SCUOLA
NORMALE
SUPERIORE

Scuola Normale Superiore di Pisa

CLASSE DI SCIENZE UMANE
Corso di Perfezionamento in Filosofia

Ph.D. degree in Philosophy

Paradoxes of the Applied Infinite

Infinite Idealizations in Physics

Candidate:
Pauline Manninne Anna van Wierst
Matricola 19261

Thesis supervisor:
Prof. Massimo Mugnai

Research advisor:
Prof. Giovanni Valente

Thesis submitted in March 2018

M.C. Escher (1960), *Circle Limit IV*. Woodcut in black and ocre

One concept corrupts and confuses the others. I am not speaking of the Evil whose limited sphere is ethics; I am speaking of the infinite.
- Jorge Luis Borges (1964), *Other Inquisitions, 1937-1952*

Contents

Acknowledgements **v**

Introduction **1**

I Infinite idealizations: Philosophical issues **7**

1 Infinite idealizations: Philosophical issues **9**
1.1 What are infinite idealizations? 10
1.1.1 A question of consistency? 11
1.1.2 Against external inconsistency: Physics as a patchwork . . 16
1.1.3 Sources of inconsistency: Modeling practice 19
1.2 What are legitimate infinite idealizations? 21
1.2.1 "Nearly infinite" systems: Working in the limits 22
1.2.2 Consistency again: Idealization versus approximation . . . 25

II Mathematics: The science of the infinite **31**

Introduction to Part II **33**

2 Ontology and the ideal in mathematics **35**
2.1 Measuring the infinite: Infinite numbers 36
2.2 The actual infinite: Set theory 42
2.3 The potential infinite: Constructive mathematics 47
2.4 The infinitely small: Nonstandard analysis 53
2.5 Idealization in mathematics 60

3 The continuum **65**

3.1 Classical conceptions of the continuum 67

3.2 Non-classical conceptions of the continuum 73

3.3 Against using the classical continuum in physical modeling 80

Conclusion to Part II **87**

III Two paradoxes of the applied infinite **93**

Introduction to Part III **95**

4 Paradox I: Phase transitions **97**

4.1 The physics 97

4.2 The philosophical debate 101

4.3 Analysis 108

4.4 Conclusion 121

5 Paradox II: Norton's Dome **125**

5.1 The physics 125

5.2 The philosophical debate 129

5.3 Analysis 137

5.4 Conclusion 143

Conclusion to Part III **147**

Acknowledgements

Writing this dissertation would not have been possible without the support of many people. Here, I want to express my graditude to them.

To start, I want to thank my supervisors. I thank prof. Mugnai for accepting me as his student, for making it possible for me to stay a significant amount of time abroad, and for giving me the freedom to pursue the current project without any prior knowledge of physics or its philosophy. I thank Giovanni Valente for having been involved with my work and my career, starting long before I asked him to supervise this dissertation. I learned a lot from his comments on my work, which were always very much to the point, and it has been a real pleasure to work with him. Most of all, I want to thank him for the fact that from the moment we met at the end of my second year in the PhD, he urged me to focus on my life *after* the PhD, whereas I myself was unconvinced that there would ever be such a thing.

Another person that I cannot thank enough is Arianna Betti. Her support has been decisive at various points in my life and at various stages in my academic career, like when she encouraged me to apply for the PhD program at the SNS in Pisa. I thank her for having given me the skills and the confidence that were necessary to bring this PhD to a good end, and for being a friend as well as a mentor and a source of inspiration.

This thesis has its origin in a talk about infinite idealization in physics by Patricia Palacios, which I attended during the wonderful MCMP Summerschool on Mathematical Philosophy for Female Students in the summer of 2015. Patricia's talk inspired me immediately, and she encouraged me to pursue the current project. Also later on I benefited a lot from talking to her about physics and infinite idealizations in general, and about phase transitions in particular.

Further, I want to thank the MCMP for welcoming me as a visitor in winter 2016-2017, and for their financial support by means of the Junior Visiting Fellowship they granted me. I benefited very much from the inspiring environment at the MCMP and from talking to many experts in philosophy of science and physics. In particular, I want to thank Alexander Reutlinger, Stephan Hartmann, and Neil Dawar for their interest in my work, and for conversations about philosophy of physics and philosophy of science in general, Erik Curiel for talking to me about physics and for commenting on a (very) early draft related to chapter 4, and Sam Sanders for an inspiring conversation about classical and non-classical mathematics, which influenced particularly section 2.5.

Another important factor in the development of this thesis was my research stay at the University of Salzburg. Also here I benefited very much from the stimulating enviroment. I thank the Philosophy Department for welcoming me as a visitor for more than a year and for having provided me with all the facilities that I needed; I thank all people in the department for the warm welcome and the pleasant time I enjoyed. I thank in particular Charlotte Werndl for her support and for the time and effort she spent commenting on my work.

A special thanks to my awesome office mates in Salzburg, Insa Lawler and Mariangela Zoe Cocchiaro, which with their friendship and words and signs of support helped me get through the last months of writing this thesis.

A very warm thanks also to my wonderful Dutch friends, Lottie, Elise, Kees, and Noël, who did not give up on me even though I did not live in the same country as them for five years, and who have been a distant but very important source of support and happiness. I feel very blessed with friends like you.

As always, my brother Rense has been very important to me during the PhD as well. I thank him for believing in me more than anybody else, for coming to visit me all over the world, and for the fact that, even though life is not always easy on him, he remains such a cheerful and sweet hearted person.

I thank also the family Ginammi, Giovanni, Luigina, and Marco, for the warm welcome into the family, for all the good care, and especially for the endless amounts of delicious polenta accompanied by a continuous flow of red wine from Val di Gerra.

Getting to know Michele Ginammi, at the very first day of the PhD, was the best thing that ever happened to me. I have been a very privileged person, spending these years with him by my side. Michele introduced me to the treasures of life in Italy. And he contributed to this thesis in many ways: he read and commented on several drafts and all material related to this thesis; he listened to me when I needed to articulate my ideas, was a sparring partner when I needed one, and remained silent when that was what I needed instead. Even more importantly, he brightened up my days and kept my world going. He took full care of our dog Luce and cooked me delicious meals when I needed to focus on my work, and he made every minute of my free time wonderful. *Ma com'è bella la vita con te!*

Introduction

> *[G]ewiß die meisten* paradoxen *Behauptungen, denen wir auf dem Gebiete der Mathematik begegnen, sind Sätze, die den Begriff des* Unendlichen *entweder unmittelbar enthalten oder doch bei ihrer versuchten Beweisführung in irgendeiner Weise sich auf ihn stützen. Noch unstreitiger ist es, daß gerade diejenigen mathematischen Paradoxien, die unsere größte Beachtung verdienen, weil die Entscheidung hochwichtiger Fragen in mancher anderen Wissenschaft, wie in der Metaphysik und Physik, von einer befriedigenden Widerlegung ihres Scheinwiderspruches abhängt, unter dieser Gattung sich finden.*
> - Bernard Bolzano (1851)

Mathematics plays a central role in science. But how is it possible that mathematics helps us understand the physical world? How the connection between mathematics and the world is to be accounted for remains one of the most challenging problems in philosophy of science, philosophy of mathematics, and general philosophy (Mancosu 2008).

Understanding how *idealization* works is an integral part of addressing this problem. Idealization is, roughly, the representation of reality in a deliberately simplified or distorted manner. Examples of idealization in science are abstracting the problem by leaving out certain features of the real situation, or approximating the real situation by using values for variables that are wrong but in some sense 'close enough' to the real situation, or by using approximating mathematical techniques (Ladyman 2008). It is a central problem in philosophy of science how to account for the fact that our best scientific theories, while invoking such simplifications or distortions of reality, nevertheless give us knowledge of the world and attain empirical adequacy.

Particularly puzzling are *infinite idealizations*, i.e. mathematical representations of a physical system in which some property of the system, which in reality is finite, is idealized as infinite. Examples are representing matter as continuous, modeling a system as having an infinitely big particle number and volume, or considering a process during an infinite amount of time. Infinite idealizations are ubiquitous in science, but pose numerous questions of a philosophical nature and often lead to *paradoxes*. These paradoxes are usually of a conceptual, rather than technical, nature: the mathematical formalism works fine, but problems arise concerning the *interpretation* of the formalism. An example is the well-known 'paradox of phase transitions': our best theories of phase transitions are in statistical mechanics, where phase transitions *only* occur in the so-called 'thermodynamic limit', i.e. in systems idealized to have an infinite particle number and volume. As it was famously put by Callender (2001), it thus seems that these theories give us a *mathematical proof* that the finite systems around us cannot display phase transitions. But we see water boil into vapor or freeze into ice on a daily basis, so something must have gone wrong here!

The paradoxes surrounding infinite idealizations in physics show that – at least in these particular cases – something is different than we expected in our thinking about how mathematics applies to the physical world. Investigating them might thus help us to better understand the applicability of mathematics in general, and the role and character of infinite idealizations in particular. Why is infinite mathematics so useful in modeling a finite world? How can an infinite value be 'close enough' to any finite value? Are infinite idealizations always used for mere convenience, i.e. to simplify the mathematics, or are there cases in which they are 'essential'? Can (apparent) reference to infinity in the physics be 'explained away' in finite terms? In recent philosophy of science, infinite idealizations and their paradoxes are discussed in order to clarify these and other questions concerning the applicability of mathematics and the character and role of (infinite) idealizations in science. Getting clear on these issues is complicated by the fact that infinite idealizations relate to other philosophical issues, such as for example intertheory relations and scientific explanation, and thus one's understanding of infinite idealizations generally depends on one's take on these other issues.

Importantly, in the philosophical debate on infinite idealizations in science, it is largely taken for granted that 'infinity' means *actual infinity*. Actual infinity is the concept which underlies set theory and – consequently – regular present-day (i.e. 'classical') mathematics and physics. However, there seems to be no *a priori* reason why it should be (mathematical physics based on) set theory which applies (best) to the physical world (cfr. Isham & Butterfield 2000). It is clear that set theory and its notion of actual infinity are *not* indispensable to science: various developments in mathematics show that much more modest means suffice. Furthermore, it has long been recognized by philosophers of mathematics and philosophically minded mathematicians that set theory and the notion of actual infinity have their own philosophical problems, and that there are good philosophical reasons to prefer alternative foundational schemes.

With this in mind, the philosophical problems with infinite idealizations appear in a different light. For if the paradoxes arise only in physics based on set theory, i.e. if they are relative to classical mathematics, then these philosophical problems are *not* a problem for idealization or applicability *per se*, but rather for a *specific* kind of idealization or applicability – i.e. the application of mathematics based on set theory to the physical world. Therefore, as I see it, we should investigate whether the paradoxes arising from infinite idealizations disappear or change if the physics is formulated in a different mathematical framework, based on a different notion of infinity. Are the philosophical problems with infinite idealizations relative to classical mathematics and its notion of infinity?

Thus, in my view there are two related problems:

A. How do infinite idealizations help us understand the finite world?

B. What is the nature of the notion of infinity used in classical mathematics?

The first question is topic of debate in philosophy of science; the second is a central topic in the philosophy if mathematics. Surprisingly, however, the two questions are hardly ever discussed in relation to one another. Sporadically, this is recognized in the literature. Fletcher (2002, p. 3) writes, for example: "These questions of the nature of infinity in pure mathematics cannot be seen in isolation from applied mathematics and physics. It is a great defect of the literature on the paradoxes of mathematical infinity that it ignores the paradoxes of physical infinity, and vice versa." Waaldijk (2005, p. 24) – after noticing that in many discussions in modern physics it is taken for granted, without explanation, that classical mathematics is the only means available when describing the real world – asserts that one "cannot emphasize enough that the foundational issues of modern mathematics are of vital importance to the foundational issues of physics (and vice versa)." One outstanding work in which these questions *are* addressed together, is the great Bernard Bolzano's *Paradoxes of the Infinite* (*Paradoxien des Unendlichen*, 1851). Although we will hardly discuss his work directly, Bolzano very much inspired this thesis, and it is to acknowledge this influence that I choose the title.

Accordingly, the aim of this thesis is to take question A and B together, and ask:

> **Central question**: Are the philosophical issues surrounding infinite idealizations in physics relative to classical mathematics and its notion of infinity?

We will set out to answer this question specifically for two cases of philosophically paradoxical infinite idealizations in physics. The **two case studies** that we will undertake in this thesis, are (1) the paradox of phase transitions, and (2) the paradox of 'Norton's Dome', i.e. a case of indeterminism in Newtonian mechanics. Both paradoxes are well-known in the philosophy of science literature. We will approach these paradoxes in this thesis by focussing on the underlying mathematical framework, and in particular, on the notion of infinity employed in the physics. We will aim to understand in which sense and for which reason the infinite idealizations in the case studies are necessary or desirable, and what makes these idealizations philosophically problematic. This will allow us to see whether developing the physics in an alternative mathematical framework would be philosophically beneficial.

To be clear, in this thesis we will undertake *philosophical reflection* on the question whether or not the philosophical paradoxes in the case studies are relative to classical mathematics and its notion of infinity. Ideally, of course, I would deepen my investigation and strengthen my argumentation by actually developing the relevant physical theories in mathematics based on alternative foundational schemes, and show how the philosophical problems change or disappear accordingly. Since it is unfeasable to undertake that work in this thesis, I will rely on the work of mathematicians, physicists, and other philosophers of mathematics and physics for the technicalities regarding the relevant mathematics and physics.

The **original contribution of this thesis** consists in bringing two fields – philosophy of science and philosophy of mathematics – together in order to think about infinite idealizations in a new way.

The thesis is **set up** as follows. It has three parts. **Part I** is called *Infinite idealizations: Philosophical issues*, and consists of one chapter with the same name. The goal is to introduce the reader to the philosophical issues around the use of infinite idealizations in science in general. We will introduce these issues by means of a summary of the debate on infinite idealizations in the philosophy of science literature. **Part II** is *Mathematics: The science of the infinite*. The goal of the second part is to introduce the reader to philosophical issues surrounding the mathematical infinite. It consists of two chapters. In chapter 2 we will discuss several different mathematical conceptions of the infinite and the corresponding mathematical systems that are built upon them. In chapter 3 we will focus on philosophical issues concerning (various developments of) the mathematical continuum. Both the first and the second part have an introductory character and serve as a preparation for **Part III**: *Two paradoxes of the applied infinite*, in which we will discuss two case studies of physical theories involving infinite idealizations which turn out to be philosophically problematic. These are: statistical-mechanical theories of phase transitions (chapter 4), and a case of indeterminism in Newtonian mechanics known as 'Norton's Dome' (chapter 5). Part II and Part III consist of multiple chapters, and have therefore their own introduction and conclusion. In the conclusion of Part III we will draw consequences from the case studies in relation to all the other material discussed in this thesis, and as such will function as the general conclusion of this dissertation.

Part I

Infinite idealizations: Philosophical issues

Chapter 1

Infinite idealizations: Philosophical issues

The aim of this chapter is to provide context for the discussion of the case studies which take central stage in this thesis. This chapter will be mainly expository; a critical assessment of the literature will be done specifically with regard to the case studies, and thus will be postponed until Part III of this thesis.

In recent philosophy of science, infinite idealizations in science in general, and in physics in particular, have received a considerable amount of attention. They are relevant, and have been discussed, in various contexts. Part of these contexts concern more general debates, such as with regard to the applicability of mathematics (for example, questions such as: how does infinite mathematics apply to the finite physical world?), the practice of idealization in general (e.g. under which conditions are infinite idealizations admissible?), the realism-antirealism debate (e.g. do theories containing infinite idealizations give us (approximate) truth?), and debates about scientific explanation (e.g. if infinite idealizations are false assumptions, then how can they be part of an explanation?). Another part of these contexts are more specific, such as the debate on intertheory relations (e.g. do newer, more refined theories reduce to older, less encompassing theories in some sort of limit?), and on emergence (e.g. if limits are essential to represent a certain phenomenon, does this mean that this phenomenon is emergent, i.e. cannot be explained in terms of what happens at a lower level?). In this literature, we find, on the one hand, authors which consider infinite idealizations (in some cases) necessary and give them substantial philosophical import (e.g. Batterman 2002, Batterman 2005, Liu 1999, Shech 2013), as well as, on the other hand, authors which deny their necessity in all cases, and search for ways to 'explain away' the apparent reference to infinity (e.g. Norton 2012, Norton 2014, Norton 2016, Bangu 2009, Pincock 2014).

In this chapter we will discuss, to start, different views on what infinite idealizations are – in particular, whether we should understand them as introducing *false assumptions* into a model, or rather as *simplifications* of a model, i.e. the elimination of details (section 1.1). We will see that part of the debate on this matter centered around the issue of *consistency*, for infinite idealizations are often understood to introduce inconsistencies within a theory (subsection 1.1.1). We will also discuss a view according to which this focus on consistency between theories is misguided (subsection 1.1.2), and an account of – on the assumption that the world is consistent and our scientific theories (to some extent) describe this world – how it can be that different scientific theories are inconsistent with each other (subsection 1.1.3). In the second section we will discuss how it can be that infinite idealizations *work*, i.e. contribute to empirical adequacy of the theory containing them (subsection 1.2.1), and an account of what distinguishes *legitimate* from *illegitimate* idealizations (subsection 1.2.2).

