fundamentally implied by the study of physics itself. Physics is an empirical subject, a subject based on what humans can observe, test, and demonstrate as consistent. The realists agree on this, saying that the consistency and predictive power at least shows *approximately* the truth about the structure of the universe. I don’t draw this conclusion and furthermore think that physicists, philosophers and scientists can occasionally reverse the logic of this argument in ways that badly hampers physics: our best physics theories tell us what is approximately true about the universe, therefore what our models conclude about places we haven’t yet and can’t observe must be the way those system function. Both Maxwell’s Demon and the Page-time paradox problems suffer from this reversal, where the truth of our theories are assumed and physicists plough on, authoring paper after paper of interesting and potentially useful mathematics that is no longer connected to the underlying empiricism that forms the necessary core of any methodological understanding of physics.

There are many avenues to potential good future research within this topic. This thesis, to maintain one cohesive whole, has not sought to chase down every possible conclusion or assumption, as to do so would be an entire career’s worth of study. Instead, this conclusion offers interesting future questions that the physicist/philosopher could ask about the scope of physics, the properties of physics or the consequences of arguments made throughout this thesis.

One concept that is worth addressing is how widespread this problem is and, calling back to Section 2.2.6, whether we should adopt constructive empiricism as a philosophy of science across all the fields of physics. This thesis limits itself to simply areas within statistical mechanics (polymer and colloid physics, the arguments surrounding Maxwell’s Demon) and black hole physics (information paradoxes, such as the Page-time paradox, and black hole entropy). I do not wish currently to make an overarching claim as to how this project would succeed in other areas of modern physics, such as quantum mechanics, but future work would address paradoxes or other research issues that exist within those fields more directly, looking for epistemic gaps or other structural issues that are highlighted as in this thesis are resolvable via constructive empiricist approaches.

Another potential future avenue of exploration involves asking whether it is entropy itself that is the problem. Across this thesis, entropy has formed the nexus of the problems we have

uncovered, and especially the question of whether an emergent or reductionist view about it. This question, along with others (is the connection between entropy and the information available to the observer, *a la* information theoretic approaches, philosophical interesting? Does every epistemic gap within physics have at its root entropy?) could form the basis of large amounts of future work, including the first concept of future work raised within this section.

A third, and most interesting to me, future pathway for research is investigating the consequences of the "inchoate intuitions of physics" on public policy and real-world outcomes, divorced from the hard science which is the scope of this thesis. In what way do realist views of physics, and science more generally, affect public policy decisions and political decision-making outside the scope of the physics laboratory. This question has been asked before, including by Nancy Cartwright in her work referenced in Chapter 2, but a novel approach would be to use constructive empiricism and, for example, the expanded concept of epistemic communities introduced in Chapter 4, to develop a framework for teaching policy decision-makers the consequences of treating "science" as equivalent to "truth" and why basing policy decisions on "science" isn't necessarily a pathway to the "correct" decision.

I would choose to frame such research around two important areas, both deeply complex and controversial: the United Kingdom's governmental response to the COVID-19 pandemic and debates about gender identity. The discourse surrounding both is deeply rooted in scientific language and claim about science. The United Kingdom government's central messaging during the first two lockdowns being that they were, and we all should, "follow the science" (which was why policies like lockdown were introduced in the first place). Regarding policy questions on gender identity, policy decisions about, for example, trans women's access to sport are framed using science as the way to decide, with other questions given less import. The level of testosterone in someone's blood is established as a way of delineating those who "are women" from those who "aren't women" in absolutist and controversial ways, and I maintain in a way that science cannot conclusively provide for as an empirical subject. In both examples the overarching orthodoxy of scientific realism is present, with science being both the true and right way of answering complex, multifaceted questions. Whether a constructive empiricist understanding of both the limits of physics and empiricism would help provide better policy decision making is, I think, deeply essential research on both a practical and moral level, as the moral consequences of getting policy wrong in these fields (up to and including people dying as a direct consequence) matters practically more than if someone's research on the Page-time paradox is based on faulty assumptions. I maintain that a root cause of the debate and discourse within these complex arguments is how we understand and view the hard sciences,

and therefore the way we as physicists think about the subject of physics directly trickles down to both popular consciousness about science, and how we engage with and respond to expert testimony.

