MASTER THESIS

Radboud Universiteit

The Demon and the Abacus

An analysis, critique, and reappreciation of digital physics

Supervisors:
Dr. Ir. F.A. Bakker
Prof. N.P. Landsman

Author:
P.M.N. Sonnemans
4738799

June 11, 2020

Scriptie ter verkrijging van de graad “Master of arts” in de filosofie
Radboud Universiteit Nijmegen

Hierbij verklaar en verzeker ik, P.M.N. Sonnemans, dat deze scriptie zelfstandig door mij is opgesteld, dat geen andere bronnen en hulpmiddelen zijn gebruikt dan die door mij zijn vermeld en dat de passages in het werk waarvan de woordelijke inhoud of betekenis uit andere werken – ook elektronische media – is genomen door bronvermelding als ontlening kenbaar gemaakt worden.

Plaats: datum:

Abstract

Digital physics (also referred to as digital philosophy) has a broad spectrum of theories that involve information and computation related alternatives to regular physics. This thesis provides an overview of the main branches of thought and theories in this field. The main claims that computation is either the ontological foundation of the universe or a useful way of describing the universe are critiqued on a philosophical level. Yet an effort is made to reappreciate the ideas in digital philosophy, especially in the way this approach can lead to a better understanding of the limitations of physics as it is constrained by formalization, computability, and epistemology.

Contents

Introduction

The Age of Enlightenment saw an explosion of mathematics, physics, logic and philosophy being practiced hand-in-hand in European academia. The hugely successful developments in mathematics and physics also found their way into the other sciences. More than that, it also again shifted the center of gravity in the age old discussion between determinism and free will. The understanding of mechanical physics, mainly through the theories of Newton and subsequent scientists, reconfirmed philosophers to favour a deterministic world view. The most iconic example of this movement was the explanation by the French mathematician Pierre-Simon de Laplace. Laplace argued that if someone knew the precise location of each particle in the universe, as well as the forces at work between these particles, the past and the future of each particle could be calculated and therefore the past and future would become indistinguishable from the present.

> Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it - an intelligence sufficiently vast to submit these data to analysis - it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. (Laplace, 2009)

This omniscient figure later became known as *Laplace's Demon*. It is interesting, however, to note that Laplace definitely was not the first to toy with this view. Before he published his famous work in 1814, philosophers and mathematicians throughout the 17th and 18th century had held very similar views (van Strien, 2014). The discovery of the power of mathematics and mechanics cross-pollinated with philosophical views on determinism throughout this age. Determinism was already an idea known to the ancients, notably by Epicureans, but this so called scientific determinism distinguished itself from the Epicureans by being backed up by the success of physics at the time. This lasted until the early 20th century. When the theories of quantum mechanical irreversibility showed that the deterministic model was no longer tenable. These theories definitely problematized the classic notions Laplace and contemporaries had held. And so, Laplace's demon faded away (Hacking, 1990).

It was only in the late 1960s that a new scientific cross-pollination started to regerminate scientific determinism. The power and success of the use of computers in the 20th century did not go unnoticed by people in the field of philosophy and fundamental physics. It was not a huge leap to hypothesize that what a computer could so easily do was actually the concept at the root of the universe and physics: *computation*, or even one step further, *simulation*. The demon is back. This time the demon is not characterized by its knowledge of mechanical physics, this time it is characterized by the fact that it computes. It is a demon with an abacus.

The scientific field that originated in the late 60s but took off in the 90s was dubbed digital physics, although others prefer the term digital philosophy. In this field we can find physicists writing about mathematics, mathematicians writing about philosophy, philosophers writing about computer science, and computer scientists writing about physics. The common denominator is the simple idea that information and computation are fundamental in physics and philosophy. Some of these ideas regard the ontological state of the universe, others the phenomenological state. This thesis will explore the ideas in the field of DP.

Related to the ideas floating around in DP is the *simulation hypothesis*. This hypothesis holds that our universe is actually being simulated somewhere else, perhaps by a different intelligent species (Bostrom, 2003) (Whitworth, 2007). The simulation hypothesis is seen as a solution to the *something-from-nothing-problem*. This idea got even more traction after Elon Musk voiced that he committed himself to it (Wall, 2018). The simulation hypothesis has even been used as an argument for the existence of God (InspiringPhilosophy, 2013). It should be clear that the simulation hypothesis is not an explicit subject of this thesis. However, all the critique that this thesis may have on DP can also be reinterpreted by the reader as a critique of the simulation hypothesis.

This thesis will provide an overview of the use of information and computation in physics and philosophy. The discussions in the DP field revolve around either or both of the concepts of information and computation. My discussion of information focuses on the question whether information is phenomenological or ontological. Or, as in the words of some of the authors we will discuss: 'bit from it' or 'it from bit'. If information is the medium, computation can be regarded as the force that sets the information into motion. The authors in the field of DP have varying ideas regarding the ontological status of computation. Some believe that computation is the ontological essence of the universe, others take a more nuanced approach where they merely see computation as a very useful tool of describing our universe. In the first section of this thesis prominent authors and their viewpoints in the field of DP are surveyed in depth. This is in service of the overall dialectic structure of this thesis: build, destroy, rebuild.

The goal of the section following the first section reviews the premises on which most of DP is built. This is for the most part computer science theory and logic. However, it is also useful to discern the underlying metaphysical premise of the authors. As unfortunately often happens, interdisciplinary ideas can easily be misinterpreted. This thesis is going to need to take a position with regard to computer science, epistemology and metaphysics. Taking such a position allows us to dissect the arguments that have been reviewed in the first section. With a foundation of philosophy, computer science and mathematics we will be able to really critique the different digital philosophical arguments. In essence, we are going to have to take apart the common arguments found in DP. This deconstruction is necessary so that when the dust settles we can look at the rubble and pick out the useful parts to see what value is contained in this rather obscure part of science and to see what avenues are worth pursuing.

The works on digital philosophy that we will be discussing are all serious academic efforts that tried to approach physics from a new angle. These revolutionary ideas are often defended by their authors with references to scientific breakthroughs in the past that were also regarded to be outlandish but are mainstream science today. Today digital physics is pretty outlandish, but at the end of this thesis I will hopefully have convinced you that the whole existence of this discipline is in itself fascinating. Not necessarily because of the conclusions that digital physics arrive at. But because these conclusions are drawn by applying the theories of one discipline to the other. Apparently we can think about the world as a simulation when looking at it purely computer scientifically. But computer science is just a different branch of mathematics, which branches from logic. And ultimately contemporary physics is rooted in that same logic. It is fascinating that the world can be looked at from a whole different angle using the same basic principles. In the end this tells us more about science itself, and about how we as humans try to make sense of this weird thing called reality.

Chapter 1

Digital Physics: An Overview

In this chapter I am going to survey some authors in the field of digital physics. The main focus of this chapter is to provide the reader with a sufficiently detailed summary of the field of DP in order to process the contents of the following chapters. Besides this explanatory function, the goal is also to categorize the standpoints of the authors regarding the ontological status of information and computation. Since a lot of publications have been made on DP, the aim of chapter 1 is therefore to explain key concepts in the field using the most striking and relevant publications.

1.1 Konrad Zuse: The principles of Digital Physics

The main work of Konrad Zuse: *Rechnender Raum*, 'Calculating Space', was written in 1969 and contained all of Zuse's ideas regarding computation. Zuse has been credited as one of the first people to have come up with an information and computation centered idea of physics. However his work was only discovered after the theory of DP was already well established by Ed Fredkin (who will be discussed later in this chapter).

Konrad Zuse was a German civil engineer who worked for the Henschel Aircraft Company in the late 1930s and in the early years of World War II. His displeasure with the need to perform huge amounts of calculations by teams of human computers lead him to build a computing machine; the Z1, between 1936 and 1938. This machine is considered to be what we can now call the first computer. The Z1 was a mechanical machine and so in 1940 he improved on his design by building an electrical version; the Z2 and in 1941 the Z3. This machine was destroyed in 1944 during an air raid. A reconstruction of the machine was made in 1961. After the war Zuse constructed the Z4 computer, which became the world's first commercially used computer. Zuse is also credited with writing the first programming language for his computers: *Plankalkül*, meaning calculus for programs. As an engineer by trade, Zuse was mainly focused on building a working machine that could perform calculations. He was not at all concerned with the theory of computation, which at that time was explored by renowned mathematicians in the academic world like Schönfinkel, Church, Post, Kleene, and Turing. Because of this, Zuse was not an established name in the academic community as a theoretical computer scientist. It is known however, that Turing and Zuse were familiar with each other's work (Zenil, 2012). Although Zuse has always been more characterized as an engineer than a theorist. This is perhaps one of the reasons why his main work on computation in physics has been relatively unnoticed in the academic community for so long.

Despite having been overlooked by the initial founders of DP, *Calculating Space* is still a great introduction to some of the basic concepts that later became central concepts in DP. In the introduction of *Calculating Space* Zuse starts off by describing the current state of science, where a close interplay exists between physicists and mathematicians. This interplay is the reason theoretical physics and experimental physics have become so successful at developing the modern accepted theories, even though these are very mathematical in nature. Zuse points out that the interplay between theoretical and experimental physics was only possible because of the experts on data processing theory. It was the data processing analysts who managed to accelerate and improve numerical calculation in order to be able to verify or disprove claims of both theoretical and experimental physics. This raises the question, according to Zuse, whether data processing is merely an effectuating part in the interplay between theory and experimental results. Or if the ideas in data processing are also successfully applicable to physical theories (Zuse, 2012, p.1).

In order to bridge what Zuse called information processing theory (but what we now call computing science) with physics he needed to establish some base observations. Zuse's goal was to place information at the center of physics. For this he turned to the most common way of representing information in computer science: Boolean algebra. In Boolean algebra information is defined by elementary true/false values called bits. The bits can then be manipulated by elementary operator,s namely: conjunction, disjunction and negation (Zuse did not mention the fourth elementary operator in Boolean algebra: identity). A computer program is generally considered to have an input and a set of rules to produce an output. This classical notion of computer programs, or algorithms, is not broad enough to cater for an information centered notion of physics, however. Nature generally does not have an input, and nature does not have an output either. Stretching the classic notion of algorithms, one can add the term *state* to imply everything that the computing machine currently has stored in memory. This means that the output now depends on what the algorithm does with both the input and the state. Consequently, the algorithm can also alter the state. An algorithm that produces no output and only has a single state as an input is called a cellular automaton. By definition it endlessly applies its algorithm to the stored state and then replaces the stored state with the resulting new state. Zuse coined the term "automaton theoretical way of thinking" with which he denoted any form of technical, mathematical or physical model that referred to a lapse of states which follow a predetermined set of rules (Zuse, 2012, p.5).

Given an initial state and a set of rules, the cellular automaton described by this is determined. This means that by definition the cellular automaton will behave in the same way every time it is ran. Zuse remarks that given a rule set of a cellular automaton, a graph can be constructed that shows the flow of different states when they resolve into one another. Every state can only have one succeeding state given a determined cellular automata. It is for this reason that every cellular automaton will eventually resolve in either a cyclic succession of states or in a single stationary state that resolves into itself [1].

As such, Zuse suggests that looking at physics in an automaton theoretical way might provide new revolutionary theories and insights. The main point of contention between this idea and modern physics is that the automaton theoretical way requires us to view nature as a series of discrete states; it requires us to consider a bit of information as an elementary particle. Modern physics is still on the fence about the discreteness of nature. Zuse does not deny the fact that modern models of physics are continuous[2]. He even states that there do not seem to be limits

[1] For more information on cellular automata see appendix B or Berto and Tagliabue (2017).

[2] Throughout this thesis the reader will find the terms 'analog' and 'continuous' and 'discrete' and 'digital' to be somewhat interchangeably used. This is because 'analog' and 'digital' are terms widely used in computer science and engineering while 'continuous' and 'discrete' are more often found within mathematics and physics. When it

or thresholds in physics besides the speed of light (Zuse, 2012, p. 15). It is not necessarily the continuous aspect of physics that is the main point of contention for Zuse, as he considers it possible to build analog computers, albeit considerably less powerful ones. It is the fact that all computers seem to have constraints on minimum and maximum values that concerns him the most. Given an infinite amount of time, a digital computer can very closely approximate a continuous function. But of course, in the real world at some point the physical constraints of the computing machine come into play.

Entropy is also a physics term highly focused on by Zuse. In chapter 2 he creatively plays with the idea that an increase in entropy in modern physics is defined by probability laws over slight deviations from classical mechanics, while in his model the increase in entropy is a result of calculation errors, or rounding errors (Zuse, 2012, p.18). He also observes that the information entropy[3] in a cellular automaton does not increase while running the calculation. However, the person running the calculation will receive a greater information value, otherwise he would not have needed to run the calculation in the first place. In other words, there is always a reason some person runs a calculation, namely to further his own understanding: thus increasing his information value. The entropy in the entire universe has therefore increased. Information entropy is a common theme that bridges contemporary physics with digital physics and it is astonishing that Zuse already mentioned his ideas on entropy.

Calculating processes that are based on classical Boolean algebra are generally not reversible. Even the simple example of a disjunction can be considered as a proof of this. "If A or B then C" is not a reversible operation: If C is encountered, one can only know that either or both A or B were true, but the exact state is not known. The concept of reversibility is very important in physics and therefore the notion of reversibility of cellular automata is an interesting idea to follow. Zuse confirms that an automaton is determined in the forward time direction while not being determined in the other direction (Zuse, 2012, p.50). If a cellular automaton would be able to simulate physics, it would mean that an (almost) infinite amount of information needs to be stored to be able to reverse the operation. Generally speaking this means that a reversible cellular automaton is mathematicaly speaking perhaps not even definable (Zuse, 2012, p.51). As a solution, Zuse is open to probability playing a factor in a cellular automaton representation of physics but leaves that question as a philosophical one without a definite answer (Zuse, 2012, p.53).