1.1 What are infinite idealizations?

Idealization, in general, can be understood as the deliberate simplification or distortion of reality with the objective of making it more tractable (Frigg & Hartmann 2006). In the literature, there have been given several classifications of different kinds of idealizations, characterized by the degree and manner in which they distort and/ or simplify. Idealizations which primarily simplify have been called *abstractions* (Cartwright 1989), *Aristotelian idealizations* (Frigg & Hartmann 2006), or *minimalist idealizations* (Weisberg 2007). These idealizations consist in 'stripping away', in our imagination, all properties from a concrete object that we believe are not relevant to the problem at hand, so that we can focus on a limited set of properties in isolation. An example is a classical mechanics model of the planetary system, describing the planets as objects only having shape and mass, disregarding all other properties (Frigg & Hartmann 2006). Another example is the *Ising model*, an important model in the study of phase transitions, in which atoms, molecules, or other particles are represented as points along a line which can be in one of two states. The model is extremely simple, building in almost no realistic detail about the substances being modeled but the interactions and structures that really make a difference, or the core causal factors giving rise to the target phenomenon (Weisberg 2007). Idealizations which primarily distort are commonly called *Galilean idealizations*, because it was characteristic of Galileo's approach to science to use simplifications of this sort whenever a situation was too complicated to tackle (McMullin 1985, Frigg & Hartmann 2006, Weisberg 2007). Examples are the assumption that the Earth is flat over a short distance or that a rolling ball is perfectly spherical (McMullin 1985). Galilean idealization is performed due to the present computational limitations, and there is the hope that ultimately the simplifying assumptions can be removed, making the theory more realistic (McMullin 1985, Weisberg 2007). In sum, whereas in primarily simplifying, *Aristotelian* idealizations certain properties which are deemed irrelevant are disregarded and this simplification is taken to be permanent, primarily distorting, *Galilean* idealizations involve false assumptions about the physical system, which one aims to ultimately remove.

So how about *infinite idealizations*? Should we understand them as simplifications, i.e. do they eliminate details, or rather as distortions, i.e. do they involve false assumptions (or both)? In the literature there seems to be no agreement on this issue, and the same infinite idealizations are by some identified as distortions and others as simplifications. An example is the analysis of water waves, in which the ocean is idealized as infinitely deep. This example was put forward by Maddy (1992) in the context of the indispensability arguments for mathematical realism, and then inspired a discussion of how to interpret infinite idealizations. The example makes it clear that infinite idealizations relate to the issue of *consistency* between different mathematical theories.

1.1.1 A question of consistency?

Do our best scientific theories commit us to the existence of mathematical objects? Much of the current thought on mathematical ontology involves in some way or another the famous Quine-Putnam *indispensability argument*.[1] The use of

[1] The argument is formulated in Putnam (1967). Putnam attributes the argument to Quine.

mathematics in science is often appealed to as the main reason to be some kind of realist about mathematical entities. The general idea of the indispensability argument has been formulated by Maddy in the following way:

> "We have good reason to believe our best scientific theories, and mathematical entities are indispensable to those theories, so we have good reason to believe in mathematical entities. Mathematics is thus on an ontological par with natural science. Furthermore, the evidence that confirms scientific theories also confirms the required mathematics, so mathematics and science are on an epistemological par as well" (Maddy 1992, p. 78)

However, scientific theories eliminate detail and precision, and often involve assumptions which are known not to be true. As Maddy famously argued, any freshman physics text is "littered with applications of mathematics that are expressly understood not to be literally true: e.g., the analysis of water waves by assuming the water to be infinitely deep or the treatment of matter as continuous in fluid dynamics or the representation of energy as a continuously varying quantity" (Maddy 1992, p. 281). She continues to note that "this merely useful mathematics is still indispensable; without these (false) assumptions, the theory becomes unworkable" (ibid). In other words, as argued by Maddy, the indispensability argument for mathematical realism is blocked by acknowledging that science involves *idealization*: if false assumptions are part of our best scientific theories, and we do not accept their truth even though they are indispensable, then the same position is available for the mathematical entities that are indispensable to those theories.[2]

Colyvan used Maddy's observation to argue that idealization comes down to *inconsistency* between different pieces of theory (2008, 2009). In relation to Maddy's example of the analysis of water waves, Colyvan points out that – besides

[2] There is another problem with the indispensability argument: *if* our best scientific theories were to commit us to the existence of mathematical entities, then to *which* mathematical entities exactly? Both Quine and Putnam were led to accept significant portions of set theory on the basis of the indispensability argument. However, as pointed out by Feferman (1998, p. 285), neither of them undertook a detailed examination of *how much mathematics is needed* for scientifically applicable mathematics to arrive at their positions, nor did they consider whether any of the alternative foundational schemes ought to be preferred on philosophical grounds. This problem with the indispensability argument will not concern us in this chapter; we will come back to it in Part II of this thesis.

assumung an infinitely deep ocean in the calculations – we also deploy sonar to determine the finite depth of the ocean. According to him, "what we are really dealing with here is a contradiction between two pieces of theory. Taking the conjunction of the two pieces of theory, we have it that oceans are both infinitely deep and not infinitely deep" (Colyvan 2008, p. 117). Colyvan is willing to bite the bullet, and accept the indispensability argument's implication that sometimes we ought to believe in the existence of inconsistent objects.

However, about the water wave case, Colyvan notes that

> "there is nothing too troubling here. The sonar theory is surely correct and the assumption of infinitely deep oceans is a mere idealisation. It is clear what we ought to believe here (and that was Maddy's point). It is just that hard-nosed Quineans, she suggests, have trouble delivering the right answer" (Colyvan 2008, p. 117, fn. 4).

Colyvan agrees here with Maddy that in the water wave case, despite the fact that there is a contradiction between two pieces of theory, *it is clear what we ought to believe*: we believe sonar theory's affirmation that the depth of the ocean is finite, and not the assumption of an infinitely deep ocean which is part of the (otherwise accurate) analysis of water waves.

This is the spirit in which many scholars think about infinite idealizations. In fact, several scholars use the fact that infinite idealizations give rise to contradictions between different pieces of theory as a *reductio* against that idea that we should take the assumptions involved in infinite idealizations as literally true. For example, Norton famously argued that if we should believe statistical-mechanical theories of phase transitions – which, as we will see in chapter 4, require the *thermodynamic limit* and model phase transitions as discontinuous changes in the system's thermodynamic properties – then we should consider the atomic theory of matter disproved:

> "If the atomic theory of matter is true, then ordinary thermal systems of finitely many components cannot display discontinuous changes in their thermodynamic properties. The changes they manifest are merely so rapid as to be observationally indistinguishable from discontinuous behavior. Indeed, if we could establish that the phase transitions of real substances exhibit these discontinuities, we would

> have refuted the atomic theory of matter, which holds that ordinary thermal systems are composed of finitely many atoms, molecules, or components. It must be feared that a similar refutation is at hand, if the positing of infinitely many components is necessary to recover other observed behaviors of phase transitions" (Norton 2012, p. 225).

Certainly, no one would consider giving up the atomic theory of matter for the reason that statistical-mechanical accounts of phase transitions require an infinite system. It is clear here what to believe: that real systems which display phase transitions are composed of finitely many components – and if statistical mechanics tells us otherwise, then we should look for a different interpretation of statistical mechanics' appeal to infinite systems, argues Norton.[3]

Similarly, Butterfield (2011) argues – against Batterman (2002, 2005, 2010) – that many limits, such as the thermodynamic limit, in which a parameter N, encoding physical degrees of freedom or some analogous concept, goes to infinity, should not be taken to be physically real. For, as he argues, as N becomes very large, the model becomes unrealistic: it runs up against either the micro-structure of space and its contents (atomism), or the macro-structure of space and its contents (cosmology). As Butterfield stresses, "these break-downs are not internal to the model, but in relation to the actual world" (2011, p. 10).[4] Again, for Butterfield, it is clear what to believe, and this is the reason not to take infinite idealizations as literally true of the concrete systems they represent.

The fact that, in cases in which infinite idealizations lead to two (seemingly) contradicting pieces of theory, it is so clear what to believe, led Pincock (2014) to argue that *we do not actually believe* both contradicting pieces of theory. In particular, he seems to argue, we do not believe the false claims which are part of the idealization. In case of the analysis of water waves, according to Pincock, "we do not ever assume that the ocean is infinitely deep" (Pincock 2014, p. 2958). This is because, as Pincock puts it, "false assumptions take us from an interpreted part of a scientific representation to an idealized representation where the partic-

[3] Norton proposes to solve this problem by distinguishing between two different ways in which infinite idealizations can be used, one of which he calls *approximations* and the other *idealizations*. We will discuss Norton's proposal in section 1.2.2.

[4] It is not clear to me what exactly Butterfield means by a model running up against "the actual world". I suppose that we should read this as these models running up against *certain propositions* which we take to be true of the actual world, or against *other models* or *theories* which we take – in some sense – to represent the actual world more accurately than these models.

ular part is no longer interpreted" (Pincock 2014, p. 2963). Thus, in Pincock's view, the false assumption of an infinitely deep ocean, i.e. $H = \infty$, changes the representation in which H denotes the depth of the ocean to one in which H is a place holder. As he puts it himself, the infinite idealization transforms the representation so that it has merely *schematic content* (ibid.). Thus, according to Pincock, infinite idealizations should *not* be understood as false assumptions, but rather as a way to ignore certain aspects of the physical system which are deemed irrelevant.

In sum, we have seen that according to Colyvan, Norton, and Butterfield, the issue of consistency raised by idealizations – i.e. the problem that the assumptions which are part of the idealization contradict propositions of other sciences – is solved by in some sense attributing less truth to the infinite idealizations. In particular Norton and Butterfield argue that we should not take infinite idealizations to be *physically real*, i.e. literally true of concrete systems. The reason for this seems to be for all three of them that, in case of contradicting pieces of theory, it is clear what to believe: *not* the infinite idealization. It seems thus that all three of them understand infinite idealizations as *distortions*, i.e. they involve assumptions which are false about the concrete systems they represent. Pincock, to the contrary, argues that in infinite idealizations the parameter taken to infinity is *decoupled from its interpretation*. Thus, Pincock seems to understand infinite idealizations not as distortions, but rather as *simplifications*, namely ignoring the parameter which is taken to infinity. There is another argument offered in the literature to the effect that – at least to some extent – infinite idealizations, in particular the thermodynamic limit of statistical mechanics, are simplifications, namely that they involve the *abstraction of finitary effects*. We will discuss this argument in section 1.2.1.

Further, the consistency discussed in the context of infinite idealizations can be distinguished in two different kinds. First, there is inconsistency between different *pieces of theory*. This is the kind of inconsistency about which Maddy wrote, and about which Colyvan and Pincock were thinking: *within* the theory of water waves there are two contradicting propositions: one asserting a finitely deep, and one an infinitely deep ocean. Butterfield and Norton (at least in the above quote) seem to be concerned with inconsistencies between *different theories*, namely infinite component limits being inconsistent with atomism or cosmology.

We could call the first kind of inconsistency, i.e. inconsistency internal to a theory or model, *internal inconsistency*, and the second kind, i.e. the inconsistency of infinite idealization with the truths of another theory, *external consistency*. We will discuss internal inconsistencies in more detail in section 1.2.2. Before turning to that, we will first discuss an argument to the effect that external consistency is *not the right way* to think about infinite idealizations – and about physics in general.

1.1.2 Against external inconsistency: Physics as a patchwork

Why would it matter that infinite idealizations lead to external inconsistencies? Without doubt, the theories with which these infinite idealizations are inconsistent involve idealizations themselves, so why would consistency between different physical theories be relevant at all?

According to Batterman (2014), these worries about inconsistency between different physical theories and about the global unity of physics stem from a misguided but highly common picture of physics – he calls it *P*hysics with a capital "P". According to this picture, microlevel theories, i.e. theories that talk about the atomic and subatomic makeup of the things we see around us, are *privileged*: they are considered to be "fundamental", and are associated with what is "real", "true", and "physical" (Batterman 2014, p. 2974). Macrolevel phenomena can according to *P*hysics be *explained* in terms of these microlevel theories. According to Batterman, many philosophers and physicists think about physics in this way (despite recent debates about emergence and failure of reduction), and it is exactly this picture which urges us to look for logical consistencies between different theories in physics:

> "If we maintain the hierarchical structure we become distracted by claims of logical inconsistency. These are, for well established theories, not in the least bit important. Physics should be seen as a patchwork of various theories each quite accurately characterizing and explaining their various domains of applicability. Typically, these are determined by the importance or unimportance of various lengths or time scales. Interesting problems arise at the borderlands between the members of this patchwork, but it is a mistake to look at those problems only

> through the lens of model theory with its measure of consistency" (Batterman 2014, p. 2991).

According to Batterman, focus on logical consistency between physical theories is a consequence of the mistaken view of physics as *P*hysics. If we take physics for what it is, namely a *patchwork* of various theories, Batterman argues, then we will see that looking for consistency is a mistake.

According to Batterman, that physics should be understood as a patchwork follows from, among other things, the fact that many macrolevel theories cannot be reduced to microlevel theories in any sense. Batterman gives an example of such a case: the micro- and macrolevel theories of breaking drops, i.e. water dripping from a faucet (Batterman 2014, cfr. Batterman 2005). As Batterman explains, in Navier-Stokes theory, which describes the shape of the drops from a macroscopic point of view, the phenomenon of the breakup – i.e. the moment in which a single mass of water changes into two or more droplets – is represented by a discontinuity that is characterized by divergences in both the fluid velocity and in the curvature of the interface at the point of snap-off (Batterman 2014, p. 2977).[5] In the microlevel theory, the theory of molecular dynamics, to the contrary, there is no such discontinuity. Thus, the (supposed) inconsistency between these two theories consists in Batterman's view in the fact that the macrolevel theory represents the breakup by a singularity, whereas there is no such singularity in the microlevel theory (Batterman 2014, p. 2980, n. 7).

Now, from the view point of *P*hysics, the continuum Navier-Stokes theory is an idealization: we know that water drops are "really" just finite collections of discrete molecules. The continuum Navier-Stokes theory is from this view point just an idealization which is appropriate to the macroscopic scale (Batterman 2014, p. 2979). However, as Batterman argues, Navier-Stokes theory has important *explanatory virtues*, namely, it explains what is known as "universality", i.e. the fact that many different systems – sometimes radically different in their microstructural makeup – can exhibit identical or nearly identical phenomenological behavior (Batterman 2014, p. 2976). There is, as Batterman writes, no way to account

[5] Batterman argues furthermore that this representation is correct: the qualitative topological changes that we see at the macrolevel *should be* represented mathematically by singularities (Batterman 2014, p. 2983). Similarly, the qualitative changes occurring at phase transitions should according to Batterman be represented mathematically by singularities. We will discuss this view in the context of phase transitions in section 4.2.

for universality from the molecular dynamical theory (Batterman 2014, p. 2983). Thus, what according to *P*hysicists are "inconsistencies" arising from idealization, have a very important theoretical and explanatory role. It is wrong, according to Batterman, to see these inconsistencies as problematic.

The fact that the microlevel theory is discrete, does not only imply that the breakup of the drop is not modeled as a discontinuity, but also that the *representation* of drops in the microlevel theory is slightly *different* from the representation in the macrolevel theory. That is, as Batterman puts it, the very idea of a well-defined drop with boundaries does not apply in the discrete theory of molecular dynamics (Batterman 2014, p. 2980, n. 7). We could thus say that the micro- and macrolevel theory both employ a (slightly) different conception of "drop" – and consequently of other concepts, such as "breakup". In noting that the mathematical counterparts of physical concepts change passing from one "patch" to another, Batterman sides with Wilson (2006). Wilson speaks of "theory facades" as the proper way to understand the organization of physical theories.[6] In the view of both Batterman and Wilson, different theoretical descriptions fit the phenomena in various ways, yet breakdown if pushed too far. As Batterman writes, such a picture of the theories of physics – as facades that work exceptionally well in restricted domains, but that cannot be extended indefinitely without conceptual change – eliminates broad worries about logical inconsistency between theories (Batterman 2014, p. 2991).

Thus, in Batterman's view, and contrary to the views that we discussed in the previous section, we should interpret the inconsistencies between different theories which result from infinite idealizations in macrolevel theories *not* as a reason to take these infinite idealizations (or the macrolevel theories to which they belong) as any less true than microlevel theories, but rather we should give up the global unity of physics and the related idea that macrolevel phenomena can be explained in terms of microlevel theories – that is, we should give up *P*hysics.

[6] We will see in chapter 5 that according to Wilson, not only physics as a whole consists of such facades, but that this is also the case for parts of physics that often – naively – are understood as one theory. This is the case, according to Wilson, for "classical" or "Newtonian mechanics", which in his view actually divides into three different formalisms: point particle mechanics, the physics of rigid bodies and perfect constraints, and continuum mechanics – each of them with descriptive gaps which reach all the way to their cores, and which often give mutually incompatible analyses of the same issue. Wilson calls Newtonian mechanics therefore "a stool constructed of six or seven legs of unequal length" (Wilson 2009, p. 174, cfr. Wilson 2013).

1.1.3 Sources of inconsistency: Modeling practice

If the physical world is consistent, then how is it possible that different theories which describe this world are inconsistent? In the literature, this question has – at least partly – been answered in terms of an analysis of the practice of modeling and theory building. It has been argued in various ways, that scientists do not usually deal with phenomena, or events in the world *simpliciter*, but rather with phenomena interpreted by means of theory and organized in stable patterns. This practice of organization of and pattern-seeking in the data makes that theories are not simply a record of the data collected in experiments: theory and data stand in a much more complex relationship. As Ladyman (2008) argues, no real system that is measured ever exactly fits the description of the phenomena that become the target of the theoretical explanation.