Overall, this thesis serves as a direct plea to both physicists and philosophers to question the underpinning assumptions and intuitions they have regarding the question of "what physics can do". This thesis contends that in not questioning the unspoken realist assumptions physics, physicists have unwittingly given themselves problems to solve that they cannot, and that these problems will remain fundamentally unresolvable until a different paradigm is established. The most convincing and agreeable route to this new paradigm is via Bas van Fraassen's constructive empiricism, and by adopting it physicists would be able to do better, more appropriate research that leads to better practical gains in physics knowledge, no matter how disheartening this may be.

Acknowledgements

It would be remiss of me to begin this set of acknowledgements without a moment of reflection. This project, which started back in September 2019, has changed dramatically over the subsequent 5 years due to dramatic events in the external world. Through COVID-19, the death and severe illness of loved family members, and unforeseeable homeless, I would have had no chance of putting together a document which reflects a dream I have held since I was a small child without the unwavering support of **Prof. Mary Leng**. Mary has been the only supervisor I have had who has seen this project from start to finish, from being originally examined in the subject of physics to now being examined within philosophy. She has my eternal gratitude and has helped me become a much better philosopher and physicist than where I began.

I am also grateful for the support of my physics supervisor, **Dr. Istvan Cziegler,** who stepped into a project half-finished and unsure of its direction to help steady the ship, provide calm advice, and re-establish a way forward.

I began this thesis with a dedication to **Prof. Tom McLeish, FRS**, who sadly passed away in March 2023 at far too young an age. It would be impossible to ever give justice to the impact he has had on my life and my faith in institutions of academia. His belief that the problems in philosophy and physics that I perceived were not only worthy of future investigation, but something he would work hard to see investigated, are the only reason I have been able to write the words you see before you. This thesis simply never happens without his belief, support, and hard work. Each day and every word I write, I miss his guidance and good nature. I hope you would be proud of what I've achieved, Tom.

I would also like to thank my peers and colleagues for supporting me both practically and intellectually throughout this project. In particular, I would like to thank by name **Joe Reed, Alex Howarth, and Dr. Catherine Yarrow**, all of whom help keep me sane during the challenges I have faced or practical support when I most needed it. Again, I do not think I would have had a chance of submitting without their invaluable support.

The University of York, both the physics and philosophy communities, deserve great praise as well for giving me a foundation and structure in which I could work on concepts I have been developing, piecemeal, over my entire life. My thesis was centrally funded by the University, the Department of Philosophy and the Department of Physics, Engineering and Technology. I thank them for their faith in me and my ideas.

Bibliography

Akers, C., Engelhardt, N. & Harlow, D., 2020. Simple holographic models of black hole evaporation. *Journal of High Energy Physics,* 2020(8), pp. 1-14.

Almheiri, A., Hartman, T., Maldacena, J. & al., e., 2020b. Replica wormholes and the entropy of Hawking radiation. *J. High Energ. Phys.,* pp. 1-36.

Almheiri, A., Hartman, T., Maldacena, J. & Edgar Shaghoulian, A. T., 2020a. *Replica Wormholes and the Entropy of Hawking Radiation.* [Online]
Available at: https://arxiv.org/pdf/1911.12333.pdf
[Accessed 17 01 2022].

Almheiri, A., Marolf, D., Polchinski, J. & Sully, J., 2013. *Black holes: complementarity or firewalls?.* [Online]
Available at: https://arxiv.org/abs/1207.3123
[Accessed 14 06 2020].

Alspector-Kelly, M., 2004. Seeing the Unobservable: Van Fraassen and the Limits of Experience. *Synthese,* 140(3), pp. 331-353.

AQA , 2015. *GCSE Physics 8463.* [Online]
Available at: https://filestore.aqa.org.uk/resources/physics/specifications/AQA-8463-SP-2016.PDF
[Accessed 20 May 2024].

AQA, 2017. *AS and A-level Physics.* [Online]
Available at: https://filestore.aqa.org.uk/resources/physics/specifications/AQA-7407-7408-SP-2015.PDF
[Accessed 20 05 2024].

Bekenstein, J., 1974. Generalized second law of thermodynamics in black-hole physics. *Phys Review D,* 9(12), pp. 3292-3300.

Bekenstein, J., 1995. Novel "no-scalar-hair" theorem for black holes. *Physical Review D,* 51(12), pp. 6608-6611.

Bekenstein, J. D., 1973. Black Holes and Entropy. *Physical Review D,* 7(8), p. 2333–2346.