Zuse was bestowed with some incredible foresight as he treated almost all aspects of what

comes down to it they all refer to the same idea, namely that analog/continuous phenomena have an infinite amount of intermediary values between two given values, while discrete/digital phenomena are indivisible.

[3]Information entropy is analogous to the term entropy used in statistical thermodynamics; it is the measure of disorder in data Shannon (1948). If an information source produces information from a low-probability event, the information content of this data value is higher

later became known as DP in his book *Calculating Space*: from the core concepts of information and information processing to cellular automata and the problems and contentions between this approach and the modern models of physics. Infinity, entropy, reversibility, probability, and time all have major implications and issues in DP. What remains is how we can classify Zuse and his ideas regarding the ontological status of information and computation. German and Zenil wrote the afterword of the translation of *Calculating Space* and state that "[...] Zuse did not hit upon the concept of universal computation, he was interested in another very deep question, the question of the nature of nature: 'Is nature digital?' He tended toward an affirmative answer, and his ideas were published, according to Horst Zuse (Konrad's eldest son), in the *Nova Acta Leopoldina* (**?**, p.60). This makes it appear as if Zuse believed that nature was at its core just information and computation: a digital view of nature. This is a very strong ontological view. I have not been able to find the literature that Adrian German and Hector Zenil were referencing, but basing myself solely on *Calculating Space* I would strongly disagree with their characterization of Zuse. In *Calculating Space* Zuse tends to focus heavily on the limitations that computation would face compared to modern physics. Even in his introduction, he clearly explains how the data processing and automaton theory aspect of physics and mathematics is merely a fruitful alley to explore. At most he says that nature can be described as a computation and he even takes caution with that statement. It would be a very big leap from this very nuanced perspective to a hardline standpoint regarding the position of information and computation in reality.

1.2 Jürgen Schmidhuber: Applying computer science

The title of Jürgen Schmidhuber's paper could not have been a clearer description of its contents: *A Computer Scientist's View of Life, the Universe, and Everything*. Published in the journal *Foundations of Computer Science* in 1997, the paper is even written in the style of a computer science paper. The content, however, is highly metaphysical. The central question that Schmidhuber puts forward is "Is the universe computable?". Nevertheless, in the preliminaries it is immediately assumed that the universe is a result of all possible programs ran by The Great Programmer on His Big Computer. From this assumption, Schmidhuber sets out to make numerous interesting observations.

The first observation is that a Turing Machine (TM) with input symbols "0", "1" and "," can calculate a *universe*, in the sense that we simply declare the result of a calculation to be a universe, given any input[4]. The output of the program is a comma separated list where each segment describes the evolution of that universe. Some inputs return finite outputs and thus finite time universes, other programs do not halt and therefore produce infinite time universes.

[4] An in depth explanation of Turing Machines can be found in appendix A

Schmidhuber shows that calculating all possible universes is relatively easy by applying dovetailing to computing universes. Dovetailing is a concept in computer science that describes how algorithms can be catered to be applicable on hypothetically infinite data. By performing one step of the algorithm on an element of the input on every second step, and every step for every next element in the steps in between, theoretically all elements are handled. What is noteworthy here is that computation time is not an issue to The Great Programmer. The process of computing all universes does not need to finish within a certain amount of time. Likewise it is not an issue for any possible observer "in" a computed universe, as for him every time evolution is sequential and does not reside in the same time dimension as The Great Programmer. A delay between evolution steps is not "noticed" (Schmidhuber, 1997, p.2). Not just time, but even determinism itself can be perceived differently by inhabitants of calculated universes and The Great Programmer according to Schmidhuber. Observers in universes may observe unpredictable behaviour but to The Great Programmer the predictability may seem obvious. As for Him[5] the collection of universes seems logical: the greater picture is more easily to be perceived as determined while the individual universe may seem *random*[6].

With universes defined in a computational way it is now possible to apply actual computer science theory to them. Randomness, initially a concept with different interpretations in physics, computer science and philosophy become unifiable across all disciplines. Schmidhuber states that the longer the shortest program computing a given universe is, the more random it is (Schmidhuber, 1997, p.3), This idea may be obvious to computer scientists but it is an idea that perhaps needs further explanation for others. The key concept at work here is *Kolmogorov Complexity*. Kolmogorov Complexity is a mathematical definition of randomness. It states that the randomness of a certain string of bits is defined as the length of the shortest program that outputs these bits. This taps into the feeling that we have when we talk about randomness. Something that may appear random might actually have an obvious cause, or in this case a program defining it. If the length of the shortest program is very short, the perceived randomness disappears as we see that our string is constructed using some pattern. However, if the length of the shortest program for a certain string is (almost) as long as the string itself, then we see that there is no more logic behind the sequence of bits than that what we initially perceived; it is perceived as *more random*. The thing that makes this mathematical definition so interesting for us is that the Kolmogorov Complexity of an arbitrary string is an incomputable function. This is computer science jargon for saying: "We have proven that we cannot instruct a computer how to perform this task". It is not just that we do not know how to tell a computer something. It is stronger than that: we *know* we cannot instruct a computer how to do something. The incomputability of the Kolmogorov Complexity means that we actually cannot let a computer decide how random a

[5] Schmidhuber does not mention God anywhere in his work, but does use reverential capitalization in order to signify The Great Programmer.

[6] We will find big picture vs. small picture universes to be a recurrent theme in digital physics

string is. This is where the philosophical question arises: "If we cannot tell a computer how to decide this, can we still know ourselves?". This notion of Kolmogorov Complexity stands at the basis of Schmidhuber's point. If the length of the shortest program computing a given universe is still quite long, it has to be more irregular and thus random. If it is very short it should appear more regular and predictable.

Schmidhuber links the incomputability of randomness of strings to the incomputability of randomness in the universe by recalling that the universe can be represented as a string. Given this proposed link between randomness in computer science and randomness in universes, Schmidhuber sees it as obvious that our physicists cannot expect to find the most compact description of our universe. This is because the Kolmogorov Complexity, the length of the shortest description of a certain unit of information, is incomputable. This does not mean that it is impossible to find a short and compact description of our universe. It does mean that finding one would have to be the result of trial and error and it means that one can never be sure that the actual shortest description has been found. This also correlates with the idea that our inability to represent our universe as a state implies *true randomness*. It all comes down to the fact that randomness is not a computable function: we can never know for sure if something is actually random.

After these very interesting observations about randomness, universes and our ability to know something about them if a universe is hypothetically represented as information, Schmidhuber moves on to question whether or not our own universe is the result of a computation by a Great Programmer. The first clue that brings Schmidhuber to say that the universe could indeed be the result of a computation is the fact that the history of our universe seems to be regular. Our universe seems to be governed by laws and we do not see a high amount of irregularity or deviation from these laws. In Schmidhuber's view this gives credit to the idea that a short algorithm is responsible for calculating our universe. Interestingly enough, Schmidhuber also decides to settle 2000 years of Western philosophy around the mind-body problem in 7 sentences. He does this by saying that in the eyes of The Great Programmer even this whole discussion itself, the sound waves of words, the mental states connected to them, are merely irrelevant sub patterns, or emergent patterns rather. By stepping back and adopting the Great Programmer's point of view, classic problems of philosophy go away (Schmidhuber, 1997, p.8).

1.3 Edward Fredkin: Proving pancomputationalism

Finally we have arrived at the birth of digital physics as a name given to a subset of science. *An Introduction to Digital Philosophy* is the overview of all the work that Edward Fredkin has performed in the field of DP in the 1990s and early 2000s even though Fredkin states that he started to think about DP as early as the mid 1950s. Edward Fredkin is a renowned academic who, amongst other positions, was the Director of the MIT Labratory for Computer Science. Fredkin's goal is arguably distinct from the previous authors. Where Zuse explores information processing theory as a way to give new insights in physics and where Schmidhuber (mostly) hypothetically explores what it would mean for a universe to be computationally representable, Fredkin's goal is to prove that our universe is actually the result of a computation, an idea that is also called pancomputationalism. Throughout the introductory pages of his article he provides arguments to back-up this claim. The rest of his article can be described as a mapping between physics and DP, indicating how every concept in physics should be translated into DP's automaton theory.

Mostly unknowingly, Fredkin follows in the footsteps of Zuse in defining a cellular automaton with bits being the fundamental units of physics. He defines three main heuristics: *simplicity*, *economy*, and *Occam's Razor* to be the guide for which all aspects of DP should follow. A cellular automaton was chosen as the model for phyics because of its simplicity and efficiency of representing a discrete physical-temporal relationship (Fredkin, 2003, p.190). It is also for this reason that he chooses a binary cellular automaton rather than a 3- or 4-state automaton. He does not exclude that in the future a different configuration state from a binary state might provide a simpler theory, but right now there is not a reason to do so. Fredkin's goal is to map (3 + 1)-dimensional space-time, CPT invariance, Plank's constant, the speed of light, the conservation laws, discrete symmetries, and known properties of particles into a DP theory. He admits that he has not found a way yet to completely formalize the current state of contemporary physics in his model, as he states that he failed to add continuous symmetries to this model among other things.

DP is discrete. This is the main reason that it is nearly impossible to incorporate continuous aspects of physics into it. Fredkin also chose to represent a mere particle model of physics in DP, having to disregard fields and waves. According to Fredkin "The principle of simplicity has driven us to reluctantly make a decision—in this paper DP is a particle model and all processes in DP are consequences of the motions and interactions of particles." (Fredkin, 2003, p.192). It is unclear whether a more complex discrete model that unifies particle and wave physics exists. In the end the discreteness of digital physics represents atomism, a philosophy that was mostly made famous by Epicurus, in its most extreme form. Where in Epicurean atomism nature was quantized in time and space, Fredkin's digital physics describes that the quantized spatio-temporal units also have a quantized state: namely on or off.

Besides the problem of analog physics being extremely hard to describe in DP, Fredkin is also faced with the same problem as Zuse, namely: reversibility. Luckily for Fredkin, by the time he developed his theory, automaton theory had become better equiped to tackle this problem. In particular Reversible Universal Cellular Automata (RUCA) were developed, initially as a new set of automata called *block cellular automata* (Toffoli & Margolus, 1987), but later Fredkin among others proved that any cellular automaton is actually transformable into a reversible cellular automaton (Margolus, 1984). In principle this is done by taking the second-order history of the cellular automata into account. The state of a cell then not only depends on the neighbours on timestep *t - 1* but also on the state of the same cell on timestep *t - 2*. This was a huge step for DP as it was now able to incorporate the notion of physical reversibility.

So far we have seen problems and solutions coming up and being resolved as soon as physics needs to be squeezed into DP. Thus far this has only been representational: looking at the world from a DP model does not necessarily require the universe to ontologically be a computation. Fredkin does seem to want to make the ontological leap, as he brings forth proofs that suppose to give "reason to believe that underlying physics there is a digital representation of information" (Fredkin, 2003, p.200). To this end, Fredkin first needs to prove that the universe is computationally universal. Computational universality is the computer-scientific concept that relates one computational device to any other: A Turing Machine is computationally universal, or Turing Complete, if it is able to simulate any other Turing Machine. In order for it to make sense to describe the universe as a computation, the universe needs to be proven to be computationally universal.

> With regard to computational universality there is a way to experimentally verify whether it is true of physics. In automata theory, we prove that a system is computation-universal by demonstrating the possibility of constructing a universal machine within that system. If, in our world, we can build and operate even one universal computer, then that is hard experimental evidence that physics must be computation-universal. This experiment has already been done and verified [Every human being and every personal computer meets the standards set by automata theory for being recognized as universal computers!]. To prove the converse, we would have had to demonstrate the impossibility of constructing a universal computer[7] (Fredkin, 2003, p.196).

[7] I am aware that this reasoning may immediately illicit some criticism. Rest is assured to the reader that this argument specifically will be dealt with in chapter 3.

Finally, Fredkin also brings forth what in his eyes are "the biggest clues for DP to be fundamentally how the universe works": namely the existence of small integers in nature. He regards all small integers that seem to be constants in physics as a clear hint towards the digital nature of reality. For the sake of the audience judging this argument for themselves I will provide the full list of Fredkin's small integer clues:

- Number of spatial dimensions: Exactly three
- Number of directions for time: Exactly two, forwards and backwards.
- Number of chiral parity states: Exactly two, left-handed and right-handed.
- Number of different electrical charge states: Exactly two, + and −.
- Number of CPT modalities: Exactly two out of eight, CPT and C − P − T.
- Number of measurable spin states of an electron: Exactly two, up and down.
- Number of spin state families: Exactly two, bosons and fermions.
- Number of particle conjugates: Exactly two: particle and antiparticle.
- Number of leptons or quarks per generation: Exactly two.
- Number of lepton and quark generations: Exactly three.
- Number of different QCD color charge states: Exactly three, R, G, and B.
- Spin of any particle that is a boson: Exactly n (n always a small integer).
- Spin of any particle that is a fermion: Exactly n + 1/2.
- Maximum number of inner-orbit electrons in an atom: Exactly two.

(Fredkin, 2003, p.200)

As already mentioned, Fredkin continues his article with a mapping of most concepts of physics to DP. This mapping is unfortunately less relevant to the content of this thesis. Fredkin often uses examples of outcast physicists coming up with new, widely rejected theories to defend his pursuit of DP. He fires back at conventional physics by saying that the amount of unanswered questions and unexplained phenomena is embarrassing (Fredkin, 2003, p.244) and that he is as justified in looking at physics from his perspective as regular physicists are at looking from theirs. He strives for a model of physics that has no unanswered questions and ultimately will be consistent with common sense. This, for him, is the goal of DP.

Fredkin's contribution to DP as a whole is substantial, with it basically laying the foundations of some key concepts. He emphasizes the cellular automaton model of physics, where others like Schmidhuber do not necessarily commit to this model. Even till this day Fredkin enjoys a certain following with this branch of digital physics. A cellular automaton view of physics seems very attractive to some physicists: even Nobel laureate Prof. Gerard 't Hooft relatively recently published a work advancing a cellular automaton model of quantum physics ('t Hooft, 2016).