The complex relationship between results of measurement on the one hand, and theory on the other, has been brought to attention by Patrick Suppes in his seminal paper "Models of Data" (1962).[7] Famously, Suppes argued that there is a *hierarchy of models* of different types that connect data to theory:

> "[E]xact analysis of the relation between empirical theories and relevant data calls for a hierarchy of models of different logical type. Generally speaking, in pure mathematics the comparison of models involves comparison of two models of the same logical type, as in the assertion of representation theorems. A radically different situation often obtains in the comparison of theory and experiment. Theoretical notions are used in the theory which have no direct observable analogue in the experimental data" (Suppes 1969, p. 25).

Understanding how theory and experiment connect to each other in Suppes' view requires understanding the links between the different kinds of models in this hierarchy.[8] There are three levels of models in Suppes' system: models of theory,

[7] The paper is reprinted in (Suppes 1969), which is the version that I will be refering to.

[8] A model of a theory according to Suppes is a possible realization in which all valid sentences of the theory are satisfied, and a possible realization is an entity of the appropriate set-theoretical structure (1969, p. 24). According to Suppes, the logician's notion of a model is the basic and fundamental concept of model and is needed for an exact statement of any branch of empirical science (cfr. 1969, p. 17).

models of experiment, and models of data. The idea[9] is that as one descents from the hierarchy, one gets closer to the data and the actual details of the experimental experience. Models of theory are at the top of Suppes' hierarchy and are the most abstract models. They contain idealizations and notions which have no directly observable analogue, such as continuous functions and discontinuities. One step down the hierarchy we find models of the experiment, which contain reference to a particular experiment, although they also contain the idealizations that are part of the model of the theory. The experimental model is to say what the theory would entail in the specific experimental set-up under consideration, and would contain, for example, possible values of the volume and pressure in this particular experiment (cfr. Harris 2003). Below the model of the experiment we find in Suppes' hierarchy the model of the data (also 'data model'). In the data model, the experimental results are summarized and put in such a form that they allow for application of analytical methods and statistical assesment of fit between predictions and actual data. Data models would for example contain readings of volume and pressure taken during a particular run of an experiment. Importantly, data models are already 'polished', in the sense that what seem to be 'experimental errors' are filtered out, and stationarity (i.e. whether the present parameters are constant over time or trials) is tested.

In recent literature, it has been stressed even more that data models are already theory-laden.[10] As is argued in this literature, in practice scientists use elements of the theory and other features of interest to guide the production of data models. There are always competing patterns in data, and scientists must choose one pattern over other competing patterns. Which pattern in the data gets singled out is guided by the interest of the scientist and incorporates already elements of theory. As argued by McAllister (1997, p. 224): "[I]nvestigators discover the patterns that are exhibited in data sets, but stipulate that some of these correspond to the phenomena."

In sum, there are various steps between the "raw" data collected in experiments and the final theory which is inferred from them at which these data are "shaped". Ascending in Suppes' hierarchy of models from the data model to the

[9] Suppes' paper is sketchy and leaves much to be clarified. A better worked-out exposition of Suppes' ideas – taking his points in some respects a bit further – can be found in (Mayo 1996). A discussion of Suppes' ideas and an account related to it can also be found in (van Fraassen 2008).

[10] See Cartwright (1999), Harris (2003), McAllister (1997), Morrison (1999).

model of theory, at each step the model becomes more and more abstract and idealized. Models of the data are different from models of the theory, for example in that models of the theory contain notions (such as, e.g., probability measures, or discontinuous changes in one or more variables) which are not directly observable and are not part of the recorded data (Suppes 1969, p. 26). But, as stressed in recent literature, already the construction of data models involves interpretation: which pattern in the data gets singled out is guided by the interest of the scientist and incorporates already elements of theory. In singling out the patterns in the data which are to be captured in the data model, and in the process of abstraction which leads from data models to models of theory, there are many choices to be made which can give rise to theories with very different characteristics. For these reasons, different theories which model a consistent world can be inconsistent with each other.

1.2 What are legitimate infinite idealizations?

We have seen in the previous section that idealizations in general, and infinite idealizations in particular, are in some sense falsifying in nature: they consist in the deliberate simplification or distortion of reality with the goal of improving tractability. But how much and in which manner may we distort and simplify reality so that we still have a legitimate idealization? What distinguishes a legitimate infinite idealization from an outright falsehood?

Physicists themselves seem not to worry about this kind of questions; they rely on an intuitive understanding of how far they can take idealizations.[11] They might – seriously or less seriously – justify idealizing systems as infinitely large by saying that finite systems are so big that they are "nearly infinite". It is a task for philosophers of science to make these intuitions explicit. Part of understanding what distinguishes legitimate from illegitimate infinite idealizations, is understanding *why* infinite idealizations *work* when they do. In subsection 1.2.1, we will see that there is indeed a sense in which concrete systems are "nearly infinite", which explains the success of certain infinite idealizations. Care is needed, however: whereas legitimate idealizations can be obtained by some limiting pro-

[11] For example, mathematically rigorous existence proofs for the thermodynamic limit are rare in physics; in general, physicists are satisfied if it is *intuitively* clear that certain properties converge and thus their limits exist (Griffiths 1972).

cesses, by others we obtain illegitimate idealizations or no idealization at all. In subsection 1.2.2, we will discuss Norton's distinction between *idealization* and *approximation*, and his proposal of how to distinguish legitimate from illegitimate infinite idealizations.

1.2.1 "Nearly infinite" systems: Working in the limits

Infinite component limits, such as the thermodynamic limit, are common place in statistical mechanics, which aims to explain the thermal properties of materials in terms of the behavior of the microscopic components of these systems. These microscopic components of the systems under consideration are usually very large in number, which means that we cannot hope or wish to follow any motion of their individual constituents, but that instead we must describe their average or typical properties through some sort of statistical treatment. Physicists commonly justify working with infinite component limits by arguing that since the systems we see around us have very large N, they are "effectively" or "nearly" infinite (Mainwood 2005, p. 26). However, to say that a number is "nearly infinite" does not make sense: *every* finite number, no matter how big, is closer to zero than to infinity.[12] So, what do these physicists mean?

According to Bangu (2009), in most cases,

> "...what motivates the physicists' relaxed attitude in this matter is not a suspect metaphysical easiness with infinities but rather an outright dismissal of the whole finite versus infinite business on the basis of considerations having to do with the limits of experimental accuracy. Since it is virtually impossible to point out observable differences between the behavior of infinite systems and systems featuring a really big number of components (of the order of 10^{23} or larger), the philosophers' worry (do finite systems really undergo [phase transitions]?) becomes immaterial. As the physicist Baierlein once joked, "It all works because Avogadro's number is closer to infinity than to ten"" (Bangu 2009, p. 495).

[12] To see this, take an interval with 0 and ∞ as it endpoints, and locate any finite number on this interval. Every finite number will be mapped to zero, for if not, by the Archimedean principle, a certain multiplication of the number will exceed infinity, which is impossible (cfr. Bell 2005, Lisker 1996). Therefore, every finite number is closer to zero than to infinity.

Thus, according to Bangu, the reason why infinite component limits work, and why, according to physicists, its use is justified, has to do with *limits to experimental accuracy*: it is, as he points out, virtually impossible to detect observable differences between the behavior of infinite systems and systems consisting of a really big number of components, i.e. of the order of 10^{23} or larger. This has to do with the fact that the thermal properties of statistical mechanical systems are a function of the number of their components, and that these properties settle down to stable values if the number of components becomes very large (cfr. Norton 2014, p. 197). Mathematically, these stable values can be approximated quite closely by taking a limit of the property, and the difference between the values of these properties for infinite systems and those for very large finite systems is often smaller than the measurement error. Therefore, the limiting properties provide a good approximation to the properties of finite systems. Thus, in other words, infinite component limits in statistical physics are justified by the facts that the properties of systems are a function of the number of components, and the values of these functions for big systems are so close to the values of these functions for infinite systems, that no difference between the limiting values and the values for big component numbers can be discovered.

It can be argued that macroscopic thermal systems are also "nearly infinite" in another sense. For finite and infinite systems are *qualitatively different* from each other (at least) in that the former, but not the latter, have boundaries. As it turns out, macroscopic systems are generally so large that boundaries have no significant effect on their thermal properties, so that boundary effects can safely be ignored (Lui 2001). Infinite component limits can thus also be considered as *abstraction from finitary effects* (Butterfield 2011, p. 19). Since boundary effects are absent in infinitely large systems, and boundary effects are in concrete systems largely insignificant, concrete systems can also for this reason be called "nearly infinite".

In sum, concrete, and thus finitely large, thermal systems can thus be said to be "nearly infinite", *not* in the sense that their component number is almost infinite, but rather in the sense that their properties are empirically indistinguishable from the properties of infinitely large thermal systems. This has to do both with the fact that the values of the functions describing their thermal properties are empirically indistinguishable from the values of these functions for infinite systems, and with the fact that for large thermal systems, boundary effects can be ignored. Thus, in this manner, idealizing systems as infinitely large can be justified.

Not all infinite idealizations involve idealizing a system as infinitely large; another example is treating collections of discrete particles as if they formed a continuous substance. Many classical continuum models of fluids and solids, for example, are obtained by taking a limit of a classical atomistic model as the number of atoms N tends to infinity (in an appropriate way, e.g. keeping mass density constant), but *without* letting the system size go to infinity (Butterfield 2011, p. 19). Such continuum models are mathematically often much more convenient than their discrete counterparts.[13] Why do infinite idealizations of this type work?

It is often the case that idealizing a molecular structure as a continuum works because it turns out that the microscopic makeup of the system is *irrelevant* for the phenomenon under consideration. The phenomenon of "universality" is an example: many different systems with completely different microscopic makeup exhibit identical behavior (Batterman 2002). An example of such universal behavior we discussed in section 1.1.2 above: Navier-Stokes theory, which describes the motions of fluids from a macroscopic point of view and models fluids as a continuum, shows universality of the shape of breaking droplets (Batterman 2005, Batterman 2014). Thus, in analogy with understanding idealizations of finitely large systems as infinitey large as abstraction from finitary effects, we can understand the idealization of discrete systems as a continuum as an abstraction of the microscopic structure.

[13] An example is given by (Butterfield 2011, section 3.3.3): Consider the mass density (i.e. the ratio of the mass of the fluid to its volume) varying within a fluid. For an atomistic model of the fluid, that postulates N atoms per unit volume, the average mass-density might be written as a function of both position x within the fluid, and the side-length L of the volume L^3 centred on x, over which the mass-density is computed: $f(N, x, L)$. Now, for fixed N, this function is very sensitive to x and L, meaning that if atoms are or contain point-particles, the function will jump when L is varied so as to include or exclude one such particle. Thus, such a function in the atomistic model will not be continuous. But by taking a continuum limit $N \to \infty$, with $L \to 0$ (and atomic masses going to zero appropriately, so that quantities like density do not "blow up"), we can define a continuous, maybe even differentiable, mass-density function $\rho(x)$ as a function of position. As such, in the continuous model we can enjoy all the convenience of the calculus.

1.2.2 Consistency again: Idealization versus approximation

How much and in which manner may we idealize, so that we still have a *legitimate* idealization? An account of what distinguishes legitimate from illegitimate idealizations has been proposed by Norton (2012, 2014). His proposal builds on the recognition – which was already longer present in philosophy of science literature[14] – that there are *two ways of using limits*: as what Norton calls *idealizations*, and as what he calls *approximations*. The key difference is referential: idealizations carry a novel semantic import not carried by approximations. As we will see, according to Norton, a minimal condition for legitimate idealizations is that they are *consistent* – or to be more precise, limits can be used as idealizations only if they give rise to a consistently describable system (Norton 2014, p. 200). In this subsection, we will discuss Norton's proposal to distinguish between idealizations and approximations in some detail. This idea will be relevant again for both our case studies: the paradox of phase transitions (chapter 4), and the paradox of Norton's Dome (chapter 5).

In order to clarify the distinction between *idealizations* and *approximations* as Norton proposes it, we will discuss an example of his (Norton 2014, section 2). Dilute gases under normal conditions of temperature and pressure conform quite closely to the ideal gas law:

$$PV = nRT \tag{1.1}$$

where P is the pressure, V the volume, n the number of moles, T the temperature, and R is the ideal gas constant. While often good, the fit of the ideal gas law to real systems is never perfect, Norton explains. Errors become larger if, for example, one consideres regimes of high pressure and high density.

Now, according to Norton, we can see the ideal gas law in two ways: as an idealization or as an approximation. In his terminology, an *approximation* is an inexact description of a target system. It is propositional. An *idealization* is a real or fictitious system, distinct from the target system, some of whose properties provide an inexact description of some aspects of the target system. (As Norton stresses, these are not definitions; they merely specify important properties.) Thus, the ideal gas law as an *approximation* is a proposition which describes

[14] E.g. Frigg & Hartmann (2006), Butterfield (2011). For an overview, see Mainwood (2005).

some target system inexactly. So seen, the ideal gas law is an empirical generalization, according to Norton. As an *idealization*, to the contrary, the ideal gas law introduces a fictitious system of which the ideal gas law is an exact description. This system consists of very many non-interacting spatially localized components. The system is fictitious since there are no real systems like that: all molecules occupy some volume of space and interact at least weakly with other molecules. This fictitious system, Norton explains, is an idealization of the target systems of real dilute gases and dilute solutions, because it looks like those systems in the particular aspects that interest us; that is, in the relationship between their pressures, temperatures and other magnitudes in the ideal gas law.

We can always demote an idealization into an approximation by discarding the idealizing system and preserving only the relevant propositions that describe it (Norton 2014, section 2). A corresponding promotion from approximation to idealization will not always be possible, for idealizations do not always exist. Whether an idealization exists or not depends, as Norton sees it, on two things: (1) whether they can exist at all, i.e. whether the limiting process results in a system that is *consistent*, and, if yes, (2) whether these idealizations have the *intended properties*, i.e. whether they resemble finite systems in the ways that interest us (Norton 2012, Norton 2014). Norton gives two simple mathematical examples to illustrate how (1) and (2) can fail, together with a third in which everything goes as desired (Norton 2012, section 3). We will repeat his examples here.

The case in which everything goes as desired is the following. Consider a sphere of unit radius. It is elongated into a capsule, a cylinder with spherical end caps, as shown in figure 1.1. Its total length grows through the sequence of cylinder lengths $a = 1, 2, 3, 4, \ldots$. In the infinite limit, i.e. letting the cylinder length go to infinity, the capsule becomes an infinite cylinder of unit radius. Now we are interested in the following property: what is the ratio of area to volume of these capsules? As Norton explains, the surface area of a capsule of cylinder length a is $2\pi a + 4\pi$, and its volume is $\pi a + \frac{4\pi}{3}$. Hence, the ratio of surface area to volume is $2\pi a + 4\pi \ / \ \pi a + \frac{4\pi}{3}$, and the ratio approaches a limiting value of 2 as a goes to infinity. The limiting system, i.e. the infinite cylinder, has a ratio of area to volume of 2. Thus, we see that the limit of the properties of the sequence of capsules agrees with the corresponding properties of the limiting system.

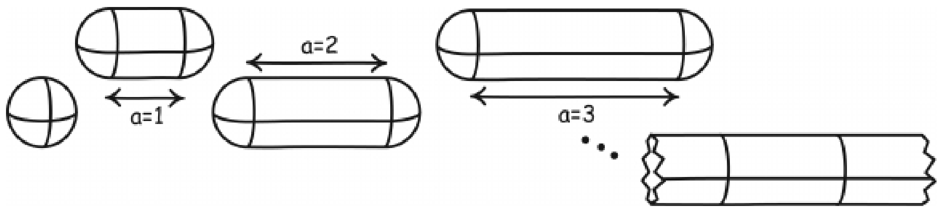

Figure 1.1: Limit property and limit system agree (after Norton, 2012)

Norton clarifies his example with a general scheme:

System 1,	System 2,	System 3,	...,	Limit System;
Property 1,	Property 2,	Property 3,	...,	Limit Property.

Limit system and limit property cohere in this case, for the limit property is the corresponding property of the limit system. Therefore, according to Norton, the infinite cylinder is an *idealization* of the larger capsules.

Things are different in the following example, i.e. a case in which there is no limit system. Consider a unit sphere whose radius r grows as $r = 1, 2, 3, \ldots$, as shown in figure 1.2. The area of the sphere is $4\pi r^2$, and its volume is $4\pi r^3/3$. Again, we are interested in the ratio between the area and the volume. As Norton explains, the ratio of surface area to volume is $4\pi r^2 \ / \ \frac{4\pi r^3}{3} = \frac{3}{r}$, and this ratio goes to zero as the radius r goes to infinity. Hence, the sequence of properties has a limiting value, i.e. zero. The sequence of systems, that is, of spheres, however, has no limit system. For, as Norton explains, a sphere is a set of points equally far away from some center; an infinitely large sphere would consist of points infinitely far away from the center, but there are no such points: all points in the space are some finite distance from the center. As Norton argues, talk of an "infinitely large sphere" is literally nonsense; an infinitely large sphere is impossible.