Belot, G., Earman, J. & Ruetsche, L., 1999. The Hawking Information Loss Paradox: The Anatomy of a Controversy. *The British Society for the Philosophy of Science,* 50(2), pp. 189-229.

Bokulich, P., 2005. Does Black Hole Complementarity Answer Hawking's Information Loss Paradox?. *Philosophy of Science,* 72(5), p. 1336–1349.

Boucher, S. C. & Forbes, C., 2024. The pragmatic turn in the scientific realism debate. *Synthese,* 203(4), pp. 1-23.

Brillouin, L., 1951. Maxwell's Demon Cannot Operate: Information and Entropy. I. *Journal of Applied Physics,* Volume 22, pp. 334-337.

Busso, R. & Penington, G., 2023. *Islands Far Outside the Horizon*. [Online]
Available at: https://arxiv.org/abs/2312.03078
[Accessed 26 06 2024].

Callendar, C., 2018. *Can We Quarantine the Quantum Blight? [Preprint]*. [Online]
Available at: https://philsci-archive.pitt.edu/15450/
[Accessed 30 07 2024].

Cartwright, N., 2022. *A Philosopher Looks at Science*. 1st ed. Cambridge: Cambridge University Press.

Cates, M. E. & Manoharan, V. N., 2015. Celebrating Soft Matter's 10th anniversary: Testing the foundations of classical entropy: Colloid experiments. *Soft Matter,* I(11), pp. 6538--6546.

Chakravartty, A., 2017. *Scientific Realism*. [Online]
Available at: https://plato.stanford.edu/entries/scientific-realism/
[Accessed 29 07 2024].

Chernobyl. 2018. [Film] Directed by Johan Renck. United States of America: HBO.

Churchland, P., 1985. The Ontological Status of Observables: In Praise of the Superempirical Virtues. *Churchland and Hooker 1985,* p. 35–47.

Churchland, P. M. & Hooker, C. A. eds., 1985. *Images of Science*. Chicago: University of Chicago Press.

Curiel, E., 2019. *Singularities and Black Holes*. [Online]
Available at: https://plato.stanford.edu/entries/spacetime-singularities/#InfoLossPara
[Accessed 15 06 2020].

Earman, J. & Norton, J., 1999. EXORCIST XIV: The Wrath of Maxwell's Demon. Part II. From Szilard to Landauer and Beyond. *Stud. Hist. Phil. Mod. Phys,* 30(1), pp. 1-40.

Eddington, S. A. S., 1948. *THE NATURE OF THE PHYSICAL WORLD*. Cambridge: Cambridge University Press.

Einstein, A., 1914. *Volume 4: The Swiss Years: Writings 1912-1914 (English translation supplement) Page 291*. [Online]
Available at: https://einsteinpapers.press.princeton.edu/vol4-trans/303
[Accessed 04 06 2020].

Einstein, A., 1920. *Relativity: the Special and General Theory*. s.l.:Methuen & Co Ltd.

Ellis, B., 1988. INTERNAL REALISM. *Synthese,* 76(3), pp. 409-434.

Encyclopedia Britannica, 2021. *Polymer*. [Online]
Available at: https://www.britannica.com/science/polymer
[Accessed 17 01 2022].

Encyclopedia Britannica, 2024. *Michelson-Morley experiment*. [Online]
Available at: https://www.britannica.com/science/Michelson-Morley-experiment
[Accessed 29 07 2024].

Englehardt, J., 2019. Linguistic labor and its division. *Philosophical Studies,* 7(176), pp. 1855-1871.

Event Horizon Telescope, 2022. *Astronomers Reveal First Image of the Black Hole at the Heart of Our Galaxy*. [Online]
Available at: https://eventhorizontelescope.org/blog/astronomers-reveal-first-image-black-hole-heart-our-galaxy
[Accessed 06 26 2024].

Feynman, R., 1964. *1964 Cornell Messenger Lectures: Seeking New Laws*. Ithaca, Cornell University.

Forbes, C., 2017. A pragmatic, existentialist approach to the scientific. *Synthese,* Volume 194, pp. 3327-2246.

Frenkel, D., 2014. Why colloidal systems can be described by statistical mechanics: some not very original comments on the Gibbs paradox. *Molecular Physics,* 112(17), p. 2325–2329.

Gautason, F. F., Schneiderbauer, L., Sybesma, W. & Thorlacius, L., 2020. Page Curve for an Evaporating Black Hole. *Jounal of High Energy Physics,* 2020(5), pp. 1-20.