1.4 Max Tegmark: Everything is mathematical

Fredkin's desire to have a physical model without unanswered questions is shared by Max Tegmark. The 2008 article by Tegmark, *The Mathematical Universe*, was not primarily written as an article in the DP tradition. The main thesis of the article is highly metaphysical and originates from the mathematical corner of science, rather than from computer science (Tegmark, 2008). Nevertheless, the implications of his metaphysics are comparable to DP and he does not leave DP untouched. The broader, mathematical, starting point of this thesis allows us to explore the implications we have seen so far even further. Not necessarily putting the emphasis on information as the building block of reality, like we have seen so far, but focusing on mathematics, information still inevitably becomes a relevant aspect of Tegmark's work.

Tegmark starts of with some relevant definitions. He defines The External Reality Hypothesis (ERH) as the idea that an external physical reality, completely independent from human beings, exists. Furthermore, he defines the Mathematical Universe Hypothesis (MUH) as the idea that reality is actually a mathematical structure. Tegmark argues that ultimately a Theory of Everything (TOE) in physics should be *baggage free*. With this he means that physics is doing great at describing better and better how the external reality works, but that it still needs *baggage* concepts to describe what it is. For example, we name different particles and forces in order to distinguish them but we are not able to explain what they ontologically are. Every evolution of physics merely dethrones the particles and forces that were held fundamental in favour of others. This is what Tegmark identifies as baggage, the fact that we need these inexplicable things in order to describe anything in physics. He points out that a baggage free TOE necessarily is a purely mathematical TOE. He then invokes the concept of mathematical identity and isomorphism to argue that if a TOE is purely mathematical, reality must by definition also be ontically mathematical. This leads him to conclude that the ERH implies the MUH. If there is an external reality, then in order for it to be completely describable without baggage, it must be a mathematical one.

This argumentation requires a very strong standpoint regarding the philosophy of mathematics by Tegmark, given how much value he attributes to it. Many have criticized Tegmark on the ground that mathematics is constructed by humans, able to describe physical phenomena well, but ultimately not able to receive any predicate of reality (Jannes, 2009). Tegmark is clear about his philosophy of mathematics being essentially very Platonic in nature. He argues that mathematics is not built by mankind but rather uncovered and hence that it is not a tool we constructed ourselves but a jigsaw puzzle we are slowly discovering the pieces of. Tegmark refers to the already proposed mystery behind the power of mathematics in the natural sciences (Wigner, 1967) to back up his view that mathematics is the fundamental object. When the universe is mathematics, it would perfectly explain why it is so well descibable by mathematics.

An important concept in Tegmark's work is the distinction between the bird and the frog. With these zoological metaphors he distinguishes the outside observer's perspective of a universe with the inside observer's perspective. The bird is seemingly more omniscient and can view the universe completely, whereas the frog is part of the universe and is constrained by the observation laws that apply. Similar ideas about different observer perspectives have been around for a while, according to him. Tegmark uses the perspectival differences to dive into the complexity of the universe. Given certain clues about the complexity of a universe, one can form an image of what a theory of everything might look like. First, he states that the bird and the frog can each independently see the universe as either having low complexity or high complexity. This results in four distinct cases. Tegmark argues that it is unlikely that the frog's perspective (i.e. our perspective) of the universe is low in complexity. This is mostly from a personal viewpoint (Tegmark, 2008, p.12), as in his eyes a theory of everything would already have existed if nature was not complex That the universe seems very complex for us is thus a given. Now what remains is either the possibility that the universe, from the bird's perspective, is either truly highly complex, or not that complex at all. If it were highly complex then the project of physics is practically doomed according to Tegmark. It is therefore hoped that the actual complexity is very low. In that case, the mathematical structure describes some form of multiverse. This conclusion is reached from the reasoning that an entire ensemble is often much simpler than any one of its members (Tegmark, 2008, p.12). This reaches back to the intuitions already brought forward when Kolmogorov Complexity was discussed. The algorithmic information content of a long bit string can be low if it is easily describable by a short algorithm. Tegmark works with the idea that a short algorithm is capable of producing individually complex universe while the total multiverse is simple. This argument is similar to the argument we have seen Schmidhuber make.

> The apparent information content rises when we restrict our attention to one particular element in an ensemble, thus losing the symmetry and simplicity that was inherent in the totality of all elements taken together. The complexity of the whole ensemble is thus not only smaller than the sum of that of its parts, but it is even smaller than that of a generic one of its parts (Tegmark, 2008, p.12).

The most intuitive example that explains this idea further is π. There are multiple ways to define π. We could see it as a number, in which case if we desired to write it down it would be infinite. We could split this number up in huge but finite chunks and present them to people individually, in which case the chunk would seem random and maybe highly complex. But we, performing this act of calculating π and breaking it up in chunks know that we are dealing with a number that has a very short mathematical definition. In the same way Tegmark argues that we are only seeing a very complex universe that if it were to be just a single chunk of a multiverse, might actually be quite simple to understand.

Touching on the subject of DP, Tegmark also discusses the notion that his mathematical universe is perhaps a computed one. He directly argues with Schmidhuber and Fredkin here on the notion of time evolution. As we recall, Schmidhuber saw a computable universe as a TM returning a comma separated list of bit strings, each bit string being the time evolution of the previous one. Fredkin saw the standard evolution of cellular automaton as the principle of time evolution in a computed universe. Tegmark's universe is not a working, evolving universe in this sense. He opts rather that a computed universe is a describable one. First of all he argues for his case on the basis that this linear time-step evolution is (understandably) the result of classical physics and that new revelations in contemporary physics are not compatible with this linear evolution. Given this, he then argues that space-time is *precomputed* and already stored somewhere. Each 4-dimensional time slice can be requested and read out if so desired. In this way a computer program describes the universe at the moment when it is required, instead of it being a program that spits out data ad infinitum. The idea that the computer program describing the universe is merely a program that can retrieve answers about a certain state at a certain time rather than actually performing an ongoing simulation seems anti-intuitive. To this Tegmark responds: "Since every universe simulation corresponds to a mathematical structure [...] does it in some meaningful sense exist 'more' if it is in addition run on a computer?" (Tegmark, 2008, p.19).

The discussion above should not be confused with Tegmark's definition of the Computable Universe Hypothesis (CUH). With this he doesn not refer to the idea of digital physics which was already part of the discussion, but he further refines the Mathematical Universe Hypothesis by requiring physical reality to be definable by computable functions. Computable in this context means that the function describing the universe should eventually halt when ran on a Turing Machine[8], contrary to Schmidhuber's definition where non-halting algorithms can describe universes just fine. The fact that Tegmark considers his universe to be described by an algorithm requires him to only consider non-halting, computable functions. His computation needs to halt in order to be compatible with the definitions he set out for time that we saw in the previous paragraph. A consequence of the CUH is that it makes the universe discrete because a continuous universe inherently contains uncomputable functions. The continuum and computability tend not to mix well because a continuus value can theoretically never be the output of a computation that halts. If the computation halts, the value becomes discrete. We see that different definitions of computable universes result in different computational requirements.

Tegmark's metaphysical view is clearly distinguishable from that of previous authors but also bears similarities. Tegmark clearly works from a purely mathematical standpoint: conjecturing that a solid Theory of Everything is required to explain everything and all and in order to do this, must be mathematical. This is only possible if the universe itself is purely mathematical. This

[8]For more information on Turing Machines and Halting see appendix A

has the consequence of supporting a multiverse theory. This is more radical reasoning than we have seen with Schmidhuber and Fredkin. Tegmark also disagrees with them on what it means for the universe to be computable. Where Schmidhuber and Fredkin favour a more classical time evolution approach, Tegmark deems time evolution in itself unnecessary. The common dividers are still the approach of physics and the universe itself as mathematical structures and, most of all, as discrete and information based.

1.5 John Archibald Wheeler: Its and bits

Out next author may seem to be an outlier when it comes to DP. Especially given his remarks in the paper I will discuss, John Archibald Wheeler would probably never have declared himself to be a digital physicist. Yet, he is referenced many times in DP literature (Zenil, 2012, Various chapters). This is because Wheeler famously puts the bit (the information) above the it (the object). Later in this thesis we are going to see if authors of DP are justified in seeing Wheeler as an intellectual ancestor. Wheeler was a renowned physicist in the 20th century as he taught at Princeton and famously coined the term *wormholes* (Jones, 2014). Later in his life Wheeler spent more time thinking about the philosophical aspects of physics. In his 1989 article *Information, Physics, Quantum: The search for links*, Wheeler layed out his philosophical considerations of physics.

Wheeler's paper has quite a remarkable structure, in the sense that he is really radical and upfront with his ideas. On the second page he lays out his main three questions and explains his main principles: four no's and five clues. It is easy to note that the three questions; "How come existence?", "How come quantum?" and "How come 'one world' out of many observer-participants?" are quite metaphysical in nature. All of the following should therefore also be regarded as a metaphysical critique of modern physics. At the center of this critique, Wheeler places information as the fundamental unit of physics. He poses that rather than a real and material observable world it is only observation itself that should be of scientific study.

> It from bit. Otherwise put, every it — every particle, every field of force, even the space-time continuum itself — derives its function, its meaning, its very existence entirely — even if in some contexts indirectly — from the apparatus- elicited answers to yes or no questions, binary choices, bits. It from bit symbolizes the idea that every item of the physical world has at bottom — at a very deep bottom, in most instances — an immaterial source and explanation; that what we call reality arises in the last analysis from the posing of yes-no questions and the registering of equipment-evoked responses; in short, that all things physical are information-theoretic in origin and

this is a participatory universe (Wheeler, 1989).

The participatory universe refers to Wheeler's Participatory Anthropic Principle (PAP). What he means is that "No phenomenon is a real phenomenon until it is an observed phenomenon"[9]. This is both a fundamental statement about (quantum) physics and a philosophical position. He places observation above reality; it is observation that constitutes anything that is thought to be real. From a philosophical viewpoint I can classify this as philosophical idealism; reality is primarily mentally constructed. The question then arises if we can classify Wheeler's idealism as either subjective or objective idealism. In this distinction, subjective idealism is mainly a form of immaterialism. It poses that minds and the mental are the only things that exist. George Berkeley is the main philosopher credited with this viewpoint but it was later also extended and serves a noteworthy place in the history of philosophy. In the words of Berkeley: "Esse est percipi", To be is to be perceived. Objective idealism is more nuanced in its view. It still rejects naturalism, the idea that minds and mental states are in essence a materialistic phenomenon, but it does not reject the view that material objects do exist. Wheeler's flavour of idealism becomes more clear when he elaborates on some of his principles. He clearly does not like what he calls the tower of turtles, most likely a reference to Terry Pratchett's book Discworld in which is explained that the earth is carried on the back of four elephants who in turn are supported by a turtle flying through space. Wheeler sees in any physical law a reliance on a different law or idea which always results in an infinite regress e.g. "what was there before the big bang?". There should be no framework of ideas underlying another which is itself again underlying another. The problem of the infinite regress, already known to Aristotle, seems like an inescapable logical annoyance unless it is resolved by posing that the regress is either finite or infinite or circular. Wheeler opts for the third option:

> To endlessness no alternative is evident but loop, such a loop as this: Physics gives rise to observer-participancy; observer-participancy gives rise to information; and information gives rise to physics (Wheeler, 1989).

To Wheeler, it is not either the mental or the material that is the only thing 'real', it is the bridge in between; information. Information is both the manifestation of physical phenomena which are only real when they are observed and the mental state of the observer. For this reason it would be justified to put Wheeler's idealism in both categories of idealism as well as neither, or not call it idealism at all. The Participatory Anthropic Principle is a step towards anthropic centeredness of metaphysics (idealism) but also a step in the opposite direction, where it is not the mental state of the observer that is real, but the information that correlates between the material and the mental.

[9]The idea of observed phenomena is also prominent in the work of Niels Bohr whose influence will be discussed later in this thesis

Interestingly enough, Wheeler seems to reject the idea of the universe as a machine. He rejects this notion as it would conflict with his principle of no infinite regress. "A machine is as justified in producing the universe as a big bang or Swiss watch-maker" (Wheeler, 1989, p.314). If we would be a computation on a machine, then in what physical reality does that machine reside? But for Wheeler the same can be said about physical laws in general, he tends to deny those as well. Nevertheless Wheeler's philosophical work shares a common theme with other authors that we discussed, with information being the key shared concept. But compare Wheeler's view with for example Tegmark's view and we will find that they are eachother's philosophical opposites. For Wheeler, the act of observing information is fundamental whereas for Tegmark, the hypothesis of an external reality is the basis of rationality, and so he concludes that reality is mathematical and arguably informational.

1.6 Summary: The tradition of Digital Physics

Through these five authors we have gained an overview of what digital physics is about. Information has become the most important building block of reality, rather than anything else. Where information is the building block, computation can be seen as the driving force. In the DP take on physics new problems arise; the discrete or continuous nature of reality, infinities, and time evolution. Perhaps a more fundamental problem is observation; what does it mean to see a universe evolve and what perspective (say bird or frog) matters? Given the time period during which DP lifted off as a theory researched by a multitude of researchers, one can also think of DP as a take on physics in the light of new discoveries in quantum physics. The same ideas as in digital physics: continuity, time, infinity, observation, play a central role in debates that emerged after quantum physics was established. Some authors see DP as a way out of the weird paradoxes and anti-intuitive nature of modern quantum physics (Zahedi, 2015). Mueller (2017) is inspired by DP and uses key concepts of it to propose a physics from an observer perspective using algorithmic information theory. Yet some others even interpret DP as the analogy between quantum computing and the quantum nature of physics (Lloyd, 2006). The differences in the takes on DP between these authors are clearly vast. The most accurate description of DP can therefore only be the common act of unifying physics with computer science and philosophy and creating a narrative around computation.