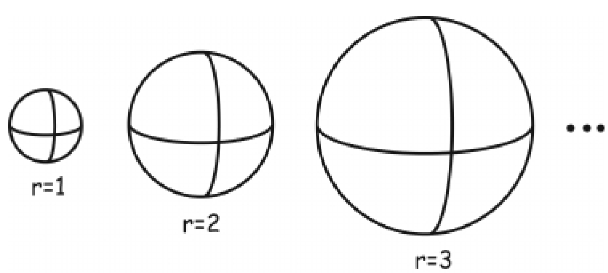

Figure 1.2: There is no limit system (after Norton, 2012)

The scheme that we have in this case is:

System 1,	System 2,	System 3,	...,	(No Limit System);
Property 1,	Property 2,	Property 3,	...,	Limit Property.

In this case there is a limit property, but it is *not* a property of a limit system, because a limit system *does not exist*. As Norton explains, the zero area-to-volume ratio is not a property of an impossible infinite sphere, but rather it is a property of the set of all finite spheres; it is the greatest lower bound of the ratios of the set's members. In this case, the limit property is thus an *approximation*: an inexact description of the properties of the later, larger members of the sequence of systems. The limit process provides no idealization, because there is no limit system to bear the limit property.

In the third example there is a limit system, but it does not have the intended properties. The example is the following. Consider once again a sphere of unit radius. Uniformly expand it in one direction only, so it becomes an ellipsoid with semimajor axis a. Continue the expansion, letting a go to infinity. The limit system is a cylinder of unit radius, as shown in figure 1.3. The volume of the ellipsoid is $4\pi a/3$. The surface area of the ellipsoid nears a value of $\pi^2 a$, arbitrarily closely for large a. We are once more interested in the ratio of surface area to volume. As Norton explains, this ratio approaches $\frac{\pi^2 a}{4\pi a/3} = \frac{3\pi}{4}$ as a goes to infinity. However, as we have seen in the first example, an infinite cylinder, which in the present example is the limit system, has a ratio of surface area to volume of 2. Thus, the limit ratio of the sequence of expanding ellipsoids is *not* the same as the corresponding ratio of the limit system, i.e. the infinite cylinder.

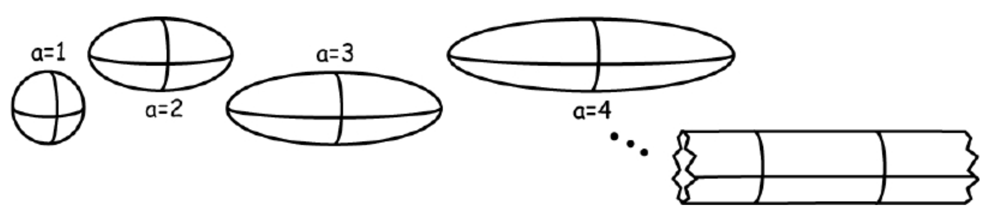

Figure 1.3: Limit property and limit system disagree (after Norton, 2012)

The scheme that we have in this case is:

System 1,	System 2,	System 3,	...,	Limit System;
Property 1,	Property 2,	Property 3,	...,	Limit Property.

However, in contrast to the first example, in this case limit system and limit property *disagree*: the limit property is not a property of the limit system. As Norton argues, the limit property can be used to provide an *approximation* of the systems leading up to the limit, for the limiting ratio of $\frac{3\pi}{4}$ is a close approximation of the area to volume ratio of very large ellipsoids. However, the limit process does *not* provide an *idealization*, for the limit system does not have the *intended* properties: an infinite cylinder has an area to volume ratio of 2, which is a poor approximation of the ratio for large ellipsoids.

To take stock, on Norton's account, limiting processes result in idealizations, i.e. limiting systems, when the properties of the limiting system are (1) consistent, and (2) intended, that is, they correspond to the limiting properties of the sequence of finite systems. When either of those conditions is not fulfilled, i.e. when there is no limiting system or when the limiting system does not have the intended properties, then one can use the limiting properties as approximations, i.e. propositions which inexactly describe large but finite systems. The kind of consistency that Norton requires as a minimal condition for idealizations, we could call – in contrast to what we called in section 1.1.1 *external consistency*, i.e. the consistency of infinite idealizations within one theory with the propositions of another theory – *internal consistency*, for it concerns the question whether the system is in itself consistent, i.e. whether or not it has contradicting properties.

In cases in which the limit system in Norton's view does not provide an idealization because the limit system does not have the intended properties, we see that – in some sense – the limit system is *qualitatively different* from the systems in the sequence up to the limit. As we have seen in Norton's third example namely, the systems with increasingly large semimajor axis a are ellipsoids, whereas the limiting system with infinite a is no longer an ellipsoid, but a cyllinder instead. This qualitatively different character of limit systems with respect to the finite systems up to the limit is in the philosophical debate on infinite idealizations related to the issue of *novelty*: infinite idealizations are often used to obtain representations which in some respect are qualitatively different from the descriptions of finite systems up to the limit. This is the case for example with systems in the thermodynamic limit of statistical mechanics: only in the infinite limit systems, and in none of the finite systems leading up to this limit, do the derivatives of the free energy change discontinuously.[15] A related question which is debated on the philosophical literature, is whether this qualitatively different, novel behavior which arises in the limit should be interpreted as *physically real*. Norton's position that such infinite idealizations are actually not idealizations, but only approximations, agrees with the position of Butterfield (2011), who argues that the physically real behavior occurs *before* one reaches the limit. The limit system itself, should according to both not be taken as physically real, but only as an approximation of finite, concrete systems. An opposite position is taken by Batterman (2002, 2005, 2014), who argues that, in virtue of this novel behavior, infinite idealizations are often *necessary* – i.e they cannot be eliminated or de-idealized – in order to get a qualitatively accurate representation of a phenomenon.

[15] This case will be discussed in chapter 4.

Part II

Mathematics: The science of the infinite

Introduction to Part II

> *If in summing up a brief phrase is called for that characterizes the life center of mathematics, one might well say: mathematics is the* science of the infinite*.*
> - Hermann Weyl (1949)

In chapter 1 we have seen that idealization is a problem for the Quine-Putnam indispensability argument for mathematical realism (section 1.1.1). But idealization is not the only problem with the indispensability argument. As we have seen, according to the indispensability arguments we should believe in the existence of those mathematical entities that are indispensable to our best scientific theories, but *which* mathematical entities are we talking about? That is, what exactly *is* the mathematics that is essential to our best scientific theories? Both Quine and Putnam were led to accept significant portions of set theory on the basis of the indispensability argument, but neither of them undertook a detailed examination of *how much mathematics is needed* for scientifically applicable mathematics to arrive at their positions, nor did they consider whether any of the alternative foundational schemes ought to be preferred on philosophical grounds (Feferman 1998, p. 285).

This tendency to assume that set theory is indispensable to science is characteristic for philosophy of science. But this assumption is mistaken: various mathematicians have shown that much more modest means suffice. To name two examples, Feferman (1998) showed, via proof-theoretical reduction, that it is possible to directly formalize almost all, if not all, scientifically applicable mathematics in a formal system that is justified simply by *Peano arithmetic*, and Ye (2011) developed almost all of the mathematics that is currently used in the physical sciences within *Strict finitism*, i.e. a fragment of quantifier-free primitive recursive arithmetic with the accepted functions restricted to elementary recursive functions. Furthermore, various arguments have been put forward to the effect that these and other alternative mathematical frameworks are *preferable* to set theory *on philosophical grounds*.

The goal of this second part is, first, to introduce set theory (and the mathematics based on it, i.e. what we will call ‘classical mathematics’) and some of the philosophical issues with it, and second, to show that there are good alternatives to set theory which in some respects are philosophically preferable. Thus, this second part is meant to show that *classical mathematics should not be taken for granted* in philosophy of science. As such, this part will serve as a *justification* for the project undertaken in this thesis, i.e. to look critically at the use of classical mathematics in two – paradoxical – cases in physics.

Chapter 2 will be an introduction to both set theory and two of its alternatives, i.e. constructive mathematics and nonstandard analysis, and will contain an analysis of the differences between these frameworks in terms of their respective ontologies. Chapter 3 will focus on issues surrounding the mathematical continuum. In this second part we will discuss some philosophical issues raised by classical mathematics and some philosophical benefits of some of its alternatives in general terms; in part III we will apply this discussion to the specific context of the case studies.

Chapter 2

Ontology and the ideal in mathematics

Mathematics is the science in which we don't know what we are talking about.*
- Bertrand Russell

** Don't care would be more to the point.*
- Martin Davis[1]

Infinity occurs in many shapes and forms in mathematics. For example, the transfinite numbers in set theory are very different from the infinity involved in a limiting process $lim_{n \to \infty} a_n$, and from the infinite and infinitesimal quantities that occur in nonstandard analysis. Historically, the infinite – both in the form of the infinitely large and of the infinitely small – has always been at the center of crises in foundations of mathematics (Dauben 1988). With the general acceptance of set theory as the foundational framework for mathematics, it might seem as if the infinite received its definite mathematical shape – and many see it indeed this way.

A minority of mathematicians, however, objects to set theory. Such objections generally arise from fundamentally differing views concerning the nature of mathematics and the objects with which it deals (Feferman 1998). In turn, these differing views concerning the nature of mathematics lead to alternative views

[1] Quotes after Dauben (1988).

about the meaning of words such as *or*, of *implication*, and of *quantification*, and consequently about what is a *proof* in mathematics.

In this chapter, we will discuss several of the different shapes that infinity takes in mathematics. We will start with an introduction of the concept of infinite *number* (section 2.1). Next, we will introduce set theory: the theory of the *actual infinite* (section 2.2), constructive mathematics: the theory of the *potential infinite* (section 2.3), and nonstandard analysis: the theory of the *infinitesimal* (section 2.4). Finally, we will compare these three mathematical systems with respect to their *ontology*, and we will see that also in mathematics there is *idealization*.

2.1 Measuring the infinite: Infinite numbers

One of the central issues concerning mathematical infinity – an issue that occupied the greatest minds in Western thought for millennia – is whether it can be *measured* (Mancosu 2009). That is to ask, in other words, can the concept of *number* be extended to infinite sets? Throughout history various thinkers have come up with different approaches to infinite sets in order to develop a general notion of number – that is, a notion of number applicable to both finite and infinite sets –, but nobody managed to do so in a mathematically satisfactory manner until Cantor developed his *set theory*.

The reason for the trouble in developing a general notion of number is the paradoxical character of the infinite. To be precise, the problem is that regarding to the notion of *size* we have two pre-theoretic intuitions, which contradict each other in case of infinite sets (cfr. Benci, Di Nasso & Forti 2006, Mancosu 2009, Parker 2013). These intuitions are:

PW The whole is strictly bigger than any of its proper parts;

HP Two sets of which the elements can be put in one-to-one correspondence are equally big.

We call the first PW for part-whole principle, the second HP for Hume's principle.[2] PW and HP are perfectly compatible as long as the sets under consideration are

[2] Hume (*Treatise* I, iii, 1), was quoted by Frege when he gave his contextual definition of cardinal number (*Grundlagen*, §63): "When two numbers are so combined as that one has always an unite answering to every unite of the other, we pronounce them equal." The label 'Hume's principle' was introduced by Boolos (1987), and is now common in neo-logicist literature (cfr. Zalta 1998).

finite, for finite sets can be put in one-to-one correspondence just in case none of them is a proper part of another. For this reason, we are inclined to apply PW and HP indiscriminately when thinking about the notion of size. With regard to *infinite* sets, however, PW and HP are in conflict: infinite sets can be put in one-to-one correspondence with a proper part of themselves, and thus a proper part of an infinite set which is itself infinite (such as the set of squares, which is a proper part of the set of natural numbers) is *smaller* than the whole set on the basis of PW, but *just as big* as the whole set on the basis of HP.[3] For long, it has been generally acknowledged that it is impossible to develop a notion of size which is applicable to both finite and infinite sets and which respects both pre-theoretic intuitions PW and HP. Existing general accounts of size either sacrifice PW or HP, or deny that it is meaningful to compare infinite sets with respect to their size (cfr. Mancosu 2009).

As is well known, Cantor's big accomplishment was to recognize that HP – i.e. one-to-one correspondence between or *equinumerosity* of sets – is a fruitful concept to study and could provide us with a measure for infinite sets just as it does for finite ones (cfr. Potter 2004, pp. 153, 155). Of the pre-theoretic intuitions concering the notion of size, Cantor fully endorsed HP, and weakened PW down to (cfr. Benci et al. 2006, Mancosu 2009):

WPW The part is not bigger than the whole.[4]

On this basis, Cantor developed two general ways of counting, i.e. two concepts

[3] A proper *paradox* can be formulated as follows (Mancosu 2009, p. 630): Assume that both PW, if A is a subcollection of B then $s(A) < s(B)$, and HP, $s(A) = s(B)$ if and only if there is a one-to-one correspondence between A and B, hold for infinite sets. Let B be the set of natural numbers. Let A be the set of even numbers. Since $A \subset B$ by PW we have that A is smaller than B, $s(A) < s(B)$. But A and B can be put in one-to-one correspondence. So, by HP, A and B are equally big, $s(A) = s(B)$. Hence $s(A) < s(A)$. Contradiction.

[4] Cantor himself was perfectly aware of the fact that his theory only *partly* reflected our pre-theoretic intuition of size, for he stressed that *in some sense* the set of all natural numbers is bigger, namely *richer*, than the set of even numbers, while *in another sense*, namely in the sense of *cardinality*, both sets have the same size: "Let M be the totality (n) of all finite numbers n, M' the totality $(2n)$ of all even numbers $2n$. Here it is definitely correct to say that according to its entities M is richer than M'; indeed, M contains in addition to the even numbers, which make up M', also the uneven numbers M''. On the other hand, it is also definitely correct that both sets M and M' [...] have the same cardinal number. *Both (propositions) are certain and they do not conflict with each other* if one carefully observes the distinction between reality and number. One should therefore say: the set M has more reality than M', because it contains as parts M' and M'' in addition; the cardinal numbers corresponding to them are however equal. When will these easy and enlightening truths be finally acknowledged by all thinkers?" (Cantor 1887, my emphasis).

of *number* applicable to both finite and infinite sets.

The first kind of number that Cantor developed is the *cardinal number* or *cardinality*, $card(A)$ of a set A. According to the notion of cardinality, two sets have the same size if and only if they are equinumerous, i.e. if there exists a one-to-one correspondence between them (cfr. Potter 2004, Fletcher 2007):

$card(A) = card(B)$ iff A and B are equinumerous;

$card(A) \leq card(B)$ iff A is equinumerous with a subset of B.

The order of a set is disregarded in case of cardinal numbers: cardinals describe the size of a set in such a way that every method of counting gives the same result (Weisstein n.d.).

The second kind of number that Cantor developed is the *ordinal number*, or *order type* of a *well-ordered set* $(A, \leqslant)$.[5] Order types encode whether structures are isomorphic, i.e. whether there exist an order-preserving mapping between them (Potter 2004, p. 179):

$ord(A, r) = ord(B, s)$ iff (A, r) and (B, s) are isomorphic;

$ord(A, r) \leqslant ord(B, s)$ iff (A, r) is isomorphic to an initial subset of (B, s).

The ordinal numbers have the following structure: every ordinal number has an immediate successor known as a *successor ordinal*; and for any infinitely ascending sequence of ordinal numbers, there is a *limit ordinal* which is greater than all the members of the sequence and which is not the immediate successor of any member of the sequence. The first ordinal number is the empty set, Ø, the *finite ordinal numbers* are those obtained by starting with Ø and repeatedly taking the successor (Bagaria 2001). In set theory the natural numbers are defined as the finite ordinals. The first transfinite ordinal number, i.e. the first ordinal number greater than all natural numbers, but not an immediate successor of any of them, is ω. Thus, after all the finite numbers comes the first transfinite number, ω, which is followed by $\omega + 1$, $\omega + 2$, $\ldots$, $\omega + \omega = \omega \cdot 2$, $\ldots$, $\omega \cdot n$, $\omega \cdot n + 1$, $\ldots$, $\omega \cdot \omega = \omega^2$, $\omega^2 + 1$, $\ldots$, ω^ω, $\ldots$ and so on and on (Ferreirós 2007).

Note that ordinals describe the 'size' of a set in the sense of its numerical position in a sequence, whereas cardinals describe its 'size' regardless of order, only

[5] A set is *well-ordered* if it is equipped with an ordering relation under which every non-empty subset has a least element.

in terms of what is called the set's "power". That these are two different notions of size is clear from the fact that a given set may be bigger than another when regarded as an ordinal, but not when regarded as a cardinal. For example, ordinal numbers $\omega+1, \omega+2, \ldots$ are bigger than ω in the sense of order, i.e. as ordinals, but they are not bigger in the sense of equinumerosity, i.e. as cardinals (Weisstein n.d.). In fact, to every infinite cardinal correspond many infinite ordinals.