Hacking, I., 1985. Do We See through a Microscope. In: *Images of Science*. Chicago: University of Chicago Press, pp. 132-152.

Harlow, D., 2015. *Jerusalem Lectures on Black Holes and Quantum Information*. [Online]
Available at: https://arxiv.org/abs/1409.1231
[Accessed 14 06 2020].

Hawking, S., 1975. Particle Creation by Black Holes. *Commun.Math.Phys,* Volume 43 , pp. 199-220.

Hod, S., 2015. Bekenstein's generalized second law of thermodynamics: The role of the hoop conjecture. *Physics Letters B,* Volume 751, pp. 241-245.

Jaynes, E. T., 1992. THE GIBBS PARADOX. In: S. C.R., E. G.J. & N. P.O., eds. *Maximum Entropy and Bayesian Methods: Fundamental Theories of Physics (An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application)*. Dordrecht: Springer, pp. 1-21.

Kelly, L., 2014. *What Is A Fruit?*. [Online]
Available at: https://www.nybg.org/blogs/science-talk/2014/08/what-is-a-fruit/#:~:text=A%20fruit%20is%20a%20mature,fertilized%20and%20turn%20into%20seeds.
[Accessed 31 05 2023].

Ladyman, J., 2000. What's Really Wrong with Constructive Empiricism? Van Fraassen and the Metaphysics. *The British Journal for the Philosophy of Science,* 51(4), pp. 837-856.

Leff, H. S. & Rex, A. F., 1990. *Maxwell's Demon: Entropy, Information, Computing*. Princeton: Princeton University Press.

LIGO Scientific Collaboration and Virgo Collaboration, 2016. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physics Review Letters,* 116(061102), pp. 1-16.

Machamer, P. & Miller, D. M., 2021. *Galileo Galilei*. [Online]
Available at: https://plato.stanford.edu/entries/galileo/
[Accessed 30 07 2024].

Maxwell, J., 1931. *Letter from Maxwell to Tait on Maxwell's demon, 11 December 1867 (P92 (a), (b))*. [Online]
Available at: https://cudl.lib.cam.ac.uk/view/PH-CAVENDISH-P-00092/1
[Accessed 02 01 2024].

Misner, C. W., Thorne, K. S. & Wheeler, J. A., 1973. *Gravitation*. s.l.:W. H. Freeman & Company.

Monton, B. & Mohler, C., 2021. *Constructive Empiricism*. [Online]
Available at: https://plato.stanford.edu/entries/constructive-empiricism/
[Accessed 19 06 2024].

Monton, B. & van Fraassen, B., 2003. Constructive Empiricism and Modal Nominalism. *The British Journal for the Philosophy of Science,* 54(3), pp. 405-422.

NASA Jet Propulsion Laboratory, 2023. *Basics of Space Flight*. [Online]
Available at: https://solarsystem.nasa.gov/basics/chapter3-2/#:~:text=Spacecraft%20operate%20at%20very%20high,navigating%20throughout%20the%20solar%20system.
[Accessed 17 04 2023].

Nunno, B. D. & Matzer, R., 2009. The Volume Inside a Black Hole. *General Relativity and Gravitation,* Issue 42, pp. 63-76.

Page, D. N., 1993. Information in black hole radiation. *Physical Review Letters,* Volume 71, p. 3743–3746.

Park, J., 1970. The concept of transisiton in quantum mechanics. *Foundations of Physics,* 1(1), pp. 22-33.

Quine, W. V. O., 1966. The Ways of Paradox. In: W. V. Quine, ed. *The Ways of Paradox and Other Essays*. Cambridge, Massachusetts: Harvard University Press, pp. 1-18.

Rindler, W., 1956. Visual Horizons in World Models. *Monthly Notices of the Royal Astronomical Society,* 116(6), p. Monthly Notices of the Royal Astronomical Society.

Shankland, R., 1964. Michelson-Morely Experiment. *American Journal of Physics,* I(32), pp. 16-35.

Smoluchowski, M., 1914. Gültigkeitsgrenzen des Zweiten Hauptsatzes der Wärmetheorie. *Vorträge über die Kinetische Theorie der Materie und der Elektrizität,* p. 89–121.

Susskind, L., 2012. *Singularities, Complementarity and Firewalls*. [Online]
Available at: https://arxiv.org/pdf/1208.3445.pdf
[Accessed 14 06 2020].