The authors in this chapter have different philosophical takes on how DP should be interpreted. Some see it as a mere alternative model to describing the universe, others see it as proof that the universe *is* a computation. This is one of the few clear distinctions we can make. Henceforth we will dub the ontological claim that the universe is a computation *strong DP*. This is in contrast with *weak DP* which merely states that physics can be *described* in a computational way. Where Zuse and Schmidhuber can be classified as mainly probing into the ideas of weak DP, Fredkin and Tegmark arguably propose strong DP. While at the core, all of the authors, including Wheeler, try to view information as the main object of research.

After reading this chapter the reader is justified in being confused about the leap from cellular automata, information and algorithms to actual physics and the nature we see around us. DP appeals to the concept of emergence to solve this problem. As emergence is not further explored in this thesis, more background on emergence can be found in the works of Prokopenko, Boschetti, and Ryan (2009), Humphreys (2016) and Yates (2017). An even greater variety of standpoints within digital physics itself can be found in Zenil (2012).

Chapter 2

The Concepts

As digital physics was being reviewed in section one, the reader might have already been infuriated with objections against it. Or not, in which case the latter part of this thesis will be an even better read. A lot of concepts, of which some are new to philosophy, play an important role in the discussions on digital physics, and anyone making an argument within digital physics equally has some (hidden) opinions and assumptions towards these concepts. I find it therefore vital that I will discuss these concepts before I move on to a full criticism of digital physics. The reader may not always agree with me on the definitions and arguments that I will make in this chapter. Nevertheless, the point of this chapter is also to split the discussion of definitions from the deeper philosophical reasoning about digital physics. Epistemology and Metaphysics are deep-rooted topics within philosophy which need such a clarified standpoint, while computation and discreteness are newer topics which have not found their way into philosophy yet. I am therefore also delighted to give these latter concepts a more philosophical depth in this chapter.

2.1 On Computation

Together with information, computation is the central defining aspect of DP. Information is described as the fundamental, indivisible object in the universe, while computation is described as the force that gives it motion. We have seen how Zuse, Schmidhuber and Fredkin specifically used computation to explain time and we have seen Tegmark using computation to strengthen his mathematical universe hypothesis with the computational universe hypothesis. The term 'computation' is typically used in the mentioned literature in the way it has been used in computer science, from where it originates. However, in this more philosophical context it deserves a dedicated deconstruction.

The theory of computation initially developed as a branch of mathematics in the early half of the 20th century. The goal during this period was to provide a formal definition of computation as part of the bigger context of foundations of the mathematics. Multiple models of computations were defined to this effect. Examples of these models are the λ-calculus, recursive functions, cellular automata and Kahn process networks. But the most popular model of computation in literature today by far, is the Turing Machine (TM). Alonzo Church and Alan Turing managed to prove that the λ-calculus, recursive functions and Turing machines are mathematically equivalent (Turing, 1937). This has lead other mathematicians to postulate the *Church-Turing hypothesis*, which states that any function that can be regarded as computable can be computed on a Turing machine (Deutsch, 1985). This is considered to be informal and as such is not provable. It merely captures the idea that what we call 'computation' is correctly formalized by a Turing machine. The Church-Turing thesis is widely accepted nowadays by the scientific community. It is for this reason that we find all the literature in digital physics (and outside of it) using the Turing machine formal definition of computation.

Several authors have pointed out (Timpson, 2013) (Copeland, 2020) that Turing provided us with a formalization of the informal process of computing, namely: "a man in the process of computing" (Turing, 1937). These authors emphasize that Turing gave us a formal way to represent a mathematician with an infinite amount of ink, paper, and time. One should therefore always keep in mind that any proof using Turing machines or any other model of computation actually assumes and is limited by: human capability. We now see that this mathematical 'definition' of computation is itself a definition of the human act of computing. This calls for a more thorough inspection.

In mathematics, functions are considered to be a formal relation between one set of elements to another. The theory of computation seeks to provide the procedure in order to actually find the result of a function if both sets are the natural numbers or sets isomorphic to these. This procedure is often called an *algorithm* and it is here that computation is considered the act of executing an algorithm. In computer science functions and algorithms are often referred to as the solutions of 'problems', but to translate that to intuitive terms algorithms can also broadly be considered to have 'input' or specific 'questions' while the result of an algorithm can be considered to be 'output' or 'answers'.

The term 'computation' is used very broadly in science today. Besides simple notions like regular computers we also see the term used in biological contexts, philosophy of mind, and of course in digital physics. This is the result of a broad interpretation of "What does it mean for something to be computing?". One take on this problem is to consider any physical system that mirrors state transitions that can be formally mapped to be a system that is computing. This take on computation is often referred to as the *mapping account* (Piccinini, 2015, p.17). A broader conception of the mapping account is defined by Piccinini as the *mechanistic account*. His main reason for extending the mapping account is that the mapping account is too restrictive in regard to complex computing systems like quantum computers and (artificial) neural networks. Piccinini explains the mechanistic account as "the manipulation of a medium-independent vehicle according to a rule" (Piccinini, 2015, p.10). This disregards the relationship that computation has with formal function definitions. Rather, instead of being able to understand and interpret the computed functions, the mechanistic account allows natural and biological systems to be called *physical computing* as they may be functionally logic and pattern abiding systems.

Consider which of the following is actually the case: Are Turing machines able to formalize phenomena in the real world, or are real world phenomena formalizable as a Turing machine? The extended[1] interpretation of the Church-Turing thesis determines which of either option one prefers. Piccini favours the latter option as he places the Church-Turing hypothesis at the center of his reasoning. Piccinini seems to want to provide a definition of computation based on all the instances where science uses the term 'computation', rather than providing a definition that is fine whilst excluding some other uses of the term. So for Piccini, all real world phenomena should in some way be formalizable as a Turing Machine. Alistair Isaac had a similar concern with Piccinini's definition, as the one thing that gets lost in the mire of the debate is *meaning* (Isaac, 2018).

When we extend the definition of computation to a mapping or a mechanical account we lose sight of the what and why of computation: the meaning gets lost. Without being able to assign a precise meaning to a computation it is indeed possible to say that anything is calculating anything. Adding meaning, or semantics, is thus needed (Fodor, 1981). If we recall that a computation can be regarded as the steps taken between asking a question and returning an answer we see the obvious importance of meaning, especially when we consider that in order to go from question

[1] "Extended" because all the Church-Turing thesis really says is that all computable functions on natural numbers are computable on a Turing Machine. Part of the problem is that scientists and philosophers try to stretch the thesis as much as possible.

to answer we have performed numerous steps in a formal system. When we ask a meaningful question we expect a meaningful answer. If instead we picture a human mathematician fervently scribbling ones and zeros on a long sheet of paper, and at some point see him jump up and exclaim that he is done, we expect that the mathematician gives us an answer that has some meaning in relation to the question, rather than that long string of numbers. We can say that the meaning of the output of a computation is related to the input. The inverse also holds, where the meaning of the input is related to the output. For what use is an answer if one does not know what question was actually posed, other than for entertainment purposes on Jeopardy? Knowing that the answer to a very profound question is '42' is not enough, it is the understanding of how that answer was returned that creates meaning.

So we can view computation as the process of answering questions or as the process of coming up with solutions to problems. This is done by first representing the question or problem in a formal system, when it comes to human or computerized calculation. An algorithm may be created to solve the formal problem and it may then be executed. Eventually the machine running the algorithm will return a result which we humans are able to give meaning to, since we know what question was asked. It is not unimportant to see that translating an informal problem to a formal representation is a reducing action. When one wants to solve a real world problem only the essential information is formalized and as such, the result will also only be phrased in terms of essential information. In this light, we can see computation as nothing more than an alternative to searching, working, thinking or any other word that describes the process in between question and answer. And it is the formalization and procedural following of mathematical primitive steps that distinguishes computation from its alternatives.

The emphasis on the meaning of computation will certainly narrow down its definition. This *question-answer semantic* take on computation certainly puts biological computation out of scope, since we cannot give any meaning to these physical computations. But our current understanding of neural networks does fall into this semantic account because of the clear meaning related to input and output. With this semantic take on computing, saying that something 'computes' is a void statement without knowing what it is that is being computed and what the answer means. Or the statement is equally void if we do not know what question was posed to begin with. Calling a process 'computing' requires that we have a semantic understanding of that process: we know how to interpret what goes in, we know how to interpret what comes out, and we can reason about the steps in between.

2.2 On Epistemology

In chapter 1 we have seen how computation in DP is sometimes used ontologically. As we recall, Fredkin's quest is to discover the cellular computation underlying physics and Tegmark explained how in the Computational Universe Hypothesis computation is incredibly important in defining the relations in reality. These claims put a strong ontological emphasis on computation. Everything is or is the result of a computation. This ontological claim is also called pancomputationalism. But as we have previously deconstructed computation to be the process between questions and answers and the semantics of it, perhaps in the philosophical context computation is more suited to say something about epistemology rather than ontology. Can the theory of computation give us more epistemological insight?

If we hold the Church-Turing hypothesis to be true, we can say that the theory of computation is able to provide us with theoretical claims about the process of answering questions and solving problems. Specifically in computability theory, verdicts are made on whether certain functions are computable or not. We can loosely consider computability theory to discern what questions actually have an answer. Epistemology is generally concerned with what knowledge is and what can be defined as knowledge. In the spirit of what computability is good at, one can give epistemology a computational twist. In this section I will explain the novel concept of *temporal epistemology*; the categories of not knowing.

Taking the idea that some problems can be definable in a formal system, the theory of computation is then able to reason about the feasibility of finding solutions to these problems. With this reasoning, certain knowledge about the problems starts to appear. It might not be hard to see how a theory that can reason about what problems can and cannot be solved is a good basis for knowledge. After all, if computer science can prove that no answer exists for a certain problem, this can be considered evidence for the impossibility to know the answer to the problem. The relationship between what is computable and what is knowable has a little more depth to it, however. Temporal epistemology can be considered a form of meta-knowledge. It aims to define if something can or cannot be known in time. Whereby 'in time' refers to any moment in time in the future, which is a particularly broad time period but the reasoning behind this definition will become apparent soon. To give structure to this epistemology, three categories of *not-knowing* can be defined. A trivial category and two *temporal-agnostic* categories.

The first category of not-knowing is the trivial category, of the problems that cán be known. This is the category of problems that are simply knowable in the present or will be known in the future. It is the category to which the majority of all problems can be assigned, as it is the category of everything that is known in the present. Traditionally, epistemology concerns itself with defining knowledge, and to that mostly within the discussion of justified, true, and believed knowledge. All those properties are properties of the present. It can therefore be said that all previously treated problems in epistemology are of the first category, as they are part of the present. Besides all that is known right now, everything that we can proof we will know in the future is also ready to be placed on this heap. This 'known in the future' definition can be treated quite trivially too: I can perhaps not know what my friend is doing now, but I might know that he is reachable on his cell phone. I know that I can call him and inquire about his activities, therefore I know that I can know what he is up to. This is perhaps a perfect non-formalizable real world example equivalent to what computer science calls computable functions. Where the theory of computing is capable of determining whether a problem is computable and thus solvable it is easily capable of placing problems in the first category. It tells us that it is easy to know the answer to any multiplication problem, that it is relatively easy to solve a 9x9 Sudoku or that it is relatively easy to find the shortest route between two addresses in New York City[2]. It therefore gives us justification to know that we can know about similar problems and solutions in the future. We know that we will know. We see the link between these mathematically formalizable problems and the non-formalizable real world examples. In both cases we have a justification for making the knowledge claim about future knowledge.

In the second category this ease of knowing about knowledge starts to become problematic. This category is populated by some problems that are treated in the field of complexity theory. To contemplate this set of problems initially, we can think of the current predicate 'knowing' and the future predicate 'not knowing'. This implies that in this category we should be able to tell what our future knowledge state is, and that knowledge state should be that we do not know. How is this conceptualizable? It means that complexity theory should provide us with problems we know we will never solve, not because an algorithm can potentially be running forever, but because the algorithm would simply be running for too long. So long in fact that it would exceed human scales of time so far that we can safely say 'never'. This means that it is the class of problems that provably have an answer but that those answers are just too complex to compute. One of those problems is the Ackermann function. It is a relatively simple and short function with two arguments for which the inputs return small integers. However, the function can be considered as super recursive[3]. Whereas a normal recursive function calls itself at some point, a super recursive function calls itself with an argument that is defined by itself. A variant of the Ackermann function, the Ackermann-Péter function is displayed below. Notice that the recursive call in the last case is nested with another recursive call.

$$
\begin{aligned}
A(0,n) &= n+1 \\
A(m+1,0) &= A(m,1) \\
A(m+1,n+1) &= A(m,A(m+1,n))
\end{aligned}
$$

[2] A reader schooled in theoretical computer science will recognize that the latter two problems are classified as NP, however my point is that even though these problems are harder, they are still computable.

[3] A normal recursive function is a function that can call itself.

The answer to the Ackermann function $A(4,1)$ is known, 65533, and is computed with relative ease. But increasing the inputs from this point on increases the complexity astronomically. To the point where it becomes bigger than the amount of particles in the universe in recursive complexity. But, it is easily shown that the algorithm will be able to produce an answer, this is just not physically possible. Some inputs of the Ackermann function are hereby a great example of problems that are known to be unknowable; the defining property of category two.

So far we have dealt with problems that are computable. And we have concluded that of these computable problems we either know that the answer can come to us in reasonable time or that the answer is so complex to compute that we would have to wait an eternity. Does this mean that a third category 'I do not know if I will know', does not exist? On the contrary, this category does exist and it is populated by some problems that are treated in the field of computability theory. In this third category, this ease of knowing about knowledge starts to become problematic. It is the category of problems of which computer science tells us that we cannot compute the answer. The prime undecidable problem is the halting problem. Proven by Turing himself, it states that there is no program that can decide whether any computer program and input will at some point in time either halt or run forever. Many other undecidable problems are reducible to this problem. In a similar way it is known that there is no program that can determine if a given set of Wang Tiles can populate a plane, but also - and this will be of importance later - that no program can compute the Kolmogorov Complexity of an object. Of similar importance is the undecidability whether a certain pattern will appear from running a simulation in Conway's Game of life.