On the assumption of the *axiom of choice*,[6] which implies that every set can be well-ordered and can therefore be associated with an ordinal number, the cardinals can be enumerated through the ordinals; in fact, the two can be put into one-to-one correspondence (Weisstein n.d.). This leads to the definition of cardinal number for a set A as the least ordinal number b such that A and b are equinumerous (Weisstein n.d., Bagaria 2001). Cantor believed that every set could be well-ordered and used this correspondence to define the $\aleph$'s, "*alephs*". For any ordinal number α, $\aleph_\alpha = \omega_\alpha$ (Weisstein n.d.). All the finite ordinals are cardinals, and the first infinite cardinal, i.e. the cardinal number of $\mathbb{N}$ is denoted by $\aleph_0$. The sequence of the well-orderable infinite cardinals looks like this (Bagaria 2001):

$\aleph_0, \aleph_1, \aleph_2, \ldots, \aleph_\omega, \aleph_{\omega+1}, \aleph_{\omega+\omega}, \ldots, \aleph_{\omega^2}, \ldots, \aleph_{\omega^\omega}, \ldots$

Importantly, Cantor used his notion of cardinality to compare relevant sets and showed that there are infinite sets of *different sizes*. As Potter (2004, p. 153) writes, this showed that the notion of cardinality did not only give rise to a *coherent* notion of size, but also to a *fruitful* one. A pivotal step in the acceptance of Cantorian set theory was his famous *diagonal argument* to the effect that the cardinality of $\mathbb{R}$ is strictly greater than the cardinality of $\mathbb{N}$.[7] This argument shows that given any enumeration of a subset of $\mathbb{R}$, one can construct a number belonging to that subset that is not in the enumeration. In other words, the proof shows that the real numbers cannot be put in one-to-one correspondence, and thus are not equinumerous, with the natural numbers, so $card(\mathbb{R}) \neq \aleph_0$.

---

[6] We will discuss the axiom of choice in section 2.2 below.

[7] The proof based on the diagonal argument was Cantor's *second* proof to the effect that the cardinality of $\mathbb{R}$ is strictly greater than the cardinality of $\mathbb{N}$; first he gave a proof based on the special properties of $\mathbb{R}$. The diagonal proof is however very appealing for its simplicity, and moreover can be generalized to make other cardinal comparisons. It is used in the proof of *Cantor's theorem*, which states that for every set S, the *power set* of S, $\wp(S)$ – i.e. the set of all subsets S' of S – has a strictly greater cardinality than S. The set $\wp(S)$ of all subsets S' of S is called the "power set" of S, because for any set S it is equinumerous with $\{0,1\}^S$ (the correspondence puts 1 if the element i belongs to S' and 0 otherwise), i.e. $card(\wp(S)) = 2^{card(S)}$. For a summary of the proof, see e.g. Feferman (1998, p. 33).

Cantor's work introduced some new concepts. Sets that are equinumerous with $\mathbb{N}$ are called *denumerable*, and sets that are equinumerous with an initial segment of $\mathbb{N}$ (possibly empty) are called *finite*. A set is called *countable* if it is either finite or denumerable, and *uncountable* otherwise (George & Velleman 2002). Cantor's diagonal argument thus showed that $\mathbb{R}$ is uncountable.[8] Other consequences of Cantor's theory are, for example, that the set of positive *even numbers* $\mathbb{E} = \{0, 2, 4, \ldots, 2n, \ldots\}$ has the same cardinality as the set of the natural numbers $\mathbb{N} = \{0, 1, 2, \ldots, n, \ldots\}$, i.e. $\text{card}(\mathbb{E}) = \text{card}(\mathbb{N})$, because the two sets can be put in one-to-one correspondence. The same is true for the *rational numbers*, i.e. the set $\mathbb{Q} = \{\frac{n}{m} : n, m \in \mathbb{Z}, m \neq 0\}$ (where $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, \ldots\}$, i.e. the set of integers) consisting of all the quotients or ratio's of integers with non-zero denominators: $\text{card}(\mathbb{N}) = \text{card}(\mathbb{Q})$ (Feferman 1998, p 32). Thus, we clearly see here that cardinality gives us a notion of size on the basis of HP (see above): the set of even number is a *proper part* of the set of natural numbers, and the set of natural numbers is a *proper part* of the set of rational numbers, but according to the notion of cardinality all three of them have the same size.

Thus, Cantor showed that it is indeed possibile to develop a mathematically satisfactory notion of size which is applicable to both finite and infinite sets. Cantor's ideas were put to use more and more in mathematics, with the result that these days they are largely taken for granted and are spread throughout the whole of mathematics.[9] As it was put by Fletcher (2007), the fact that one can do calculations with infinities, i.e. that one can show, for example, that

[8] A natural question to ask is whether $\mathbb{R}$ is, after $\mathbb{N}$, the second smallest infinite set, i.e. whether $\text{card}(\mathbb{R}) = \aleph_1$ (Feferman 1998, p. 38). The conjecture that this is the case is called the *continuum hypothesis*. It was conjectured in 1878 by Cantor, but despite considerable effort he was unable to prove it. Today, as a result of the work of Gödel in 1938 and Cohen in 1963, we know that, if set theory is consistent, then the continuum hypothesis is not provable nor disprovable (George & Velleman 2002, p. 85).

[9] Serious worries arose when, at the beginning of the 20^{th} century, paradoxes appeared in set theory by taking its ideas to what appeared to be their logical conclusion. One example is *Russell's Paradox*, also known to Zermelo: consider the property of sets of not being members of themselves. If the property determines a set, call it A, then A is a member of itself if and only if A is not a member of itself. In order to avoid such paradoxes, the notion of set was adjusted as to exclude certain collections, like the collection of all sets, the collection of all ordinals numbers, or the collection of all cardinal numbers. Such collections are called *proper classes* (Bagaria 2001). When Zermelo developed an axiom system allowing one to develop Cantorian theory in full while avoiding all known paradoxical constructions, some measure of confidence in set theory was restored (Feferman 1998, p. 29). The now common axiomatization is called *Zermelo-Fraenkel set theory*, abbreviated ZFC (where C stands for the axiom of choice).

$5(\omega+1)(\omega^2 3+4)=\omega^3 3+\omega 4+5$ and $\aleph_{37}$ x $\aleph_{20}=\aleph_{37}$, convinced mathematicians that transfinite numbers were just as real as finite numbers, and that Cantor's theory represented a genuine advance over all previous thinking about infinity.

However, there are still a number of thinkers who object to the diffusion of set theory in mathematics. Partly, these objections stem from fundamentally differing views concerning the nature of mathematics and the objects with which it deals (Feferman 1998, pp. 29-30).[10] For another part, objections stem from dissatisfaction with the weird arithmetic properties of Cantor's transfinite numbers, and with the fact that the part-whole principle is not respected in the notions of size which underly set theory (Mancosu 2009).

These latter objections to set theory inspired recent mathematical developments that develop anonther general notion of number by generalizing the part-whole principle to infinite sets (see Mancosu 2009). The most noteworthy of those is Benci's *theory of numerosities* (Benci 1995, Benci & Di Nasso 2003, Benci et al. 2006, Benci, Bottazzi & Di Nasso 2014). Numerosities are, just as cardinals and ordinals, an extension of the notion of 'number' to infinite sets, but in such a way that the part-whole principle (PW) is completely respected. Hume's principle, however, is weakened to (Benci et al. 2006):

WHP Equinumerous sets are in one-to-one correspondence.

Numerosities have some nice properties which make that they are closer to our intuitions regarding the concept of number than are Cantor's ordinals and cardinals. For example, the numerosity of a proper subset is strictly smaller than the numerosity of the whole set, the numerosity of a disjoint union is the sum of the numerosities, and the numerosity of a Cartesian product is the product of the numerosities (Benci & Di Nasso 2003). All the standard algebraic laws for addition and multiplication hold for numerosities. As Mancosu writes, the theory of numerosities generalizes finite arithmetic much more thoroughly than Cantor's theory of ordinals or cardinal numbers (Mancosu 2009, p. 641).

In addition, numerosities can be interpreted as the set of hypernatural numbers, and hence be used to construct *infinitesimals.* Set theory, on the other hand, does not provide a natural way to introduce infinitesimal analysis. As Benci et al. (2006) put it themselves, numerosities offer a way of unifying the construction of

[10] We will discuss such objections in the next section.

the infinitely large and the infinitely small. Moreover, numerosities can be used to assign uniform probability distributions over infinite domains, such as for example the fair infinite lottery on $\mathbb{N}$ (Benci, Horsten & Wenmackers 2016). The theory of numerosities has however some less desirable features as well, such as for example the fact that the numerosities one obtains heavily depends on the choice of ultrafilter. As Mancosu (2009) shows, depending on whether the ultrafilter one chooses contains the even numbers or the odd numbers, it will turn out that this will affect such properties as whether the numerosity of the natural numbers is even or odd.

Both set theory and the theory of numerosities are based upon the idea that the finite and the infinite can be treated on the same footing. A precondition for the acceptance of both of them is that one should accept the consistency of the notion of infinite set, and thus the notion of actual infinity. In the next subsection, we will go on to discuss set theory, and consider some objections against it.

2.2 The actual infinite: Set theory

The essence of set theory is the study of infinite sets, and therefore it can be defined as the mathematical theory of the actual – as opposed to potential – infinite (Bagaria 2001). It studies well-determined collections called *sets*, of objects that are called *members* or *elements* of the set. In set theory, sets are given axiomatically, so their existence and basic properties are postulated by the appropriate formal axioms. The axioms of set theory imply the existence of a set-theoretic universe so rich that all mathematical objects can be construed as sets. Set theory is *foundational* in the sense that any mathematical object whatsoever can always be viewed as a set (or a proper class[11]). The properties of the object can then be expressed in the language of set theory. Any mathematical statement can be formalized into the language of set theory, and any mathematical theorem can be derived, using the calculus of first-order logic, from the axioms of ZFC, or from some extension of ZFC. It is in this sense that set theory provides a foundation for mathematics (Bagaria 2001, section 5).

According to Potter, one measure of the success of set theory is that it is nowadays a commonplace in the toolkit of most pure mathematicians; another

[11] See footnote 9.

measure of its success is that hardly any mathematician now thinks the existence of infinite sets might be logically inconsistent, or even incoherent. As Potter points out, if some are finitists and oppose to the notion of infinite set, it is because they are unconvinced by the positive arguments for the existence of infinite sets, not because they think there is a negative argument which shows that there are none (Potter 2004, p. 205, cfr. p. 69). However, there are still a number of thinkers who object to the panoply of set theory in mathematics. According to Feferman (1998, pp. 29-30), objections to set theory consist in fundamentally differing views concerning the nature of mathematics and the objects with which it deals.

What are the features of set theory which are considered problematic by its critics? Here we will discuss nine of them:[12]

1. *Sets as independent existents.* According to set theory, mathematics is the study of a fixed universe of actually-infinite mathematical objects, existing independently of our ability to construct them. Since sets are supposed to be part of an external, objective reality, it is commonly held that the notion of actual infinity *presupposes Platonism* about mathematical objects (see e.g. Weyl 1949, Fletcher 2007, Feferman 1998).[13] Platonism in set theory reveals itself most obviously in the *axiom of extensionality* (if two sets A and B have the same elements, then they are equal – or in other words, a set is determined entirely by its members and is to be regarded independently of any specific means of determining what these members are) and in the *axiom of choice* (see number 5 below; Bagaria 2001, Feferman 1998).

2. *Actual infinity.* It is essential to set theory that infinite sets exist as completed objects. As pointed out by Potter (2004, pp. 70, 72), there is very little that can be offered as a *justification* for the assumption of the existence of an infinite set (or an equivalent form of the axiom of infinity); the only possibility would be to justify it relative to some other infinitary theory

[12] The first seven objections are discussed in Feferman (1998, ch. 2), the last two in Benci & Di Nasso (2003) and Mancosu (2009).

[13] Alternatively, set theory can be justified by *formalism* or *pragmatism* about (higher) mathematics (Feferman 2009). On such accounts, however the applicability of mathematics is even more problematic (see for example Horsten 2007, Pincock n.d.). For this reason we will discard these alternatives in this dissertation, however, the topic of alternatives to Platonism about set theory and their relation to the applicability problems evidently merits further discussion.

(such as Euclidean geometry), but this undermines the foundational role of set theory.

3. *Arbitrary (sub)sets.* For each set, any arbitrary combination of its elements is supposed to exist as a well-determined set in its own right. As pointed out by Feferman (1998, p. 45), this should *not* be taken to mean that *only definable* subsets of a set are assumed to exist. Moreover, the *axiom of separation* (for every set A and every given property – given by a formula ϕ of the first-order language of set theory -, there is a set containing exactly the elements of A that have that property; thus, separation is not a single axiom but an axiom schema, that is, an infinite list of axioms, one for each formula ϕ; Bagaria 2001, section 2.1), which can be used to define subsets of a given set, allows for *impredicative definition* in case the formula ϕ contains unbounded quantifiers ranging over the whole universe of sets, i.e. to define an object by reference to a totality which includes the object to be defined (Crosilla 2009, section 1.3.1).[14]

4. *Power sets.* For each set S, the totality $\wp(S)$ of arbitrary subsets of S is supposed to exist as another set, called the *power set*. The power set of an infinite set can never by explicitly constructed: in the case of an infinite set we have no idea of what counts as an arbitrary subset of it. As a consequence, there seems to be no way of generating all the subsets of an infinite set, and so we have no way to form the set of all of them (Crosilla 2009, section 1.3). Moreover, by *Cantor's theorem*, for any set S the cardinality of $\wp(S)$ is strictly greater than that of S, and thus formation of power sets leads to higher and higher infinities. In addition, according to Feferman (1998, p. 45), the existence of power sets *justifies impredicative definition*, for a subset S' of a set S can be singled out by reference to the power set (for example, as the intersection of all subsets X of S satisfying a given condition).

5. *The axiom of choice.* For every set A of pairwise-disjoint non-empty sets, there exists a set that contains exactly one element from each set in A. The objection against this axiom is that it asserts existence of an object without

[14] We will discuss impredicative definitions in a bit more detail in relation to *predicativism* in section 3.2.

providing any means to construct it uniquely (Feferman 1998, p. 40). It has rather unintuitive consequences, such as the *Banach-Tarski paradox*, which says that the unit ball can be partitioned into finitely-many pieces, which can then be rearranged to form two unit balls (Bagaria 2001, section 2.1). The axiom of choice is equivalent to the *well-ordering principle*, which asserts that every set can be well-ordered, i.e. it can be linearly ordered so that every non-empty subset has a minimal element (Bagaria 2001, section 2.1).

6. *Relations and functions as sets.* A relation R between elements of a set S_1 and those of a set S_2 is in set theory simply an arbitrary set of pars in S_1 x S_2. Similarly, functions from S_1 to S_2 are many-one relations. This reduction goes far beyond the mathematician's experience; in particular, the reduction of functions to sets goes far beyond the previous conception of functions as given by laws (Feferman 1998, p. 45).

7. *Objectivity of truth and classical logic.* The meaning of statements is given by their truth conditions, that truth is independent of our means of knowing it. A statement is true or false independently of our ability to verify it, because there is supposed to be a certain *fact of the matter* (Fletcher 2007). The *Law of the Excluded Middle* (LEM) is accepted for every proposition P. As a consequence, *proof by contradiction* is an accepted method. Coupled with LEM, the usual laws of quantification yield the possibility to establish an existence result $\exists x P(x)$ from a contradiction derived from $\forall x \ \neg P(x)$, without any indication as to how to produce an element x to satisfy $P(x)$ (Feferman 1998, p. 45).

8. *Weird algebraic properties for infinite sets.* The idea behind transfinite numbers is that they allow one do arithmetic with infinite numbers in a way that is analogous to the arithmetic on finite numbers (see section 2.1 above). However, Cantor's theory has weird algebraic properties. For example, if a and b are cardinal numbers, then (Benci & Di Nasso 2003, Benci et al. 2006):

$$a + b = a \times b = max(a, b) \tag{2.1}$$

whenever a is infinite and $b \neq 0$. That is, when counting cardinal numbers one may add elements to an infinite set without making the set bigger. Furthermore, if one adds an element at the bottom of an infinite well-ordered set, its ordinal number does not change, while its ordinal number becomes bigger if the element is added at the top (Benci & Di Nasso 2003):

$$1 + \omega = \omega < \omega + 1 \tag{2.2}$$

9. *Part-whole principle not respected.* The set theoretical conceptions of size and number do not satisfy the intuitive principle that the whole is always bigger than a proper part of it (see section 2.1 above).

Note that these differing views concerning the nature of mathematics and the objects with which it deals also influence the conception of what is a *proof* in mathematics. This issue rises, for example, with regard to Cantor's diagonal argument: if one does not accept completed infinity, then the diagonal argument merely shows that given any collection of real numbers, a real number which is not yet in the collection can be constructed (Potter 2004, p. 138, cfr. Feferman 1998, p. 46). That is, if one does not accept completed infinity, then the diagonal argument does *not* show that there are *more* real numbers than natural numbers.[15]

Following up on the above objections to set-theory, alternative schemes for the foundations of mathematics have been developed, with the aim of demonstrating that everyday mathematics can be accounted for in a direct and straightforward way on philosophically acceptable, non-Cantorian grounds. The results that have been obtained in this direction led Feferman to conjecture that

> "a case can be made that *higher set theory is dispensable in scientifically applicable mathematics*, that is, in that part of everyday mathematics which finds its applications in other sciences. Put in other terms: the *actual infinite is not required for the mathematics of the physical world*" (Feferman 1998, p. 30, emphasis in the original).