Susskind, L., Thorlacius, L. & Uglum, J., 1993. Complementarity, The Stretched Horizon and Black Hole. *Physical Review D,* 48(8), pp. 3743-3761.

Susskind, L., Thorlacius, L. & Uglum, J., 1993. The stretched horizon and black hole complementarity. *Physical Review D,* 48(8), p. 3743–3761.

Szilard, L., 1964. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Behavioural Science,* 9(3), pp. 301-310.

't Hooft, G., 2000. *The Holographic Principle*. [Online]
Available at: https://arxiv.org/abs/hep-th/0003004
[Accessed 15 06 2020].

Teller, P., 2001. Whither Constructive Empiricism?. *Philosophical Studies: An International Journal for Philosophy in the Analytic,* 106(1), pp. 123-150.

Terhal, B. M., 2004. Is entanglement monogamous?. *IBM Journal of Research and Development,* 48(1), pp. 71-78.

The University of York, 2024. *MPhys (Hons) Physics*. [Online]
Available at: https://york.ac.uk/study/undergraduate/courses-2024/mphys-physics/#course-content
[Accessed 20 05 2024].

University of Cincinnati, 1998. *Chapter 1 Polymer Physics - The Isolated Polymer Chain*. [Online]
Available at: https://www.eng.uc.edu/~beaucag/Classes/Physics/Chapter1.html
[Accessed 24 07 2024].

Unruh, W. G. & Wald, R. M., 1995. On Evolution Laws Taking Pure States to Mixed States in. *Physical Review D,* 52(4), pp. 2176-2182.

Unruh, W. G. & Wald, R. M., 2017 . *Information Loss*. [Online]
Available at: https://arxiv.org/pdf/1703.02140.pdf
[Accessed 09 06 2020].

van Fraassen, B., 1980. *The Scientific Image*. Oxford: Oxford University Press.

van Fraassen, B., 1985. Empiricism in the Philosophy of Science. *Churchland and Hooker 1985,* p. 245–308.

van Fraassen, B., 1989. *Laws and Symmetry*. Oxford: Clarendon Press.

van Fraassen, B., 1995. Against Naturalized Epistemology. In: P. Leonardi & M. Santambrogio, eds. *On Quine*. Cambridge: Cambridge University Press, p. 83.

van Fraassen, B., 2001. Constructive Empiricism Now. *Philosophical Studies,* Volume 106, p. 151–170.

van Fraassen, B., 2004. REPLIES TO DISCUSSION ON THE EMPIRICAL STANCE. *Philosophical Studies,* Volume 121, pp. 171-192.

van Fraassen, B., 2004. *The Empirical Stance*. s.l.:Yale University Press.

van Fraassen, B., 2005. The day of the dolphins: Puzzling over epistemic partnership. In: A. D. Irvine & K. A. Peacock, eds. *Mistakes of Reason: Essays in Honour of John Woods*. Buffalo: University of Toronto Press, pp. 111-133.

Wald, R., 2010. *General Relativity*. Chicago: University of Chicago Press.

Wallace, D., 2017. *Why Black Hole Information Loss is Paradoxical*. [Online]
Available at: https://arxiv.org/abs/1710.03783v2
[Accessed 11 06 2020].

Wallace, D., 2018a. The case for black hole thermodynamics Part I: phenomenological thermodynamics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics,* Volume 64, pp. 52-67.

Wallace, D., 2018b. The case for black hole thermodynamics Part II: Statisitical Mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics,* 13 06, Volume 66, pp. 103-117.

Wilkins, A., 2024. *'Islands' poking out of black holes may solve the information paradox*. [Online]
Available at: https://physics.berkeley.edu/news/%E2%80%98islands%E2%80%99-poking-out-black-holes-may-solve-information-paradox
[Accessed 26 06 2024].

Wivagg, D. & Allchin, D., 2002. The Dogma of "The" Scientific Method. *The American Biology Teacher,* 64(9), pp. 645-646.

Worrall, J., 1989. Structural Realism: The Best of Both Worlds?. *Dialectica,* 43(1), pp. 99-124.

Yale University, 2024. *Probability Distribution of End-End Distances*. [Online]
Available at:
https://www.eng.yale.edu/polymers/docs/classes/polyphys/lecture_notes/3/handout3_wse1.html
[Accessed 07 24 2024].