A naive and rigid computer scientific interpretation would say that one cannot know the answer to undecidable problems, which would dub this third category the category of problems that I will never know the answer to. This, however, is false. This is because the computer-scientific interpretation is misleading when it comes to knowledge. It states that certain problems are formally undecidable[4] but that does not imply that no knowledge can be gained whatsoever. Given the halting problem, we know that it is formally undecidable whether any program stops within a finite amount of time or not. But perhaps, we can decide to just start a program and leave it running. Potentially, for an infinite amount of time we would be staring at the program not knowing if it stops or not. This would not really help us. But it is not impossible for the program to stop, therefore after an instance, a minute, a year or 2500 years, it could stop. And if it does stop, we have gained knowledge. Where mathematics and computer science is not able

[4] It is not unimportant to place a footnote with the definition of undecidability in this context. Undecidability namely only holds within a given formal system. This means that if a problem is proven to be undecidable within mathematics, it does not mean that it is undecidable outside of mathematics as well. This nuance that undecidability has only strengthens the point of this paragraph that we can actually acquire knowledge about problems that are formally undecidable. Where we could formally only conclude that we cannot know the solution to a certain problem, there are still tedious but informal ways to potentially come to an answer.

to *a priori* gather this knowledge, simply running the program could. The same is true for all undecidable functions and all the examples in this section. We could just hit 'enter' on a Game of Life simulation and wait till we see a given pattern appear. We could wait potentially forever but that does not change the fact that theoretically at some point in the future we will get a result. What matters is that it is not impossible to acquire the knowledge within a finite amount of time. It is however not knowable if this knowledge will be acquired. The definition of the third category is therefore the opposite of its naive interpretation. In the second category we place undecidable problems, where it is undecidable if we can gain knowledge in the future: 'I do not know if I will know'.

2.3 On Metaphysics

Since DP throws out metaphysical claims all over the place I am forced to take a metaphysical standpoint. But taking a standpoint on metaphysics, specifically on the ontology of the universe is not an easy feat. If I want to react to Tegmark's claim that the universe is mathematical, or to Fredkin's claim that the universe is a computation, or to Wheeler, who says that information is the fundamental ontological unit, it would seem that I need a response in the form of "No, that what is, is actually x". But doing that, no matter the arguments, would always open up possibility for even more critique. And this in itself would perhaps not be the strongest form of critique against the metaphysical standpoints in DP either. A stronger critique would tackle any statement of the form "That what is, is actually x". The only thing I can do here is resort to a form of metaphysical agnosticism. Specifically, Robert Anton Wilson formulated the definition of *model agnosticism*, which is quite convenient for me to use here. In the footsteps of the reflections of Niels Bohr on the Copenhagen interpretation he noted that any (scientifically) created model of the world should not be confused with the world itself (Wilson, 1986). Any model would be a reduction of the modeled, and anything that is lost cannot be regained. The map is not the world!

2.4 On Discreteness

The main criticism on digital physics from the scientific community - mainly from physicists - is that the problems it has with handling continuity in physics are too problematic for it to be a viable alternative to contemporary physics. Proponents of DP like Fredkin acknowledge the shortcoming in regards to continuity and do attribute it to the limitations of computing. This line of reasoning will not be part of my critique on DP. A case can even be made that both proponents and opponents fall short in their view on continuity and discreteness. Where some say that nature is fundamentally analog and some say that computation is fundamentally discrete (and therefore not unifiable), both may in fact be wrong.

The relation between discreteness and computing is mostly brought forward by the models of computation we have, namely in the academic scene the Turing Machine, and in day-to-day use the personal computer. But these examples can be regarded as mere anecdotal evidence of discrete computing. Computing itself is not confined to the discreteness that these systems exhibit. Analog computers have existed and do still exist, even though they have limited use. Let us look at a hypothetical analog computation device in order to grasp analog computation.

In order to build our hypothetical analog computation device we are going to need an analog principle to work with. For this we are going to use wave interference, the physical manifestation of which will be water in a long rectangular basin. At each end of the basin we will have a device that creates waves and a floating buoy that is attached to the side of the basin so that it can only move vertically along the height of the water. Attached to the buoy is a piece of chalk that leaves traces on the side of the basin. Now in order to start the computation we create a wave of a certain height simultaneously at each end of the basin. When the waves collide with each other at the center the principles of wave interference come into play. If both waves have the same vertical direction we will see constructive interference take place, and otherwise we will observe destructive interference. When the waves reach the opposite ends of the basin the buoy will float up and leave a trace of chalk on the side. The position of the chalk mark, mainly in relationship to the bottom of the basin, is the output of our computation. Our analog computation has succeeded! But what did we achieve? The semantics of the computation need to provide clarity. The distance between the bottom of the basin (or some other base-line like the level of undisturbed water) and the mark is an analog output. This distance is measurable up to a theoretically infinite precision. But if we want to use this output for anything we will need to perhaps make a cut out or hold a rope next to the basin and cut it at the exact position. We might be able to bring this rope to yet another analog computation device that is able to compute with its length, and then that will again produce an analog output.

We can say that the information produced by our wave creating floating buoy computer is analog in nature. But no matter how elaborate our analog computing system is we will always have results that we define as analog. If at any point we want to articulate this output, say it out loud or send it to someone over a digital medium, we have to make it discrete. We have to conclude that theoretically infinitely precise distance is actually for instance $1,3987$ meters. No matter how precise we measure it it will always turn into something discrete and we lose some information. It seems that continuity is easily created by definition, though in practise easily lost through communication. This is why analog computers are fading in the digital age. We cannot communicate analog values efficiently[5]. And this is also why analog and digital models do not

[5]We actually can if we consider measuring the voltage on an electrical wire as the analog value, but this has some practical disadvantages that I will not further discuss here.

really mix very well. The information loss is irreversible.

Digital physics is called *digital* physics for a reason. The underlying ideas of pancomputationalism are not refutable by arguments regarding the analog or discrete nature of reality in itself. It is just that no one has really defined a computable physics based on analog computations. We have seen that computation was a formalization by Turing of that what a human computer could do with a pen and paper. Writing, just like speaking, is a discrete communication medium. So if all formalizations of computation are equivalent according to the Church-Turing thesis, then computation in the formal sense is discrete. So the continuous argument against DP is really only an argument against the shortcomings of *formalized* computation. Analog computation would not be susceptible to this argument, but then again, analog computation has its own practical shortcomings. But in this way it is just as easy to ascribe continuity to nature itself. Everything is possible in definition an the hard reality is that communication is not always continuous.

So disregarding digital physics merely on the fact that digital computation cannot be unified with continuous physics is rather shortsighted. We have to realize that analog and discrete values do not mix because that is their precise definition. Analog computation exists, which strengthens the pancomputational metaphysical position. It is, however, true that most literature in *digital* physics is based on discrete computing, which limits the relation with contemporary physics only to theories that define nature to be discrete. Let this then be a kind tip to digital physicists: analog computing may be a viable computational model of physics as well.

Chapter 3

The Critique

The theses and aspects of DP have been layed out in chapter 1 and the most important concepts of have been layed out in chapter 2. Now the time to pull DP apart has finally arrived. In this section I will go through the different variants and aspects of DP. As seen in the final section of chapter 2, I will not be using any arguments from the field of physics against DP. I will keep my argumentation strictly philosophical.

3.1 Is the universe a computation?

I have earlier refered to the claim 'The universe is a computation' as strong digital physics, and in other literature this claim is also refered to as pancomputationalism. From the authors that we have discussed in chapter 1, Edward Fredkin and (up to a certain extend) Max Tegmark have been identified as supporters of this claim. In the previous chapter we have actually already discussed the two concepts that are combined in this claim: computation and metaphysics. The attentive reader has already gathered the two main arguments that are going to be brought in against strong digital physics from the contents of chapter 2. Let us start with the computational argument.

To say that the universe is a computation puts a lot of emphasis on the meaning behind a computation. In the previous chapter we already discussed that one should not use the term 'computation' too lightheartedly. It does not always make sense to call a system that appears to exhibit state transitions of a regular pattern a computation. We generally define 'computation' as a human act, and most formal definitions are also modeled after a human computer finding answers to formal questions. To identify anything as a computation requires a semantic explanation - which most pancomputationalist theories do not have, they mainly focus on a mechanistic account of computation - and so this argument against pancomputationalism comes down to a debate on the definition of computation: semantic or mechanistic. That would ultimately not be a convincing argument for die-hard pancomputationalists. It would result in an agreement to disagree.

In the previous chapter we discussed that critiquing an ontological-metaphysical claim should not be done on the basis of postulating a counterargument on the same level. For that reason, I cannot follow a reasoning along the lines of "The universe is not a computation, but instead it is x". This makes any subsequent critique rather weak. If I cannot come up with a better argument then what can I say at all? Well, I can point out the anthropocenteredness of the postulate: computation is a very human concept, a model. And when one takes a model agnostic standpoint the weakness of the claim 'The universe is a computation' becomes apparent. For that matter any ontological-metaphysical claim is anthropocentric. Obviously the ontology of metaphysics is a topic that has been haunting philosophy forever and this thesis will therefore not settle that discussion. But it is good to note that digital physics actually takes a very strong position in this discussion. If one only starts from a perspective of physics, mathematics, or computer science, which are objective sciences, then it is not unremarkable that the resulting theories have very objective claims. Let this then be a call upon these theorists, that centuries of philosophical discourse preceded them with arguments that are also very relevant against them.

The fact that I cannot bring arguments of the same objective level as the authors I have discussed is not necessarily a weakness. As I do see that if I would bring equal level arguments

I would fall into exactly the same pitfalls that I try to argue against. Therefore, I cannot throw out pancomputationalism completely. But, what I can do is argue against the general reasoning used by the discussed authors. The rest of this chapter is dedicated to critique this reasoning we commonly find in digital physics.

3.2 Ontological arguments in science

"The universe is x". This is the general form of the three ontological claims we have come across so far. In the introduction of chapter 1 we saw various simulationist claims that the universe is actually a (computer) simulation ran "somewhere else". Edward Fredkin stays clear of the what or how, but he does argue for the universe to be a computation. Finally, we have seen Max Tegmark argue for the universe to be a mathematical structure. All these claims possess a certain transitive relation with each other: If the universe is a simulation then it must be a computation. Similarly, if the universe is a computation then it must be a mathematical structure. We have seen that the arguments for each individual claim are quite different from one another. Yet, they seem to result in a connected ontological claim. The difference in arguments for these claims, while arriving at similar results, must certainly indicate that they approach the nature of reality correctly, right? It is finally time to dive into the logic and see what is really going on with these arguments.

We have discussed the arguments of Fredkin and Tegmark at greath lengths in chapter 1. Fredkin's and Tegmark's arguments can individually be criticized and there are perhaps many valid ways of doing so. I will, however, show here that both Fredkin and Tegmark essentially commit the same fallacy in their argument. Doing so, a single philosophical argument suffices to argue against both Fredkin and Tegmark, and perhaps also against other ontological arguments in science. To recap and to further my efforts I will provide both Fredkin's argument for a computational universe and Tegmark's argument for a mathematical universe in a condensed and schematized form. Starting with Fredkin's argument for universal-computation:

1. A system can be shown to be computation-universal if one can construct a universal machine within that system.
2. We can actually build and operate a universal computer in our world.

Physics must be computation-universal.[1]

In chapter 1 we initially glanced over this argument. Fredkin views this argument mainly as "hard experimental evidence" (Fredkin, 2003, p.196). In this reasoning, us observing ourselves

[1]Note that this claim in itself is not necessarily an ontological claim. However, in the context of the second premise, it cannot be seen as anything but ontological.

building a computer is apparently the experimental fact. Either way, Fredkin's conclusion is highly ontological in nature. It is this weird tension between empiricism and ontology that I will further explore. Before doing so we have to note that Fredkin also comes with an argument that supposedly is "the biggest clue for DP to be fundamentally how the universe works" (Fredkin, 2003, p.200): this is the argument that the vast amount of small integers in physics hint towards the digital nature of reality (as discussed at the end of section 1.3). I call this Fredkin's *indication argument*, as this argument does not follow a conclusion given certain premises, but is meant to invoke a strong feeling with the reader that certain aspects of reality do indicate the correctness of Fredkin's main thesis. We can see that Tegmark actually follows a similar pattern in his reasoning:

1. If two objects are mathematically isomorphic[2] (a one-to-one correspondence between two objects that preserve their structures), then there is no meaningful sense in which they are not one and the same.
2. Our External reality is mathematically isomorphic to a theory of everything that describes it.

External reality is a mathematical structure.

Tegmark's argument has the same features as Fredkin's argument. Both Tegmark's and Fredkin's arguments are forms of *Modus Ponens* and therefore logically valid. But he also operates with a strange tension between empirical and ontological reasoning. And he too has his own indication argument, namely that the ability of mathematics to describe the universe so perfectly surely points us towards the conclusion that the universe must be mathematical (Tegmark, 2008, p.4).

To resolve both indication arguments first, we can see that both arguments do not really work without their ontological companions. To say that small integers in physics point towards the universe as a computation is a huge leap that on its own does not make sense. The same goes for Tegmark's leap from mathematical describability to mathematical ontology. The authors are probably aware of the weakness of these isolated arguments and therefore use them in conjunction with their stronger arguments. As already stated, in the case of both indication arguments the conclusion does not follow from the predicates and they are therefore clearly meant to support the ontological arguments. I will therefore move on to the main ontological arguments.