[15] Note furthermore that Cantor's diagonal argument is a *reductio ad absurdum* and that therefore the acceptance of the conclusion relies on the *Law of the Excluded Middle*. There is also another proof to the effect that the reals are of a higher-order infinity than the rationals (which is not a *reductio*), namely Borel's proof to the effect that any countable subset of reals is *infinitesimal* with respect to all reals (Chaitin 2006, section 2.2). However, Borel also proved that most reals are inaccessible to us, since they can never be picked out as individuals using any conceivable mathematical tool. In Borel's view all these reals, and thus *most* reals, are "mathematical fantasies" (Chaitin 2006, section 2.5).

2.3 The potential infinite: Constructive mathematics

In the previous section we have seen that even though set theory is accepted by the majority of the mathematical community and permeates the whole of present-day mathematics, this does not mean that this *status quo* goes undisputed. Some mathematicians take concepts such as Cantor's transfinite numbers, axioms such as the axiom of choice, and paradoxes such as the Banach-Tarski paradox as signs that the discipline has strayed too far from the realm of ideas accessible to the human mind (Maloney 2008). Constructivism can be seen as a conservative attitude towards this issue, aiming to build and to keep mathematics within the realm of what is accessible to the human mind.[16] According to constructivists it is epistemologically problematic or meaningless to talk of some particular mathematical object if that object is not *constructed*. In classical mathematics such talk is allowed, for it's logical principles – i.e. classical logic – permit to conclude that a certain object exists if the assumption of its non-existence leads to a contradiction. Constructivism, in other words, rejects the classical mathematical practice of taking mathematical objects to exists independently of our ability to find or construct them.[17]

A particular notion that is considered problematic is set theory's notion of infinity as *actual infinity*.[18] In fact, constructive mathematics can be seen as the project of showing that actual infinity is not indispensable for mathematics. Constructivism rejects the notion of actual infinity, and adopts *potential infinity* instead. In the constructive view, infinite sets and infinite objects are never 'finished', but rather are perceived as *developing* in the course of time (Waaldijk 2005, p. 24).

To illustrate the difference between actual and potential infinity, we will start by considering some examples of mathematical concepts which change their meaning according to whether infinity is understood actually or potentially. As an actually infinite collection, the natural numbers are understood as the *set* of all

[16] There are several versions of constructivism, and the notion of infinity and its potentiality takes in each of these theories a (slightly) different shape. We will not go into these differences in this section; the interested reader can consult (Bridges & Palmgren 1997).

[17] 'Finding' and 'constructing' an object is for constructivists the same thing (cfr. Schechter 2001).

[18] Some versions of constructivism do accept the natural numbers as an actually infinite collection. In this section, we will be talking about versions of constructivism that do not. We will say a bit more about different versions of constructivism in section 3.2.

natural numbers $N = \{0, 1, 2, 3, \ldots\}$, thought of as given once and for all. As a potential infinity, instead, the natural numbers are understood as a *generating process* $0 \to 1 \to 2 \to 3 \to \ldots$, which can be continued as long as one pleases. Importantly, thought of as a potentially infinite collection, at any stage of the generating process, the collection of natural numbers has only a finite number of elements. Also other mathematical concepts in which a reference is made to the notion of infinity can be understood in different ways. For example, the sum of an infinite series $\sum_{n=1}^{\infty} a_n$, is as an actual infinity understood as the addition of *infinitely many quantities*, and as a potential infinity instead as the *limit of finite sums*, $\sum_{n=1}^{N} a_n$. Similarly, an integral $\int_a^b f(x)dx$, is as an actual infinity, understood as the addition of *infinitely many infinitesimal areas* $f(x)dx$, and as a potential infinity defined in terms of the *supremum and infimum of finite sums of areas*. We could thus say that in case of potential infinity, the reference to infinities is merely apparent, infinities are 'explained away' in finite terms (cfr. Fletcher 2007). Now, in modern analysis infinitesimals and infinite quantities are banished in favor of arbitrarily small positive numbers ϵ and arbitrarily large natural numbers N. Did modern analysis banish actual infinity? No, underlying these uses of potential infinity are two uses of actual infinity: the concept of an *actually infinite set* (e.g. the set of all natural numbers, N, the set of all real numbers, R) and the concept of *infinite quantifiers*, ranging over actually infinite sets ('for all x,...', 'there exists a δ ...'). Thus, in modern analysis – and consequently in the physics that uses it – potential and actual infinity are intertwined (cfr. Fletcher 2007).

The difference between actual and potential infinity is perhaps most clear in the concept of the continuum. Classically conceived, real numbers are (following Cantor) equivalence classes of completed Cauchy sequences of rationals $\{r_n\}$, each of which represents a determinate position on the rational line, and which, taken together, form an uncountable set: the classical continuum.[19] Constructively conceived, real numbers are *uncompleted* Cauchy sequences of rationals $< r_n >$, each of which, at each stage in their determination, represents an *open interval* on the rational line, and which, taken together, form an *indefinitely refinable coordinate system* within the continuum.[20] Thus, classical real numbers are actually infinite

[19] We will discuss Cantor's conception of the continuum – as well as other conceptions – in more detail in section 3.1.

[20] Note that the criterion for Cauchy convergence, too, differs in the constructive from that

objects: they are infinite Cauchy sequences of rationals $\{r_n\}$ which are assumed to be completely given. Constructive real numbers, instead, are potentially infinite objects: they are Cauchy sequences of rationals $< r_n >$ which can be indefinitely extended. Constructive real numbers are never completely given and are said to *generate* the real number to which they converge. The dimensionless points which constitute the classical continuum are within constructive mathematics conceived as *unattainable theoretical limits* to the process of dissection.

Many theorems of classical real analysis have a constructive counterpart – although sometimes with a strengthened hypothesis or a weakened conclusion (cfr., e.g., Bridges & Dedui 1997). There are, however, classical theorems which do not have a constructive counterpart, that is, theorems which cannot be proved in constructive mathematics – i.e. using intuitionistic logic – and it can be proved that they cannot.[21] Also the converse obtains: constructive theorems which are classically false. Here we will discuss one such example, which is particularly relevant to the case of phase transitions: the constructive theorem to the effect that *any function defined on every real number is continuous.* To be precise, this is not in all systems of constructive mathematics a theorem. It is a theorem in Brouwer's intuitionistic analysis, where it is a consequence of the representation of real numbers as choice sequences. Brouwer also assumes the principle of bar induction, which leads to the stronger theorem that any function on an interval $[a, b]$ is uniformly continuous (see e.g. Brouwer 1927/ 1967, **?**).[22] In Bishop's constructive analysis there is no continuity theorem, but all the obvious ways of defining a discontinuous function fail (see e.g. Beeson 1985).

Why are there no discontinuous functions defined on every real number in constructive mathematics? This may be recognized in the following way. Recall that constructive real numbers are uncompleted Cauchy sequences of rationals $< r_n >$. Suppose that f is a function from $\mathbb{R}$ to $\mathbb{R}$, and take $x =< r_n >$ and

in the classical framework: In classical mathematics, a sequence of rationals $\{r_n\}$ satisfies the Cauchy criterion for convergence if for any positive real number ϵ, *there exist* a natural number $N(\epsilon)$, such that for every $n > N(\epsilon)$ and every $p > 0$, $|a_{n+p} - a_n| < \epsilon$. In constructive mathematics, on the other hand, a sequence $\langle r_n \rangle$ of rationals satisfies the Cauchy criterion for convergence if for every natural number k, *we can effectively construct* a natural number $N(k)$, such that for every $n > N(k)$ and every $p > 0$, $|a_{n+p} - a_n| < \frac{1}{k}$.

[21] See Schechter (2001) for an example.

[22] It must be noted though, that in his Cambridge lectures, Brouwer also considered 'measurable functions': functions that are defined almost everywhere on $\mathbb{R}$, but need not be continuous, and thus could be used as a way of modeling discontinuous phenomena (van Dalen 1981).

$f(x) =< s_n >$. At any particular stage n, we have a finite amount of information about the terms of $< r_n >$ (plus the information that $< r_n >$ is Cauchy), which determines an interval within which x can develop in subsequent stages of its generation – say from $x - \frac{1}{j}$ to $x + \frac{1}{j}$. This implies that likewise, at any stage n we have a finite amount of information about the terms of $< s_n >$, which determines an interval within which $f(x)$ can develop in subsequent stages of its generation – say from $f(x) - \frac{1}{k}$ to $f(x) + \frac{1}{k}$. Thus, $f(x)$ can be computed within some approximation $\frac{1}{k}$ for some positive integer k if x is known within $\frac{1}{j}$ for some positive integer j. This means that f must be continuous. This is not a rigorous argument, but should suffice to make the continuity theorem plausible (a similar argument see e.g. Beeson 1985).

Another approach to see why there are no discontinuous functions defined on every real number in constructive mathematics would be to consider why any attempt to formulate a discontinuous function will fail. Consider some function f defined by the equation:

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Recall that a function is a rule f which enables us, when given a real number x, to compute another real number $f(x)$, in such a way that when $x = y$, then $f(x) = f(y)$. Thus, in order to compute our $f(x)$, we must for any given real number be able to determine whether $x \leq 0$ or $x > 0$. Suppose we have some $x =< r_n >$ of which we know that up to some stage n, all digits of x are 0. At this stage n, we cannot determine whether $f(x)$ is 0 or 1. In fact, no finite amount of computation will guarantee that we will be able to tell whether $x \leq 0$ or $x > 0$, so we have no guarantee that we will ever be able to compute $f(x)$. Thus, the above function f is not a function according to the constructivist, since it does not tell us how to compute it at the point of discontinuity. A similar argument applies to all discontinuous functions (see e.g. Beeson 1985).

It is however possible to obtain discontinuous transitions in constructive mathematics if we do *not* consider functions defined on every real number, but instead allow *partial functions*.[23] To take a simple example:

$$g(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

We can constructively define such an g as a map $g : S \to \mathbb{R}$, where $S = \{x \in \mathbb{R} \mid x < 0 \text{ or } x = 0 \text{ or } x > 0\}$. Now, classically S is all of the reals, but constructively S is *only a subset* of the reals. The function g is thus *partial* in the sense that it is not defined on all of $\mathbb{R}$. That this is the case, follows from the finite amount of information we have about constructive real numbers at each stage of their developments. To see this, consider a constructive real number $x =< r_n >$ of which we know that up to some stage n all digits of x are 0: we have no guarantee that we can ever determine which of $x < 0$, $x = 0$, or $x > 0$ is the case. Thus, an element $x \in S$ is a real number *together with* the information whether x is negative, zero, or positive. Note that $\mathbb{R} \backslash S = \emptyset$, i.e. the complement of S is empty: there is no real number which is neither negative, zero, nor positive. Thus, S is not all of $\mathbb{R}$, but neither is there a real number which is not in S. As Bauer puts it, S should be regarded as $\mathbb{R}$ *with extra information* (personal communication, see also Bauer 2012).

The most well-known – logical – difference between classical and constructive mathematics is without doubt the invalidity in constructive mathematics of the classical *Law of the Excluded Middle* (LEM). The invalidity of LEM in constructive mathematics is a consequence of the view of infinite mathematical objects as ever developing and never completely given: in constructive mathematics it cannot, in general, be determined whether a proposition about them is true or false. For example, for a constructive real number $x =< r_n >$ of which we know that up to some stage n all digits of x are 0, we have no guarantee that we can ever determine whether $x = 0$ or $x > 0$. As a consequence, the classical law of *trichotomy* fails for constructive real numbers. That is to say, in classical mathematics, but not in constructive mathematics, it is the case that for all real numbers x, y either $x < y$, $x > y$, or $x = y$. It has been argued in the literature that in virtue of this failure of trichotomy, constructivism fits well with the practice of measurement in physics, in

[23] I thank Andrej Bauer for making me aware of this.

which we always have some measurement error (e.g. Bauer 2012, Dummett 2000). We will discuss this in section 3.3.

It is important to note that the invalidity of LEM in constructive mathematics does *not* imply that there is no use for *contradiction* in constructive mathematics. As was nicely clarified by Bauer, there are *two kinds* of proof which rely on the derivation of a contradiction, and only one of them is constructively invalid (Bauer n.d.). The first kind of proof goes as follows: to prove p, assume $\neg p$ and derive a contradiction. This principle of reasoning can be written as $\neg\neg p \Rightarrow p$, and as such it is evident that this principle is justified by – and, in fact, is equivalent to – LEM and is consequently constructively invalid. This is the kind of proof which can be rightly called *proof by contradiction.* The other kind of proof which relies on the derivation of a contradiction is just a *proof of negation.* The reasoning of such a proof goes as follows: to prove $\neg p$, assume p and derive a contradiction.[24] This is just the inference rule to prove a negation and is valid both classically and constructively. As Bauer writes, proofs by contradiction can often be avoided, proofs of negation cannot, and it is important to distinguish between the two. In particular, with regards to a constructive perspective on physics, it should be clear that there is no general reason why non-existence proofs – which play in physics an important role – cannot be proved constructively.

Note further that given the invalidity of LEM in constructive mathematics, double negation elimination is invalid and thus a proposition P and its double negation $\neg\neg P$ are *not* equivalent (cfr. Bauer 2012). Whereas P holds if there is evidence supporting it, $\neg\neg P$ holds if there is no evidence that there is no evidence of P, i.e. $\neg\neg P$ holds when P cannot be falsified. Therefore, when we have $\neg\neg P$, we can say that P is *potentially true.* As Bauer explains, there is a translation of classical logic into intuitionistic logic, called *the double negation translation*, which transforms a given proposition by placing double negations in front of every quantifier and logical connective. In terms of our terminology, it inserts the adverb "potentially" everywhere, so the intuitionistic mathematician can interpret the utterances of his classical colleague as statements about potential truth. Unfortunately, there is no such translation for the opposite direction (Bauer 2012).

[24] Note that a proof by contradiction can only be obtained from a proof of negation by plugging in $\neg p$ in place of p, if double negation elimination – or LEM – is assumed. Hence this trick works only classically.

In sum, constructive mathematics differs from classical mathematics before all else in its conception of infinity. Whereas in classical mathematics an infinite collection is thought of as given once and for all, in constructive mathematics a collection is infinite if it can be indefinitely extended, and hence the notion refers to arbitrarily large but finite quantities. As a consequence, constructive real numbers are Cauchy sequences of rationals $< r_n >$ which can be indefinitely extended and are never completely given. At each stage of their generation, constructive real numbers have an extension. An important consequence of the constructive conception of the continuum is that all real-valued functions are continuous.

For a long time – basically, since the appearance of Brouwer's dissertation in which he first proposed intuitionism (a version of constructivism) as a foundational framework for mathematics, until 1967 – it was thought that constructive mathematics was *much too weak* to be useful in the sciences (Bridges & Dedui 1997). In 1967, however, this situation changed abruptly thanks to the appearance of Errett Bishop's *Foundations of Constructive Analysis.* Bishop developed, more or less from scratch and without commitment to the (quasi-)metaphysical principles which underlay the work of Brouwer, large parts of modern analysis by rigorously constructive methods. Why, then, isn't constructivism more popular than it in fact is? I think we can agree with Dummett (1977) that, at least for an important part, constructivism is unpopular because *from purely mathematical considerations* it is pointless. The advantage of constructive over classical mathematics regards primarily *philosophical considerations.* Since the problems surrounding infinite idealizations which are discussed in this thesis are indeed philosophical rather than technical, constructive mathematics might have relevant benefits over classical mathematics in this context. We will discuss applicability of constructive mathematics in physics with regard to case study 1, statistical-mechanical theories of phase transitions, in chapter 4.

2.4 The infinitely small: Nonstandard analysis

In early stages of calculus, up to the first years of the nineteenth century, infinitesimal quantities were widely used to develop many of the classic results of analysis. Traditionally, an *infinitesimal quantity* is one which, while not necessarily coinciding with zero, is in some sense smaller than any finite quantity. An *infinitesimal number* is one which, while not coinciding with zero, is in some sense smaller than any finite number. This sense has often been taken to be the failure to satisfy the *Principle of Archimedes*, which amounts to saying that an infinitesimal number is one that, no matter how many times it is added to itself, the result remains less than any finite number (Bell 2005).

Infinitesimals have been very useful in practice, but their apparently inconsistent character has been bothering mathematicians ever since their earliest appearance in the mathematics of the Greek philosopher Demokritos c. 450 BCE (Bell 2005). In Euclidean mathematics they were banished, only to reappear in the sixteenth century (idem.). Famously, Berkeley in the 18th century called them "ghosts of departed quantities", in the 19th century they were execrated by Cantor as "cholera-bacilli" infecting mathematics, and in the 20th Russell condemned them as "unnecessary, erroneous, and self-contradictory" (Bell 2005). Finally, infinitesimals were supplanted in the foundations of analysis when the rigorous concept of *limit* made their appearance superfluous. With the work of Bolzano and Weierstrass in the 19^{th} century, in modern analysis infinitesimals and infinite quantities are banished in favor of arbitrarily small positive numbers ϵ and arbitrarily large natural numbers N.[25]

This situation changed in 1960 with the work of the mathematical logician Abraham Robinson, which used the concepts and methods of the then recently developed branch of mathematical logic called *model theory* to provide a suitable framework for the development of the differential and integral calculus by means of infinitely small and infinitely large numbers (see Robinson 1996, p. xiii). A basic fact in model theory is that every infinite mathematical structure has nonstandard models, i.e. non-isomorphic structures which satisfy the same elementary properties. The existence of nonstandard models was first shown by Thoralf Skolem in the late twenties, which first, in the fifties, led to an intensive study of nonstandard models of arithmetic. Robinson had the idea of systematically applying that model-theoretic machinery to analysis. By considering nonstandard extensions of the real number system, he was able to provide the use of infinitesimal numbers

[25] Physicists, however, never abandoned the use of infinitesimals as a heuristic device for the derivation of correct results in the application of the calculus to physical problems. Thus, the proscription of infinitesimals did not succeed in extirpating them; they were, rather, driven further underground (Bell 2005).

with rigorous foundations, thus giving a solution to a century-old problem (Di Nasso 1999).