[2]It can already be debated if Tegmark's usage of the term 'isomorphism' is correct. However, I will procede to follow Tegmark's usage as this mathematical discussion does not lie within the scope of this thesis.

It is the combination of empiricism and ontology that allows me to categorize the ontological arguments of the authors. Both authors appeal to an analytical[3] premise and an empirical premise to reach an ontological conclusion. However, even if all premises are true it does not follow that the ontological conclusion is true. In both of the arguments shown above the first premise is analytical. It can be considered a definition that is widely used and accepted in contemporary science. Fredkin invokes one of the definitions of a Turing Machine while Tegmark invokes an interpretation of mathematical isomorphism. The second premise then can be considered the step required to pull this analytical statement into the real world. To cross this bridge, the authors need an empirical premise. For Fredkin: We can build physical instances of the Turing Machine. For Tegmark: A theory of everything describes reality. The conclusion of both arguments is seemingly ontological, as we end up with a conclusion that sticks an analytical definition onto our world. However, because of the huge leaps between very different domains of science and philosophy, I argue that this type of reasoning is fallacious.

Analytical statements have no value outside of their domain. To say that the applications of logic, mathematics, and theoretical computer science in the real world are useful would be an understatement. They have all been of incredible importance. But concepts from these sciences are usually *applied*, and applied is what these sciences are meant to be. We can, for example, *describe* a physical wave in water with a sine function, but no right minded person would stand on a beach and point at the water exclaiming that they "*are* sines". All the analytical statements surrounding wave mathematics are applicable in the real world, but these statements hold no ontological value.

The closest actual fallacy to this misuse of analytical statements would be a *reification*. In a reification abstractions are treated as if they were a concrete physical entity. Alfred Whitehead similarly talks about the "Fallacy of misplaced concreteness"(Whitehead, 1925). The fallacy in the arguments that we have seen is a special case of the reification fallacy. It is specifically the reification of analytical scientific definitions that appear to be problematic. We can therefore speak of the *analytical reification fallacy*. This happens when analytical statements result in ontological claims about non-abstract matters. The analytical statement exceeds the boundaries of its domain. In the conclusion the abstract becomes concrete while the concrete becomes more abstract. The result is an abstract-concrete limbo; it simply does not make sense from either angle. Because part of the analytical definition overlaps with the real-world definition; it does not follow that the analytical value can be directly applied to the real world.

The analytical reification fallacy is not limited to the two arguments in DP. It pops up in more

[3]The first premises are refered to as analytical as the entire premise is based on mathematical definitions. That a universal machine can be built within a computation-universal system is a definition of a computation-universal system. Likewise, that two mathematically isomorphic structures are the same is a definition of mathematical isomorphism, albeit this is a mathematical interpretation.

places in science. Fore example, John Lucas made an ontological argument against determinism and for human free will using Gödel's incompleteness theorem (Lucas, 1961)[4]. In this argument the same fallacy can be seen unfolding; First an analytical premise is posed, then an empirical premise is used to bridge to an ontological conclusion. The result is a misuse and invalid domain crossing of Gödel's theorem.

1. Formal systems are either incomplete or inconsistent according to Gödel's theorem. Therefore, there are problems that computers cannot solve.
2. Humans can solve problems that deterministic formal systems cannot solve.

Humans are not deterministic machines; they have free will

The analytical premise in Lucas' argument is the definition of Gödel's theorem, which is a statement considered to be correct within mathematics. He then invokes the emperical observation that humans can solve problems that machines apparently cannot solve due to Gödel's incompleteness theorem, to eventually conclude that humans cannot be subject the same theorem, and thus not be deterministic. Lucas arguably makes even more questionable deductions, but the analytical reification fallacy is still clearly found in the reasoning.

It is no coincidence that the term *ontological argument* is already a renowned term in philosophy. As it appears, the fallacious reasoning that I described in this section is actually the same structure of reasoning Anselm of Canterbury used in his ontological argument for the existence of God. The arguments of Fredkin and Tegmark are ontological in that their conclusion is ontological, but the similarity of using analytical premises between Anselm's ontological argument and the discussed authors makes this term even more fitting. Roughly, Anselm's argument is as follows:

1. By definition, God is a being which none greater than can be imagined
2. A being that exists in reality is greater than a being that solely exists in the mind.
3. God exists as an idea in the mind

God exists in reality.

We observe the exact same structure being used in the ontological argument as with DP arguments seen so far. First we see an analytical claim, then an empirical observation, followed by an ontological conclusion. To claim that Anselm's ontological argument is just as fallacious as the DP arguments would be a very quick reduction of a thousand years of philosophical and theological reasoning as the philosophical discussion is mainly on the theological status of the first two premises. That is for theologians to debate and hence is outside the scope of this thesis.

[4] It should be noted, however, that Lucas retracted his statements years later.

What remains is the similarity between ontological arguments in science and this centuries-old ontological argument. At least when it comes to science, this reasoning is a fallacy.

3.3 How viable is weak DP?

When we classify strong DP as the ontological claim that the universe is a computation then the argument in my previous section dealt with the philosophical issue of making such a ontological-metaphysical claim. Weak DP, i.e. the idea that the universe is *describable* as a computation, is still on the table. As we have seen in chapter 1, it was Konrad Zuse who first proposed to describe physics with an "automaton theoretical way of thinking" (Zuse, 2012, p.5) and Fredkin joined Zuse in his methodology. Schmidhüber chose a different approach that did not necessarily rely on cellular automata but used a more classical computer-scientific approach. But since a computation in a cellular automaton is computationally speaking identical to Schmidhüber's classical approach[5], we can view them as the same. Max Tegmark was the odd one out with his Computational Universe Hypothesis. The main feature that distinguishes him from the others is that he does not pose a symmetry between the time evolution of the computation and the time evolution of the universe. For Tegmark, the universe's full time lapse is already 'precomputed' and stored. We cannot regard Tegmark's theory as weak DP in any way as it does not necessarily try to be a viable alternative to contemporary physics. In this section we will look into the viability of the effort that is being made to redefine physics in a computative way.

Perhaps the most important question we first have to ask ourselves is: how does weak DP differ from regular physics? Indeed, in the end they are both formal systems that try to describe all the workings and processes of the real world. If one of the arguments against DP is the restriction of digital computation, then we have seen in chapter 2 that since analog computation exists, DP as a whole is still in the running. To reinforce the idea that DP is not all that different from regular physics we can even look at our ability (and the fact that we make great use) of programming the physical *laws* into a computer and running a simulation with it. Is this not exactly what DP is, and are they therefore not two sides of the same coin? What difference is left?

The key difference lies in the things that Tegmark described as *baggage*. He saw that mathematical formulas in physics always described the behaviour of something we labeled while not explicitly defining that label in mathematical terms. A set of formulas that worked on the label or object 'molecule' were not the same as the formulas on the object 'electron'. Let alone the mechanical formulas that work on the object 'football' and 'earth'. Very reductionistically, practicing physics is nothing but applying mathematical formulas to abstracted objects. Whether that object is a football or an electron does not matter because the formulas are just different; they

[5]Provided that the Church-Turing thesis is correct.

work because they are catered towards a context and a closed system. Now what is the object for formal descriptions in DP? Information comes to mind represented formally by ones and zeros. But with that, DP dug a very complex hole for itself in which it now not only needs to provide an algorithm that describes the same behaviour of a football as regular physics, it now also has to provide a way to translate 'football' into ones and zeros. And yes, physical formulas do not actually take footballs, what they take is the properties of an object, like mass, volume, material density, air resistance, surface friction etc. But all these terms are just as much anthropic baggage as the term 'electron'. We realize now that the project of DP and Tegmark's project lie even closer together, as also any theory in DP needs to formalize the baggage away.

Up to a certain extent, DP actually does not have to formalize all the baggage. This is because it is hypothesized that objects like electrons and footballs emerge from an accurate computational TOE. Instead of going deeper and deeper through multiple baggage layers, from molecules to atoms to protons to quarks to fields (like the natural development of physics), it strives to produce that bottom layer of information from which all these baggage layers emerge. It tries to leap over a century of physics and a century to come, holding the hidden axiom that (quantum) physics actually has an end-layer of baggage.

So far this invokes the feeling that actually finding a computational TOE is hard, yet not impossible. But how hard is it exactly? In order to get to a concrete answer we have to recall chapter 2, in which I explained that the problem of knowing if a certain pattern appears in Conway's Game of Life is undecidable. This problem is actually more generalizable and we know that knowing if a certain pattern appears given any cellular automaton is undecidable. I also argued to view computability and complexity theory as epistemological arguments: we can know what we can know based on these theories. Now let us say that we are able to formalize physics in an informational way (defining nature in terms of bits), then the goal of DP is to find an algorithm that manipulates these bits in a way that is identical to reality. In this case the undecidability of the Game of Life problem actually tells us that we cannot reasonably find an algorithm that does so.

Why is this the case? DP actually tries to play Game of Life with the universe. It tries to define a set of rules on a cellular automaton that, given a starting position, perfectly imitates the universe. The problem is, however, that even if we can define the current state of our universe, we cannot create an algorithm that decides if this state appears from a given configuration and algorithm. Even if we hypothetically had a digital TOE in our hands, there is no formal procedure that can decide if this digital TOE eventually could arrive at a state that is similar to our current universe's state. In short: we cannot verify if a given algorithm is actually the algorithm that describes our universe. This does not mean that such an algorithm does not exist; it merely means that we cannot think of a systematic way to find it[6]. Therefore, finding such an algorithm can be

[6]Note that the scenario described here is the ultimate scenario. It already assumes that a formal and finite state

classified as a problem of the third category we discussed in chapter 2; we do not know if we will ever know such an algorithm.

Finally, even if we would have such an algorithm, one could already grasp the impossibility of building a machine that can actually run this algorithm for semantic benefit. It has even been proven that it is impossible to construct such a machine (Wolpert, 2008, p.23). This proof can be Intuitively translated to the idea that a machine that can output the entire state of the universe at any particular time also has to take itself into account, which would result in a certain infinite regress. To be precise, this machine can be built outside of our universe, but not inside of it (Rummens & Cuypers, 2010). To be a bit tongue and cheek, we can draw a philosophical conclusion from this: if we define *everything* as the complete state of the universe at a certain point in time, then the pseudo-philosophical question "*Can we know everything?*" is provably answerable with: "*No*". One can view this as an epistemological constraint. But one can also keep this constraint in mind while dealing with regular physics.

So, 'weak DP' makes sense on paper, but the intricacies of decidability and the computability of algorithms make it so actually verifiably simulating our universe is argueably impossible. There are many logical hurdles that prevent it from being done. It is interesting to think about universes being simulated on a computer, but actually taking the state of our universe and describing it as bits, and then verifying if a hypothetical theory of everything is provably *our* universe's theory of everything is impossible. Even though weak DP seems an interesting alternative to contemporary physics, the scientific method of contemporary physics is in contrast to weak DP clearly workable and provides results. Weak DP will therefore never be an alternative that is viable enough to replace it.

3.4 The use of Occam's Razor

We have seen that using Occam's Razor is also a somewhat shared theme in the DP literature. Both Schmidhüber and Tegmark apply Occam's Razor in their reasoning. The use of the razor by these authors however, justifies a dedicated section in this chapter.

> Any particular universe evolution is highly unlikely if it is determined not only by simple physical laws but also by additional truly random or noisy events. To a certain

representation can be made *and known*. Furthermore, if there is an algorithm that only works on the formal state of the entire universe then this endeavour is already doomed by our inability to know the state of the entire universe at one particular moment. Even if there is a local computable algorithm (an algorithm that changes a bit only depending on the value of bits very close-by) then it would mean that there is a certain floor to the relevant information that can be stored in a finite amount of space, which would imply that classical physics is able to find an indivisible particle. These scenarios show that other constraints make the goal of DP even harder.

> extent, this will justify "Occam's razor" which expresses the ancient preference of simple solutions over complex ones, and which is widely accepted not only in physics and other inductive sciences, but even in the fine arts. (Schmidhuber, 2000)

> From an esthetical point of view that favors simple explanations of everything, a setup in which all possible universes are computed instead of just ours is more attractive. It is simpler. (Schmidhuber, 1997)

> Naively, you might think that a single number is simpler, but the entire set can be generated by quite a trivial computer program, whereas a single number can be hugely long. Therefore, the whole set is actually simpler... (Similarly), the higher-level multiverses are simpler. [..] A common feature of all four multiverse levels is that the simplest and arguably most elegant theory involves parallel universes by default. To deny the existence of those universes, one needs to complicate the theory by adding experimentally unsupported processes and ad hoc postulates: finite space, wave function collapse and ontological asymmetry. Our judgment therefore comes down to which we find more wasteful and inelegant: many worlds or many words. Perhaps we will gradually get used to the weird ways of our cosmos and find its strangeness to be part of its charm. (Tegmark, 2008)

In essence, both authors make the same argument: A program that describes a single universe is longer than a program that describes multiple universes, because the simpler is to be preferred (Occam's Razor): the multiverses exist.

What is happening here, is that Occam's Razor is being employed in argumentative reasoning to prove the existence of multiple universes. This is an ontological conclusion. Readers of section 3.1 can already see the fallacy looming except that it does not completely fall into the analytical reification fallacy. In this case Occam's Razor is being used in conjunction with analytical facts to make an ontological claim. This means that this argument has strayed so far that even the empirical component has been thrown out of the window. Instead, Occam's Razor or the appeal to simplicity is meant to replace it. A reader schooled in philosophy does not need to be reminded that Occam's Razor is merely a heuristic principle and not a logical axiom. It is a tool to select the plausible from the almost infinite amount of possible hypotheses. Insofar as it is a heuristic during the development of a scientific model it is reasonably used (Gauch, 2003). But it cannot be used as an argumentative statement in that model. Occam's Razor is meant to select between hypotheses that make the same prediction and cannot to be used to choose from hypotheses that make different predictions (Gibbs, 1996). Since Schmidhuber and Tegmark argue between two different predictions: one or many universes, the use of Occam's Razor is already fallacious. To take it a step further, one could actually argue that Schmidhuber and Tegmark *violate* Occam's Razor. As the razor points out that entities should not be multiplied without necessity, while the authors have no problem multiplying universes.