Robinson's *nonstandard analysis* (NSA) is an extension of mathematical analysis embracing both *infinitely large* and *infinitesimal* numbers in which the usual laws of the arithmetic of real numbers continue to hold. By an infinitely large number is meant one which exceeds every positive integer; the reciprocal of any one of these is infinitesimal in the sense that, while being nonzero, it is smaller than every positive fraction $\frac{1}{n}$. There are a number of ways of presenting nonstandard analysis. Here, I will follow Bell (2005) in presenting a sketch of one of them.

The nonstandard universe $*\mathcal{U} = (*U, * \in)$ is an extension of the standard set-theoretical universe $\mathcal{U} = (U, \in)$, i.e. a set U containing the classical real line $\mathbb{R}$ which is closed under the usual set-theoretic operations of union, power set, Cartesian products and subsets, where $\in$ is the usual membership relation on U. $\mathcal{L}(U)$ is the extension of the first-order language of set theory to include a name $\mathbf{u}$ for each element u of U. Using the well-known *compactness theorem* for first-order logic, $\mathcal{U}$ is extended to $*\mathcal{U}$, called a nonstandard universe, satisfying the following key principle:

> **Saturation Principle**. Let Φ be a collection of $\mathcal{L}(U)$-formulas with exactly one free variable. If Φ is finitely satisfiable in $\mathcal{U}$, that is, if for any finite subset Φ' of Φ there is an element of U which satisfies all the formulas of Φ' in $\mathcal{U}$, then there is an element of $*U$ which satisfies all the formulas of Φ in $*\mathcal{U}$.

The saturation property, as Bell explains, expresses the intuitive idea that the nonstandard universe is *very rich* in comparison to the standard one. While in the standard universe $\mathcal{U}$ there may exist, for each finite subcollection F of a given collection of properties P, an element of U satisfying the members of F, there may not necessarily be an element of U satisfying *all* the members of P. In the nonstandard universe $*\mathcal{U}$, to the contrary, the saturation principle guarantees the existence of an element of $*U$ which satisfies, in $*\mathcal{U}$, all the members of P. For example, suppose the set $\mathbb{N}$ of natural numbers is a member of U; for each $n \in \mathbb{N}$ let $Pn(x)$ be the property $x \in \mathbb{N}$ & $n < x$. Then clearly, while each finite subcollection of the collection $P = Pn : n \in \mathbb{N}$ is satisfiable in $\mathcal{U}$, the whole collection is not. In $*\mathcal{U}$, to the contrary, the whole collection P is satisfiable; an element of *U satisfying P in $*\mathcal{U}$ will then be a "natural number" greater than every member of $\mathbb{N}$, that is, an infinite number.

From the saturation property it follows that $*\mathcal{U}$ satisfies the important

> **Transfer Principle**. If σ is any sentence of $\mathcal{L}(U)$, then σ holds in $\mathcal{U}$ if and only if it holds in $*\mathcal{U}$.

The transfer principle, as Bell writes, asserts that all first-order properties are preserved in the passage to or "transfer" from the standard to the nonstandard universe.

The members of U are called *standard sets* or *standard objects*; those in $*U - U$ *nonstandard sets* or *nonstandard objects*. Thus, $*U$ consists of both standard and nonstandard objects; the members of $*U$ will also be referred to as **-sets* or **-objects*. Since $U \subseteq *U$, under this convention every set (object) is also a *-set (object). The *-members of a *-set A are the *-objects x for which $x* \in A$.

If A is a standard set, we may consider the collection Â, called the *inflate* of A, consisting of all the *-members of A (this is not necessarily a set nor even a *-set). The inflate Â of a standard set A may be regarded as the same set A viewed from a nonstandard point of view. While clearly $A \subseteq$ Â, Â may contain nonstandard elements not in A. It can in fact be shown that *infinite* standard sets always get "inflated" in this way. Using the transfer principle, any function f between standard sets automatically extends to a function, also written f, between their inflates. Thus, if $\mathcal{A} = (A, R, \ldots)$ is a mathematical structure, we may consider the structure $\hat{\mathcal{A}} = (\hat{A}, \hat{R})$. From the transfer principle it follows that $\mathcal{A}$ and $\hat{\mathcal{A}}$ have precisely the same first-order properties.

Now suppose that we have a set U of which the set $\mathbb{N}$ of natural numbers is a member. Then also the set $\mathbb{R}$ of real numbers is a member of U, since each real number may be identified with a set of natural numbers. $\mathbb{R}$ may be regarded as an ordered field, and the same is therefore true of its inflate $\hat{\mathbb{R}}$, since the latter has precisely the same first-order properties as $\mathbb{R}$. $\hat{\mathbb{R}}$ is called the *hyperreal line*, and its members *hyperreals*.

A standard hyperreal is just a real, for emphasis we shall refer to it as a standard real. Since $\mathbb{R}$ is infinite, nonstandard hyperreals must exist. The saturation principle implies that there must be an infinite (nonstandard) hyperreal, that is, a hyperreal a such that $a > n$ for every $n \in \mathbb{N}$.[26] In that case its reciprocal $\frac{1}{a}$ is infinitesimal in the sense of exceeding 0 and yet being smaller than $\frac{1}{n+1}$ for every $n \in \mathbb{N}$. In general, we call a hyperreal an *infinitesimal* if its absolute value $|a|$ is less than $\frac{1}{n+1}$ for every $n \in \mathbb{N}$. In that case the set I of infinitesimals contains not just 0 but a substantial number (in fact, infinitely many) other elements. Clearly I is an additive subgroup of $\mathbb{R}$, that is, if $a, b \in I$, then $a - b \in I$.

The members of the inflate $\hat{\mathbb{N}}$ of $\mathbb{N}$ are called hypernatural numbers. As for the hyperreals, it can be shown that $\hat{\mathbb{N}}$ also contains nonstandard elements which must exceed every member of $\mathbb{N}$; these are called infinite hypernatural numbers.

For hyperreals a, b we define $a \approx b$ and say that a and b are *infinitesimally close* if $a - b \in I$. This is an equivalence relation on the hyperreal line: for each hyperreal a we write $\mu(a)$ for the equivalence class of a under this relation and call it the *monad* of a. The monad of a hyperreal a thus consists of all the hyperreals that are infinitesimally close to a: it may be thought of as a small cloud centred at a.

A hyperreal a is *finite* if it is not infinite; this means that $|a| < n$ for some $n \in \mathbb{N}$. Finiteness is equivalent to the condition of *near-standardness*; a hyperreal a is near-standard if $a \approx r$ for some standard real r. In other words, every finite hyperreal lies infinitely close to a standard real.

It is useful to think of the notions of monad and near standardness by thinking of different perspectives: a near standard hyperreal a, $a \approx r$ for some standard real r, as well as all the other hyperreals different from r in the monad of r, may be considered equal to r *from a standard perspective but not from a nonstandard perspective*. Keisler introduced the concept of an "infinitesimal microscope" to capture this notion: through an infinitesimal microscope we may see that there is some infinitesimal difference between a and r, but without the microscope this is invisible (Keisler 1986, cfr. Keisler 2007). Keisler illustrated the hyperreal line with its monads as if seen through an "infinitesimal microscope" as in figure 2.1:

[26] As Bell adds, it follows that $\hat{\mathbb{R}}$ is a nonarchimedean ordered field. One might question whether this is compatible with the facts that $\hat{\mathbb{R}}$ and $\mathbb{R}$ share the same first-order properties, since the latter is archimedean. In fact it is consistent, because the archimedean property is not first-order. However, while $\hat{\mathbb{R}}$ is nonarchimedean, it is *-archimedean in the sense that, for any $a \in \hat{\mathbb{R}}$ there is $n \in \hat{\mathbb{N}}$ for which $a < n$.

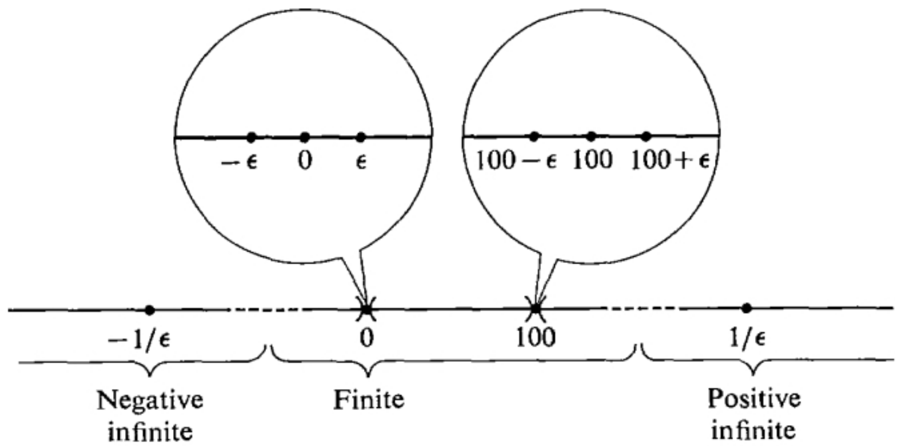

Figure 2.1: Hyperreal line (Keisler 1986, p. 25)

When translating results obtained about the finite nonstandard reals to propositions about the standard reals, the nonstandard hyperreals can be discarded with by means of the *standard part function*, *st*. The standard part function is a function from the finite hyperreal numbers to the real numbers, which "rounds off" a finite hyperreal to the nearest standard real. The function *st* is represented in figure 2.2 by a vertical projection. Again, an "infinitesimal microscope" is used to view an infinitesimal neighborhood of a standard real number r; α, β and γ represent typical infinitesimals (Bascelli *et al.*, 2014).

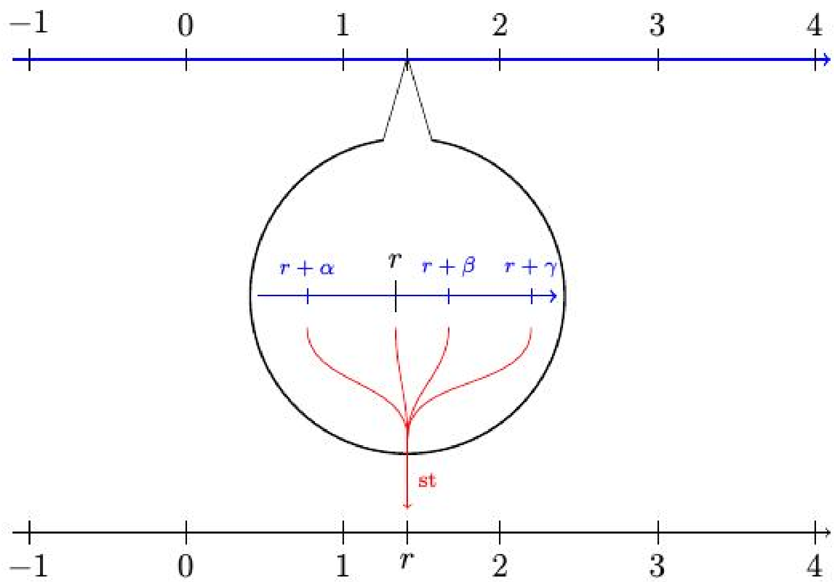

Figure 2.2: The standard part function, *st*. From Wikipedia, after Bascelli *et al.*, 2014.

Much of the usefulness of nonstandard analysis stems from the fact that statements of classical analysis involving limits or the (ϵ, δ) criterion admit succinct, intuitive translations into statements involving infinitesimals or infinite numbers, in turn enabling comparatively straightforward proofs to be given of classical theorems (Bell 2005). However, the hyperreals are also interesting from the perspective of modeling practice, given that they have different properties from the standard reals. For example, since infinitesimals fail the Archimedean principle, the hyperreals are not *complete*,[27] and can serve to build discrete models of physical systems. We will discuss applicability of nonstandard analysis in physics with regard to case study 2, indeterminism in Newtonian mechanics ("Norton's Dome"), in chapter 5.

[27] We will discuss completeness in section 3.1.

2.5 Idealization in mathematics

So far we have discussed classical mathematics (set theory), constructive mathematics, and nonstandard analysis. The reader might have gotten the impression that the two alternatives to classical mathematics that we discussed, i.e. constructivism and nonstandard analysis, are each other's opposites: whereas constructivism is very careful not to introduce nonconstructable elements in the theory, nonstandard analysis happily extends classical mathematics, by means of including infinitely large and infinitely small numbers at the fundamental level. This is in fact a common view about the relation between these three mathematical systems. We find this view for example expressed by Bishop:

> "[Constructive mathematics and nonstandard analysis] are at opposite poles. Constructivism is an attempt to deepen the meaning of mathematics; nonstandard analysis, an attempt to dilute it further" (Bishop 1972, quoted in Sanders 2017, p. 19).

Meaning, for Bishop, means *computational content*. That is, according to Bishop, mathematics is meaningful when every statement has numerical meaning and every object has an algorithmic description, i.e. it has been built using (and can in principle be reduced to) *algorithmic reasoning* (Sanders 2017, p. 17). Nonstandard analysis is *not* meaningful in Bishop's view, because, as he sees it, the presence of *ideal objects* (in particular infinitesimals) in nonstandard analysis yields the *absence* of computational content (Sanders 2017, Sanders 2018).

Sanders shows that this is not completely true: there *is* substantial computational content in nonstandard analysis, for working with infinitesimals comes down to making explicit calculations. As he puts it, infinitesimals actually provide an elegant shorthand for expressing computational content (Sanders 2017, p. 24). Sanders thus challenges the binary view that mathematics is either constructive or not. In his view, rather than being altogether non-constructive, classical nonstandard analysis inhabits the twilight zone between the constructive and the non-constructive: it is not the former as it (explicitly) includes the law of the excluded middle, but it also has 'too many' constructive properties to be dismissed as merely the latter (Sanders 2018, p. 49).[28]

[28] It should be kept in mind, however, that nonstandard analysis – at least, in its usual development – involves quite *non-constructive* axioms. The usual development of Robinson's

So, how can we understand Bishop's view that nonstandard analysis amounts to the delution of meaning?[29] Why are infinitesimals and infinite numbers suspect? To better understand the different views on mathematical ontology, and in particular why, according to some, ideal elements such as infinite numbers and infinitesimals are suspect, we will discuss these views in relation to the notion of *idealization* in mathematics.

It is quite common in mathematics to speak about "ideal" (or "imaginary", or even "impossible") elements in contradistinction to "real" elements. This linguistic practice can be explained by means of some kind of distinction, for example of a metaphysical, epistemological, or psychological nature (cfr. Cantù 2013). Metaphysically, it has been common to distinguish between *formal* and *real* elements, where the latter as opposed to the former have a counterpart in the physical world. For example, one might distinguish between a line segment, which is a representation of something real, and an infinite straight line, which has no counterpart in the physical world (Cantù 2013). From an epistemological perspective the distinction might be traced between different *grades of knowledge*: higher or lower certainty, greater or smaller perspicuity, higher or lower difficulty, or greater or smaller efficiency of proofs. For example, one might distinguish between theories that can be proved to be consistent and theories that cannot, or between finitary and infinitary theories (Cantù 2013). From a psychological perspective there is a distinction between elements that can be *intuitively conceived*, and elements that cannot be intuited or represented. For example, the concept of unit and the process of iteration are often conceived as intuitively conceived. According to Bishop "[t]he positive integers and their arithmetic are presupposed by the very nature of our intelligence and, we are tempted to believe, by the very nature of intelligence in general"; therefore, in his view, "[t]he development of the positive integers from the primitive concept of the unit, the concept of adjoining a unit, and the process of mathematical induction carries complete conviction" (Bishop

nonstandard analysis proceeds via the construction of a nonstandard model using a *free ultrafilter*, the existence of which is only slightly weaker than the axiom of choice in ZFC (Sanders 2017, p. 20, n. 20; cfr. section 2.2 in this thesis). Alternatively, there are constructive developments of NSA which do not make use of free ultrafilters (e.g. Palmgren 1998, see also the references in Sanders 2017, section 3.3).

[29] To be clear, I will *not* be concerned with exegesis of Bishop's writings; rather, I will consider how this idea fits with the constructive conception of mathematics as we have discussed in this thesis (see section 2.3.).

1967, quoted in Bridges & Palmgren 1997, section 3.1).