In all of these discussions we have not used physical arguments against DP. We have dissected the reasoning in these arguments and pointed at their faults. I have been agnostic about all the conclusions or even on possibility of multiverses, but I have shown that the DP reasoning for these conclusions is at least not correct.

Chapter 4

Rebuilding

So far we have discussed, analysed and critiqued digital physics. If one thing has becoming clear from chapter 3, it is that it does not really make sense to think computationally about our universe. Does this mean that all the effort that has been made, trying to unify physical concepts with algorithmic theory, has been a waste? Perhaps not entirely, because this mountain of research is still applicable to universes we simulate ourselves, and for that goal it basically has not been used yet. Or the links that have been established between the different sciences are at least interestingly enough to pursue. In this final chapter, I will argue why the research in this field has not completely been in vain, and pick out those elements of DP that are highly worth discussing.

4.1 Computing Universes

DP has gone out of its way to give computer-scientific representations of concepts from physics and philosophy. This means that we can now also point out these concepts in computer science itself. When we look at an algorithm, we can now not only talk about complexity and computability, but we can also talk about concepts like reversibility[1], entropy, or even perceived randomness. These concepts are already applicable to information itself, but DP has linked them conceptually to their counterparts in physics. It does this by saying "if we pretend this algorithm is actually computing a universe, then that universe has metaphysical properties x, y and z". We can pick any algorithm, give it an input, set it off running, and then claim that we are simulating a universe. This may sound crazy but it would actually put our own ideas of concepts in physics to the test. We now have the ability to look at algorithms and we can try to link algorithmic properties to metaphysical concepts. Creating hypothetical universes in order to study them can be called *hypothetical metaphysics*. What we are doing here is tying definitions in physics and computer science together to see that they should have the same conceptual philosophical meaning. There are more interesting mathematical theories that grasp the imagination of philosophers, such as chaos theory, which can grouped under hypothetical metaphysics be a very fruitful exercise to study as they provide a fruitful understanding of the interplay between mathematical concepts and philosophical implications.

What we can actually conclude about simulated universes tells us foremost something about our own thinking about universes. This grants us the ability to grasp the limitations about any science that tries to reflect on our own universe. Simulating hypothetical universes is therefore more of an exercise in cosmological understanding. Computer science and physics work still work with the same mathematics, within the same formal system. So if we provably cannot simulate something, what does this tell us about our own cosmological theories?

Computer science and physics formally have the same scientific root, namely mathematics and logic. Where physics excels at applying mathematics to explain observations in the real world, computer science excels at exploring the boundaries that restrict mathematics and logic. In chapter 2 I connected computer science with epistemology. The unification of physics and computer science in the field of DP connected concepts of both fields. When those connected concepts ever raise the flag of uncomputability in computer science, this would immediately reflect on the physical models in the same way; there are parts of the model that we cannot know. It is this idea exactly that Stephen Wolfram has shown in one of his more works: computer science can provide us with knowledge about the limitations of physics (Wolfram, 2012). If there

[1] The concept of reversibility already has implications on information and time, but Bennett (1973) has shown that our act of computation is tied to physical energy and entropy as well.

is anything that has showed such a limitation for contemporary physics in this thesis, then it is the undecidability of Conway's Game of Life which implies that finding a baggage-free theory of everything is uncomputable (section 3.3) and therefore we do not know if we can ever know of such a theory. It has thus shown us *why* we have so much trouble finding a TOE. It strengthens our broad understanding of science.

4.2 The crossroads of sciences

Physics, mathematics, computer science and philosophy are the four scientific disciplines that have been heavily featured in this thesis so far. These sciences are not unfamiliar to each other. Physics goes hand in hand with mathematics, computer science is the scientific child of engineering and mathematics, computer science is a tool for physics to discover things more efficiently, and finally philosophy has some entry points in all of the three other sciences. In this thesis, I have not been restricted to reasoning of one or more sciences on their own. Thinking about DP requires one to think of these disciplines in a more unified way than before. It is the relation between these sciences that is often overlooked. In this thesis I have tried to show that these relations are often the most valuable things to look at in order to come to a better understanding. In this unified way of thinking the knowledge and formalization of physics becomes more constrained by reasoning *about* knowledge and formalization. We can hypothesize about reality, but we are eventually constrained in how we can formalize and reason about it. As a part of this broader project, I have shown in chapter 2 how computability theory and epistemology are actually closely connected.

This unified thinking shows that what seemingly has little consequence in one discipline actually has major implications in another. The debate about whether nature is fundamentally analog or discrete is almost secluded to physics, but in the unified picture we can also see how discreteness or continuity is also preferentially perceived by our mathematical way of thinking, or problematized by our computer scientific efforts to encapsulate the fundamental workings of the universe. Randomness may be observed in physics but randomness is defined in mathematical and computational concepts. The same goes for concepts like time, determinism, reversibility or entropy. A mathematical definition of one concept constrains us from thinking about the same concept differently in another field. Formalizing reality comes with its own problems. It is only by looking at all these sciences as one, that these problems can be distilled and somewhat handled. We take a step back to see where our issues really stem from. Physics, mathematics, computer science and philosophy can be thought of as four sides of a table rug, where if one corner is pulled (one concept in one science is redefined), the other corners slightly but forcefully shift position as well. Eventually we must realize that formalization itself is restrictive, and we must deal with the consequences. This does not imply that formalization is bad, far from it, but we must always realize that formalization itself is a reducing act, as I have thoroughly explained in chapter 2, and therefore a constraining act. Theoretical computer science is able to make these formal limitations more visible, and this has profound philosophical value and allows for more insightful physical understanding.

4.3 Observering bits

In the spirit of treating these sciences in a unified way, we will take a last look at Wheeler's ideas. As we recall from chapter 1, Wheeler based his suggestion to shift towards an observer-based and information-based approach of physics on his annoyance with some fundamental issues physics runs into from time to time. Infinite regress, the ontology of laws, the continuum, space, time etc. Wheeler's greatest act was stepping back and noticing that this external reality we are trying to describe is presented to us by means of information. It is information about this reality that we observe. So for Wheeler it was obvious that we should at least be aware of this and perhaps explore this information medium in general. We are not creating a model of the real world, but we are creating a model of information. This justified Wheeler in claiming that the information, the bits, were just as much our reality from a physical perspective as that real world.

Julian Barbour (2015) tried to argue against this observation- and information-centric view. He mainly questions why any definition of information or bit is sufficient to constitute a shift in focus, because information simply cannot be a fundamental physical entity as it even still acknowledges the existence of a real object that is the source of that information.

> But whatever authors may mean by information, quantum states still give us probabilities for outcomes in the form of factual information about things. Moreover, the probabilities themselves are determined by observation of things. I therefore conclude that things are the ground of being and constitute the ontological basement. Reality creates information and is separate from it. Once this has been recognized, we see that, for all its importance, information theory can in no way change what has always been the starting point of science: that structured things exist, in the first place in our mind and, as a reasonable conjecture given the remarkable correlations in our mental experiences, in an external world. (Barbour, 2015, p.6)

In general, Barbour argues as if Wheeler does away with the real world, elevating bits to a pure ontological status. Against this, he argues that bits do not solve any infinite-regress problem and that they themselves cannot be an atom of reality. "A 'bit' has no meaning except in the context of the universe" (Barbour, 2015, p.7). This argument would be correct if Wheeler had actually made such claims. It underlines that Wheeler's shift of focus to observation and information is an act of analysing physics itself rather than reality. We do not need to worry about what definition of information we use for the information we observe in physics: any definition of information from other fields would only cloud the discussion. We observe information period, that is the definition [2]. Smilga (2013) summarized Wheeler's aim in an almost perfect way: "So he [Wheeler] obviously had the suspicion that physics may not reflect structures of nature but rather structures inherent to information".

Wheeler's views and suspicions can be a great starting point for looking at physics in a more unified way with the other sciences I have mentioned. We should not only try to derive models from the real world about the real world, but also see how the structures of our thinking (mathematics) and the limitations of our reasoning (computer science) affect them. This is a turning point in scientific thinking. Being aware of the fact that physics essentially deals with information between reality and observation allows us to extend our reasoning and understanding when dealing with theories in any of these disciplines. Most authors in DP, independently of their work on either strong or weak DP, have this unified understanding. Mueller (2017) for example, emphasizes how subjective observation should play a role in theories in physics. In this way, it all comes back to Niels Bohr's philosophy of physics:

> Physics is to be regarded not so much as the study of something a priori given, but rather as the development of methods of ordering and surveying human experience(Bohr, 1963).

Or similarly by Niels Bohr, taken from Petersen (1963):

> It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature.

To this extend, I have tried to show that computer science and philosophy also have invaluable contributions to show "what we *can* say about nature".

[2]More discussion on its and bits can be found in Leifer (2015)

Conclusion

Digital Physics is a branch of physics that aims to provide a computation-based model of our understanding of the universe. Pioneered by Konrad Zuse in 1960, it initially went mostly unnoticed, but it gained traction and wider support some years later as a possible solution to problems brought about by quantum physics. Within the field of DP the views on the actual status of computation itself are diverse. Some view it as a mere descriptive tool for the universe while others see computation as the ontological base of the universe. The common denominator in their theories is the thought that information would be a more suitable object of physics, and that computation is the force that alters this information.

Regarding the ontological status of computation, otherwise refered to as strong DP in this thesis or pancomputationalism, I have shown that any objective counterargument would not overthrow its ontological claims. When one theory poses computation or mathematics to be the ontological status of nature, an alternative ontological claim and its justification would never end the debate. The same goes for any critique regarding the definition of computation. The approach that remains is to be model-agnostic about these theories: the map is not the world. What one can show though, is that the argumentative justifications for claiming either computation or mathematics to be the fabric of the universe that are given by different authors is in itself fallacious (although the claim itself may be true). I have shown that some key authors in DP commit an analytical reification fallacy by mixing terminologies from different domains. They generally take an analytical concept and an empirical observation to arrive at an ontological conclusion. This type of reasoning is eerily similar to the structure of reasoning Anselm of Canterbury used for his ontological argument: the existence of God. This may perhaps be a valid debate in theology, but it is surely a fallacy in science.

What remains is the idea that computation could more accurately describe nature, i.e. weak DP. This effort is motivated by a desire to lose the unformalized 'baggage' in physics and to move towards a pure theory of everything. I have shown that this effort is actually harder by not only needing to come up with an algorithm, but also with a data representation of real states. Not only does this make things more complicated, the task is also provably very hard. This is because the problem of finding if a certain state configuration appears from a given cellular automaton is undecidable, and in order to verify if a certain cellular automaton is an accurate representation of reality this problem would need to be solved. Finally, even if we would happen to have such an algorithm, we cannot physically construct a machine that executes it for both practical and theoretical reasons.

Criticizing DP in such a radical way would make anyone feel that the entire field is wasted brainpower. However, I argue that this is not entirely the case. DP has tied concepts from physics, mathematics, philosophy and computer science together into a more unified scientific discipline. Doing this allows us to reason more broadly about these concepts. Fundamentally, the exact sciences all rely on the same mathematical axioms and logic. The objective of physics is to use these axioms to formalize the workings of nature. Computer science, on the other hand, specializes in what we can and cannot formally compute. I have argued that there is a link between computer science and epistemology in that what we can *compute* also tells us something about what we can *know*. Looking at it this way, we can apply these restrictions and limitations of computer science to physics and philosophy as well. This elevates us to a better understanding of why certain principles in physics are troublesome and why certain principles seem ununifiable with each other, perhaps mainly because of the formal approach we take. Taking a step back, taking observation and information into account, looking at the limitations of formal systems we can discover, even our structure of thinking, we can understand why things appear strange and weird to us.

Laplace envisioned that one who would know the complete state of the universe, together with the rules that determined the evolution of the states of universe, would instantly be enlightened with the knowledge of past, present, and future. Given the way mathematics has evolved, this is an understandable analytical hypothesis. Most, if not all, authors in DP would not disagree with Laplace's thought experiment, as they too believe in the mathematical supremacy of man over the cosmos. They too would envision a demon able to compute the states of the universe even on a simple abacus. With that, the demon seems to have power over the universe. But perhaps we are let astray in this thought experiment by the supposedly endless power of mathematics, not aware of its limitations. After all, the anthropocenteredness of this thought experiment should not be forgotten. Combining physics and mathematics with philosophy and computer-science can bring us more understanding of these limitations. Taking steps to understand what constraints come along with our formalizations of the real world will ultimately bring us closer to understanding our relation to this real world.

List of Abbreviations

- **DP**: Digital Physics/Digital Philosophy - Both names for the collection of scientific theories that revolve around information and computation as fundamental to physics.
- **PAP**: Participatory Anthropic Principle - The principle that no phenomenon is a real phenomenon until it is observed.
- **ERH**: Extern Reality Hypothesis - The idea that an external reality independent of human observation exists.
- **MUH**: Mathematical Universe Hypothesis - The hypothesis that the universe is ontologically mathematical.
- **CUH**: Computational Universe Hypothesis - The hypothesis that the mathematic universe consists only of (Turing) computable definitions.
- **TOE**: Theory of Everything - A hypothetical theory that unifies all fields of physics
- **TM**: Turing Machine - A mathematical model of computation invented by Alan Turing. It is proven that any other mathematical model of computation is rewritable to a turing machine. The turing machine is the most popular way to model computation in computer science today.

Appendix A - Basic Principles of Computer Science

At its core, computer science studies the act of computing. Especially within theoretical computer science, the efficiency of algorithms on formal models of computation are analyzed. Computation models formalize the steps required to take, to get to a result of a mathematical function. Mathematicians needed such a model in order to make claims about the computational process required for evaluating mathematical functions. For example, the theoretical performance of an algorithm can be proven without having to depend on practical implementations. The most often used model of computation of the last 50 years is the Turing machine.

The Turing machine (TM) is the mathematical representation of a mechanical machine that has four elements: a tape, a head, a register and an instruction table. The tape is of infinite length and has numbered cells that can contain any symbol (but usually only 0 and 1). The tape head is able to move to any cell, to the left and to the right, and read the value of the cell into the register. Based on this value, a lookup is performed in the instruction table. The instruction can either be to write the current cell with a different value or do nothing. Typically the instruction then also commands to which cell the tape head should move to next.

Many variants of this basic TM model exist. Models with multiple tapes, models with read-only tapes etc. All these other models are proven to be *equivalent* to the basic TM. This means that there is some translation that enables one to rewrite a special TM into a basic TM. This begged mathematicians in the early part of the 20th century the question: "Which models of computation can be rewritten to a Turing Machine?". Surprisingly, all of the existing models of computation (λ-calculus, recursive functions, cellular automata and Kahn process networks) were rewritable to a TM form. This lead to the *Church-Turing thesis*, which roughly states that any model of computation can be rewritten into a TM. This also implies that all models of computation can be rewritten to any other. In general it states that that what we call *computation* is exactly a process that is redefinable as a Turing Machine. This thesis, however, is not proven. The thesis is generally accepted within computer science up until this day based on the lack of a counterargument. And because the definition of ’computation’ by the Church-Turing thesis uses the Turing Machine as the pivotal computational model, computer science as a whole does so as well.

A very important proof that Alan Turing managed to deliver with his Turing Machine in 1936 was the undecidability proof for what later became known as the *halting problem*. In the halting problem it is asked to define an algorithm (thus a specific Turing Machine configuration) that takes as input *any* algorithm and decides if this algorithm, if executed, will run forever or at some point output a value and halt. Turing proved that such an algorithm *cannot exist*. He did this by hypothesising an algorithm H that can solve the problem. H is then wrapped in an algorithm G as a subroutine, where G inverts the output of H. If H outputs *false*, for "no this input program does not halt", then it halts. If H outputs *true*, for "yes this input program halts", then it goes into an infinitely looping subroutine. The novelty of the proof now comes from the hypothesis that G can be fed into itself as an input. A logical inconsistency arises as logically G now outputs that the input program halts if it does not halt, and that it does not halt if it does halt. This contradiction proves that the initial assumption was incorrect. H cannot exist and thus the halting problem is undecidable.

Appendix B - Cellular Automata

A cellular automaton is a model of computation that works differently from a TM, but where a translation still exists. A cellular automaton is usually defined as a grid of cells (most commonly in 2 but in any finite amount of dimensions) with a certain state, usually 0 or 1, i.e. on or off. Instead of a register that shifts through the cells to change the state like a TM, in a cellular automaton all cells are pushed into a next state simultaneously. This is done according to a certain rule. Usually the rule looks at the state of the cell in relation to its neighbours to determine its next state, but this is not required.

A popular cellular automaton is *Conway's Game of Life*. The idea behind this game is that it mimics life based on population. This cellular automaton has the following rule set:

1. If a live cell has fewer than two neighbours, the cell dies.
2. If a live cell has two or three neighbours, it lives.
3. If a live cell has more than three neighbours, it dies.
4. If a dead cell has three neighbours, it lives.

The reader is certainly invited to play the Game of Life using any of the many instances to be found online! The fascinating property of this rule set is that given any arbitrarily chosen starting pattern, it creates a seemingly highly complex simulation, while still based on very simple rules. It is this observation that eventually lead DP researchers like Edward Fredkin, Gerard 't Hooft and Stephen Wolphram to believe that a cellular automative principle (based on simple rules, producing vast complexity) was at the root of physics.

References

Barbour, J. (2015). Bit from It. In A. Aguirre, B. Foster, & Z. Merali (Eds.), *It from bit or bit from it?* (pp. 197–211). Springer. Retrieved from http://link.springer.com/10.1007/978-3-319-12946-4_17 doi: 10.1007/978-3-319-12946-4{_}17

Bennett, C. H. (1973). Logical Reversibility of Computation. *IBM Journal of Research and Development*, *17*(November), 525–532.

Berto, F., & Tagliabue, J. (2017). Cellular Automata. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2017 ed.). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/fall2017/entries/cellular-automata/

Bohr, N. H. D. (1963). The Unity of Human Knowledge. In *Essays 1958–1962 on atomic physics and human knowledge* (pp. 8–16). Wiley.

Bostrom, N. (2003). Are you living in a computer simulation? *Philosophical Quarterly*, *53*(211), 243–255. Retrieved from https://www.simulation-argument.com/

Copeland, B. J. (2020). The Church-Turing Thesis. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Summer 202 ed.). Metaphysics Research Lab, Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2019/entries/church-turing/

Deutsch, D. (1985, 7). Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *400*(1818), 97–117. Retrieved from http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1985.0070 doi: 10.1098/rspa.1985.0070

Fodor, J. A. (1981, 1). The Mind-Body Problem. *Scientific American*, *244*(1), 114–123. Retrieved from http://www.nature.com/doifinder/10.1038/scientificamerican0181-114 doi: 10.1038/scientificamerican0181-114

Fredkin, E. (2003). An Introduction to Digital Philosophy. *International Journal of Theoretical Physics*, *42*(2), 189–247. Retrieved from https://link.springer.com/article/10.1023/A:1024443232206 doi: https://doi.org/10.1023/A:1024443232206

Gauch, H. G. (2003). *Scientific method in practice*. Cambridge: Cambridge University Press.

Gibbs, P. (1996). *What is Occam's Razor?* Retrieved from http://math.ucr.edu/home/

baez/physics/General/occam.html
Hacking, I. (1990). *The Taming of Chance*. Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/product/identifier/9780511819766/type/book doi: 10.1017/CBO9780511819766
Humphreys, P. (2016). *Emergence*. Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780190620325.001.0001/acprof-9780190620325 doi: 10.1093/acprof:oso/9780190620325.001.0001
InspiringPhilosophy. (2013). *Digital Physics Argument for God's Existence.* Retrieved from https://www.youtube.com/watch?v=v2Xsp4FRgas&t=8s
Isaac, A. M. C. (2018, 7). Physical Computation: A Mechanistic Account. *The Philosophical Review*, *127*(3), 426–431. Retrieved from https://read.dukeupress.edu/the-philosophical-review/article/127/3/426/135289/Physical-Computation-A-Mechanistic-Account doi: 10.1215/00318108-6718882
Jannes, G. (2009, 4). Some Comments on "The Mathematical Universe". *Foundations of Physics*, *39*(4), 397–406. Retrieved from http://link.springer.com/10.1007/s10701-009-9286-9 doi: 10.1007/s10701-009-9286-9
Jones, M. (2014). *John Wheeler's Participatory Universe.* Retrieved from https://futurism.com/john-wheelers-participatory-universe
Laplace, P. S. (2009). Essai Philosophique sur les Probabilités. In *Essai philosophique sur les probabilités* (5th ed., pp. 1–67). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/product/identifier/CBO9780511693182A005/type/book_part doi: 10.1017/CBO9780511693182.001
Leifer, M. S. (2015). *It From Bit or Bit From It?* (A. Aguirre, B. Foster, & Z. Merali, Eds.). Cham: Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-12946-4 doi: 10.1007/978-3-319-12946-4
Lloyd, S. (2006, 7). Programming the universe: a quantum computer scientist takes on the cosmos. *Choice Reviews Online*, *43*(11), 43–6586. Retrieved from http://choicereviews.org/review/10.5860/CHOICE.43-6586 doi: 10.5860/CHOICE.43-6586
Lucas, J. R. (1961, 4). Minds, Machines and Gödel. *Philosophy*, *36*(137), 112–127. Retrieved from https://www.cambridge.org/core/product/identifier/S0031819100057983/type/journal_article doi: 10.1017/S0031819100057983
Margolus, N. (1984, 1). Physics-like models of computation. *Physica D: Nonlinear Phenomena*, *10*(1-2), 81–95. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/0167278984902525 doi: 10.1016/0167-2789(84)90252-5
Mueller, M. P. (2017, 12). Law without law: from observer states to physics via algorithmic information theory. Retrieved from http://arxiv.org/abs/1712.01826
Petersen, A. (1963, 9). The Philosophy of Niels Bohr. *Bulletin of the Atomic Scientists*, *19*(7), 8–14. Retrieved from http://www.tandfonline.com/doi/full/10.1080/00963402.1963.11454520 doi: 10.1080/00963402.1963.11454520

Piccinini, G. (2015). *Physical Computation*. Oxford: Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199658855.001.0001/acprof-9780199658855 doi: 10.1093/acprof:oso/9780199658855.001.0001

Prokopenko, M., Boschetti, F., & Ryan, A. J. (2009, 9). An information-theoretic primer on complexity, self-organization, and emergence. *Complexity*, *15*(1), 11–28. Retrieved from http://doi.wiley.com/10.1002/cplx.20249 doi: 10.1002/cplx.20249

Rummens, S., & Cuypers, S. E. (2010). Determinism and the paradox of predictability. *Erkenntnis*, *72*(2), 233–249. doi: 10.1007/s10670-009-9199-1

Schmidhuber, J. (1997). A computer scientist's view of life, the universe, and everything. In *Foundations of computer science* (pp. 201–208). Retrieved from http://link.springer.com/10.1007/BFb0052088 doi: 10.1007/BFb0052088

Schmidhuber, J. (2000, 11). Algorithmic Theories of Everything. Retrieved from http://arxiv.org/abs/quant-ph/0011122

Shannon, C. E. (1948, 7). A Mathematical Theory of Communication. *Bell System Technical Journal*, *27*(3), 379–423. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6773024 doi: 10.1002/j.1538-7305.1948.tb01338.x

Smilga, W. (2013). *It from Bit, or Bit from It?* Retrieved from https://fqxi.org/community/forum/topic/1807

Tegmark, M. (2008, 2). The Mathematical Universe. *Foundations of Physics*, *38*(2), 101–150. Retrieved from http://link.springer.com/10.1007/s10701-007-9186-9 doi: 10.1007/s10701-007-9186-9

't Hooft, G. (2016). *The Cellular Automaton Interpretation of Quantum Mechanics* (Vol. 185). Cham: Springer International Publishing. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/3319412841http://link.springer.com/10.1007/978-3-319-41285-6 doi: 10.1007/978-3-319-41285-6

Timpson, C. G. (2013). *Quantum Information Theory and the Foundations of Quantum Mechanics* (1st ed.). Oxford: Oxford University Press. Retrieved from http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199296460.001.0001/acprof-9780199296460 doi: 10.1093/acprof:oso/9780199296460.001.0001

Toffoli, T., & Margolus, N. (1987). *Cellular Automata Machines: A New Environment for Modeling (Scientific and Engineering Computation)*. Cambridge, MA: MIT Press.

Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230–265. Retrieved from http://doi.wiley.com/10.1112/plms/s2-42.1.230 doi: 10.1112/plms/s2-42.1.230

van Strien, M. (2014, 3). On the origins and foundations of Laplacian determinism.

Studies in History and Philosophy of Science Part A, *45*(1), 24–31. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0039368113001325 doi: 10.1016/j.shpsa.2013.12.003

Wall, M. (2018). *We're Probably Living in a Simulation, Elon Musk Says.* Retrieved from https://www.space.com/41749-elon-musk-living-in-simulation-rogan-podcast.html

Wheeler, J. A. (1989). Information, Physics, Quantum: The Search for Links. In *Proceedings iii international symposium on foundations of quantum mechanics* (pp. 354–368). Retrieved from https://jawarchive.files.wordpress.com/2012/03/informationquantumphysics.pdf

Whitehead, A. N. (1925). *Science and the Modern World*. London: Cambridge University Press. Retrieved from https://archive.org/details/b29978531

Whitworth, B. (2007). The physical world as a virtual reality: a prima facie case. *Research Letters in the Information and Mathematical Sciences*, *11*, 44–60. Retrieved from https://arxiv.org/ftp/arxiv/papers/0801/0801.0337.pdf

Wigner, E. (1967). *Symmetries and reflections*. Cambridge, MA: MIT Press.

Wilson, R. A. (1986). *Cosmic Trigger Volume I: Final Secret of the Illuminati.* Retrieved from http://rawilson.com/trigger1.html

Wolfram, S. (2012). What Is Ultimately Possible in Physics? In *A computable universe* (pp. 417–433). World Scientific. doi: 10.1142/8306

Wolpert, D. H. (2008, 7). Physical limits of inference. *Physica D: Nonlinear Phenomena*, *237*(9), 1257–1281. Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0167278908001243 doi: 10.1016/j.physd.2008.03.040

Yates, D. (2017, 2). Demystifying Emergence. *Ergo, an Open Access Journal of Philosophy*, *3*(20200523), 809–841. Retrieved from http://hdl.handle.net/2027/spo.12405314.0003.031 doi: 10.3998/ergo.12405314.0003.031

Zahedi, R. (2015, 1). On Discrete Physics (Digital Philosophy/Digital Cosmology) and the Cellular Automaton: A Perfect Mathematical Deterministic Structure for Reality as a Huge Computational Simulation. *SSRN Electronic Journal*. Retrieved from https://www.ssrn.com/abstract=2624143 doi: 10.2139/ssrn.2624143

Zenil, H. (2012). *A Computable Universe*. World Scientific. Retrieved from https://www.worldscientific.com/worldscibooks/10.1142/8306 doi: 10.1142/8306

Zuse, K. (2012, 12). Calculating Space (Rechnender Raum). In A. German & H. Zenil (Eds.), *A computable universe* (pp. 729–786). World Scientific. Retrieved from http://www.worldscientific.com/doi/abs/10.1142/9789814374309_0036 doi: 10.1142/9789814374309{_}0036