Where exactly the border between real and ideal elements is located, differs between mathematical frameworks. For, as we have seen above, in (classical) nonstandard analysis, the nonstandard objects – i.e. the infinitesimal and infinitely large numbers – are typically conceived as ideal, in contradistinction to the real numbers which are conceived as just that: real. In constructive mathematics, to the contrary – as we have seen in section 2.3 – the classical real numbers are conceived as *ideal* limits of the process of dissection. Thus, to answer the above question how we can understand the view that nonstandard analysis is a delution of meaning, we note that from the standard constructive perspective classical nonstandard analysis can be seen as *double* idealization: already the classical real numbers (limits of infinite sequences of reals) are ideal objects, and the nonstandard elements (infinite numbers and infinitesimals) are another step away from what is accessible to the human mind.[30]

Not only do classical and constructive mathematicians place the border between real and ideal elements differently, the distinction also has a different *justification*. As we discussed in section 2.3, constructivism can be seen as a reaction to the abstract nature of classical mathematics (set theory) and its aim is to build and keep mathematics within the realm of what is accessible to the human mind. For this reason, restrictions are placed both on the objects studied and on the methods of proof which may be applied: only those objects which are epistemologically accessible are accepted, and the methods admitted guarantee that this property is preserved when new objects are created from existing ones. The real-ideal distinction in constructive mathematics is thus (primarily) epistemologically motivated: real-ideal in constructive mathematics – i.e. constructed versus non-constructed – is linked to higher versus lower certainty. This seems to be different in case of (classical) nonstandard analysis, for if one already accepts the infinitary objects and methods of classical mathematics, then the addition of infinite and infinitesimal elements – which are consistent with classical mathematics and use the same methods (classical logic) – does not seem epistemologically problematic. Rather, it is suggested that infinite and infinitely small numbers in nonstandard analysis are considered *ideal* in the sense that they are not intended by initial

[30] Again, I am not arguing that this is Bishop's view - nor the view of any actual constructivist for that matter -, but merely that this is a way in which we can conceive the constructivist's objection against (the ideal objects in) classical nonstandard analysis.

naive intuitions (Reeb 1981, quoted in Fletcher *et al.* 2017, section 8.2). The real-ideal distinction in nonstandard analysis seems thus to have to do with the distinction between the intuitable and the non-intuitable and can be said to be *psychological* in character (cfr. Cantù 2013).[31]

It is important to note that the introduction of ideal objects – even when they are regarded as epistemologically suspicious – is (usually) *not* considered to be something in itself 'wrong' or 'bad'. Bishop, for example, saw it this way:

> "...idealistic mathematics is [not] worthless from the constructive point of view. This would be as silly as contending that unrigorous mathematics is worthless from the classical point of view. Every theorem proved with idealistic methods presents a challenge: to find a constructive proof" (Bishop 1967, quoted in Sanders 2017).

Thus, in Bishop's view classical – ideal – mathematics is *not worthless*, but rather presents the challenge for the constructivist of finding a constructive proof and thus giving it an epistemologically secure basis. Note that the constructivist does have the means to talk about ideal mathematicals in his theory: as we have seen in section 2.3, on the intuitionistic understanding of truth, $\neg\neg P$ holds if P cannot be falsified. Thus, of classical truths which in constructive mathematics cannot be proved nor disproved, a constructivist can say that they are *potentially true* (Bauer 2012, section 1). Similarly, those objects of which the existence cannot be proved constructively (i.e. objects which cannot be constructed), but neither can be disproved, can be said to *exist potentially* in constructive mathematics.

In virtue of the use of intuitionistic logic there are thus *two levels* of existence in constructive mathematics. Something similar holds for nonstandard analysis, where there are *two universes*: the standard universe (which contains only the standard reals) and the nonstandard universe (which contains besides the standard reals also the nonstandard reals: infinitesimals and infinitely large numbers; see section 2.4). This *layered ontology*, with a accompanying distinction between real and ideal objects, is something that constructive mathematics and nonstandard analysis have in common. This in contradistinction to set theory, which has a very *flat ontology*. The flat ontology in set theory is a consequence of the use of

[31] This applies, of course, *only* to the use of infinitesimals after Robinson. For before Robinson with his standard analysis gave a mathematically rigorous treatment of infinitesimals, the worry was indeed about their consistency and thus epistemological (see section 2.4).

the law of the excluded middle – which permits to conclude that a certain object exists if the assumption of its non-existence leads to a contradiction – and of the fact that all elements – in particular the finite and the infinite – are treated on the same footing (see section 2.2).

In sum, even though constructive mathematics and nonstandard analysis *in some sense* are each others opposites – namely, in that the former has a restricted and the latter an extended domain with respect to set theory – *in another sense* they are more similar to each other than to set theory. Namely, whereas set theory has a *flat* ontology, both constructive mathematics and nonstandard analysis have a *layered* ontology with a distinction between real and ideal elements.[32]

[32] I thank Sam Sanders (personal communication) for suggesting me to think about the distinction between set theory, constructive mathematics, and nonstandard analysis in this way.

Chapter 3

The continuum

> *The idea of never reaching the area of the circle, no matter how far one might go in the sequence of polygons, although one approaches it arbitrarily closely, strains the power of imagination to such a degree that it will tend, at all cost, to bridge this gap extending, as it were, between reality and the ideal. Under this psychological pressure the – infinitely small or infinitely large? – step is taken that leads to the assertion: the circle is a polygon with infinitely many infinitely small sides.*
> - W.G. Hankel (1874)[1]

A prominent example of a mathematical entity that is considered to be indispensable to science is the real number system or continuum. In the physical sciences, many things – entities, such as electrons and planets; constraint surfaces, such as table tops and domes; and dynamical variables, such as forces, positions, and velocities – are mathematically represented by functions defined on $\mathbb{R}^n$, the n-dimensional space of real numbers (Stemeroff & Dyer 2016). Why do we model physical quantities using $\mathbb{R}$? "One main, if not compelling, reason for taking [physical] quantities to have real-number values is that results of measuring them can apparently always be reduced to the position of some sort of pointer in space and space is modelled using $\mathbb{R}$" (Isham & Butterfield 2000).

That space is modelled using $\mathbb{R}$ is no coincidence. For the mathematical continuum – insofar as it is right to speak about *the* mathematical continuum,

[1] Quoted in Weyl (1949, p. 43).

see section 3.1 below – is developed precisely to reflect our intuitions about space. According to our intuitive conception, space is infinite not only in the sense that it never comes to an end; at every place it is also, so to say, *inwardly infinite*: it is capable of infinite division. Points in space can only be fixed step-by-step by a process of subdivision which progresses *ad infinitum* (cfr. Weyl 1949, p. 41).

The history of the mathematical continuum goes back to the pure geometry of the ancient Greeks. For them, real numbers are given as the ratio of two given segments, and thus it is up to geometry to tell us what numbers exist (Weyl 1949, pp. 38-39). The discovery of the irrationality of the ratio $\sqrt{2}$ of the diagonal and side of a square made it clear that the fractions are not the only possible quantities measuring ratios of line segments, and thus not the only "real numbers" (ibid). These "gaps" on the continuous number line which are not filled by fractions (rationals) are the *irrational numbers*. Applying the idea of existence to all the points in space, the Greek geometers recognized that the continuum must be "completed" with irrational numbers in order to make it "gapless". Thus, in this manner the continuum is supposed to reflect our intuition of the *completeness* of space: just as space, according to our intuition, is complete in the sense that it has no gaps which do not correspond to a physical point, the mathematical continuum or real line is complete in the sense that it has no gaps which are not covered by a mathematical point or real number.

Only in the 19^{th} century did mathematics go beyond the ancient Greeks (Weyl 1949, p. 39). At that time it was recognized that in the Euclidean development of geometry completeness was *assumed*: several of its proofs assume the existence of certain points that are supposed to be there, but their existence is not guaranteed by the Euclidean postulates. For example, in Euclid it is assumed that a continuous curve joining the center of a circle to a point outside of it meets the circumference (Weyl 1949, p. 40). However, the curve and the circle might be "gappy" – i.e. *incomplete* – and thus they might fail to have a point in common.

In response, several improved – set-theoretic (Cantor, Dedekind) and geometric (Hilbert) – characterizations of the continuum were developed. In particular, when it comes to completeness, there are several different methods of filling the "gaps" in the rational line, leading to various and under certain conditions equivalent versions of the completeness axiom, and thus various versions of the (classical) continuum. In additon, a minority of mathematicians argued that the classical continuum does *not* reflect our intuitions about space and other physical quantities accurately, and developed alternative, non-classical – e.g. intuitionistic – conceptions of the continuum.

In this chapter we will discuss different conceptions of the continuum. We will start by discussing six classical conceptions of the continuum (section 3.1), and then proceed by discussing three non-classical conceptions (section 3.2). In the last section of this chapter we will consider some arguments against (the necessity of) using the classical continuum for modeling in physics (section 3.3).

3.1 Classical conceptions of the continuum

It is common to refer to *the* continuum, as if it is a uniquely determined concept. But – even if, as we will do in this section, we restrict our attention to classical mathematics – the mathematical continuum takes different forms: in geometry that of the straight line, in analysis that of the real number system (which can be characterized in several different ways), and in set theory as the power set of the natural numbers or as the set of all infinite sequences of zeros and ones (Feferman 2009). It is often assumed that these formulations somehow express the same concept and thus can be identified with one another – in many cases without argument. Here we will discuss several conceptions of the continuum and see to which extent they are instances of the same concept. We will follow Feferman (2009) in discussing six conceptions of the classical continuum. Three non-classical conceptions will be discussed in section 3.2.

The *Euclidean continuum* exists as a part of the framework of plane geometry, and by extension spatial geometry (Feferman 2009, p. 7). The epitome of the continuum in Euclidean geometry is the indefinitely extended straight line. As mentioned in the introduction, the continuum is supposed to reflect our intuitions about space, but, as Feferman explains, in Euclidean geometry these intuitions are Janus faced (Feferman 2009, pp. 7-8). On the one hand, our language leads us to treat points and lines as objects for which the basic relation is that of a point lying on a line, e.g. in such statements as, “For any two distinct points there exists a unique line on which they lie”. On the other hand, our intuitions of points as being dimensionless – in Euclid’s terminology, as having “no parts” – and of lines as having “no breadth”, requires us to imagine entities which have no substance. Alternatively, we can think of points as being "pure locations", but also this makes talk of a specific point lying on a line sound odd. Dispite the fact that our intuitions must do double duty, Euclid's postulates – except, perhaps, the parallel postulate – are sufficiently intuitive and can immediately be recognized as valid. The main thing to be emphasized about the conception of the continuum as it appears in Euclidean geometry is that the concept of set is *not* part of the basic picture. In general, figures in Euclidean geometry are not to be identified with their sets of points (Feferman 2009, p. 8).

Euclid's axioms were put into a corrected and more analytically perspicuous form by Hilbert in his *Grundlagen der Geometry* (1899). The *Hilbertian continuum* is here developed in the fifth axiom group – that he calls the *Axioms of Continuity* – which are meant to make explicit what Euclid tacitly assumes: the completeness of the real line. Dispensing with Hilbert's own formulations, we may summarize the group in a single axiom of completeness as follows (Moore 2007):

> **Completeness**. Let all the points on a line be divided into two non-empty sets S_L and S_R such that every point in S_L lies to the left of every point in S_R. Then S_L has a rightmost point if S_R has no leftmost one.

As with the Euclidean line, the Hilbertian continuum is not conceived of independently, but only within the framework of plane (and spatial) geometry as a whole. In contrast with the Euclidean line, however, set theoretical concepts *are* part of this picture. As Feferman (2009, p. 13) puts it, the Hilbertian conception of the continuum is a hybrid of geometrical and set-theoretical notions.

The notion of completeness in Hilbert's system of geometry is directly informed by the completeness condition earlier developed by Dedekind (*Stetigkeit und Irrationalzahlen*, 1872).

> "I find the essence of [completeness] in [. . .] the following principle: 'If all points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division of all points into two classes, this severing of the straight line into two portions'" (Dedekind 1872, after Moore 2007, p. 72).[2]

[2] Where I inserted "completeness", Dedekind uses the word "continuity". But "continuity" is

In *Dedekind's continuum*, points are conceived of as elements of a linearly ordered set going from "left" to "right". Any two points in that set should determine a line segment, and since there are incommensurable lengths such as the diagonal of a square with unit length, not all these points can correspond to rational numbers (Feferman 2009, pp. 10-11). In fact, Dedekind notes that "there are infinitely many points [on the line] to which no rational number corresponds" (Dedekind 1872, quoted in Bell 2005, section 5). He then goes on to fill-up the gaps in the rational numbers through the creation of new "point-individuals" by means of the method of *cuts*: a *cut* is a partition (A_1, A_2) of the rational numbers, such that every member of A_1 is less than every member of A_2. Each rational number corresponds to a cut, but there are infinitely many cuts that do not correspond to a rational number; each cut that does not correspond to a rational number creates an irrational number. Dedekind thus aims to "fill up" the rational numbers, so that *every* point on the line corresponds to a real number. The domain of all the cuts, and thereby the associated domain of all the real numbers, can be ordered in such a way that it has the completeness property (Bell 2005, section 5).

The *Cantorian continuum*, as we have seen in section 2.3, consists of equivalence classes of Cauchy sequences of rational numbers. Cantor analyzed the continuum in terms of infinite *point sets*, where he took the elements of these sets, i.e. points of a line, to correspond to real numbers. A sequence $a_1, a_2, \ldots, a_n, \ldots$ of rational numbers is a Cauchy sequence – or, in Cantor's terms, a *fundamental sequence* – if for every rational $\epsilon > 0$ there exists an integer N such that whenever $n \geq N$, $|a_{n+m} - a_n| < \epsilon$. Any sequence $< a_n >$ satisfying this criterion has a certain limit, which he denotes by b. Cantor denotes the collection of all such b corresponding to fundamental sequences by B, and defines a total ordering and an arithmetical structure on B. In particular, Cantor defines $b = b'$, where b is the limit of $< a_n >$ and b' the limit of $< a'_n >$, to be the case when the difference between $< a_n >$ and $< a'_n >$ converges to zero. Cantor now goes on to prove that there is a *one-to-one correspondence* between the line and B. That is, he shows that each point on the line corresponds to an element of B, but the converse – "the geometry of the straight line is complete" – he assumes as an axiom:

> "[T]o make the geometry of the straight line complete is only to

put forth as that geometrical property that the rational numbers lack, and should therefore be understood as that which we call "completeness" (cfr. Moore 2007, p. 72).

> add an axiom, which simply consists in [declaring that] any numerical quantity belongs to a certain point of the straight line [...] I call this theorem an axiom because it is in its nature to not generally be provable" (Cantor 1872, quoted after Scoville 2012).

The idea here is that after this bijection is established, we have a rigorous foundation to justify arithmetic operations on the real numbers (Scoville 2012). Thus, the set B is just a set of symbols (refering to limits of Cauchy sequences) on which you can "do" arithmetic, because they are in one-to-one correspondence with (subsets of) the real line. In this manner, Cantor freed the concept of set from its geometric origin, and paved the way for the emergence of the general abstract set central to todays mathematics (Bell 2005, see section 2.2 of this thesis). Point sets soon came to be studied in their own right, and evolved into the mathematical discipline we now call *topology*. Moreover, Cantor's analysis of the continuum and the use of the idea of one-to-one correspondence made it for the first time possibile to compare the *sizes* of point sets in a definite way (Bell 2005, see section 2.1 of this thesis)

The idea behind Cantor's definition of the real numbers as fundamental sequences or Cauchy sequences of rational numbers can be used for the set-theoretic conception of the continuum as *the set of all subsets of natural numbers* (Feferman 2009, section 3.6; cfr. Feferman 1998 p. 35). Namely, we can associate each element x of $\mathbb{R}$ with the subset $\mathbb{Q}_x$ of $\mathbb{Q}$ consisting of all the rationals r with $r < x$, that is, in other words, every real number is associated with the set of all rational numbers smaller than it. Since the rational numbers are dense in the real numbers,[3] every real number x is then uniquely determined by the subset $\mathbb{Q}_x$ of such rational numbers.

Another set-theoretic conception of the continuum is that of *the set of paths in the full binary tree* (Feferman 2009, section 3.5). A full binary tree T is obtained from an initial node by successive branching to the left or right, such as in figure 3.1:

[3] That the rational numbers are *dense* in the reals means that if a and b are real numbers with $a < b$, then there is a rational number $\frac{p}{q}$ such that $a < \frac{p}{q} < b$. For a proof, see e.g. Trench (2013, theorem 1.1.6). Note that this implies that whereas there is an *uncountable* number of "gaps" in the real numbers (i.e. the irrational numbers), between *any two* of these gaps there is a rational number; but at the same time the rationals are countable, whereas the irrationals are uncountable and thus there are *many more* irrationals than rationals. This is in my opinion one of the most counter-intuitive consequences of set theory.

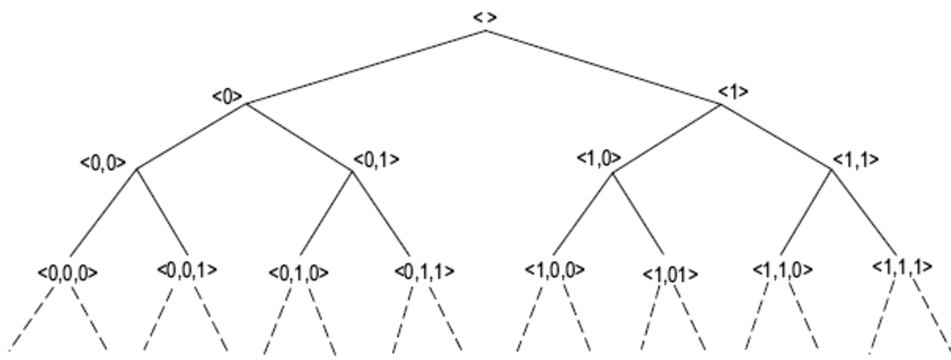

Figure 3.1: Full binary tree.

Every element or node of the tree has either zero or two successors, typically called "left" and "right". Denoting "left" by 0 and "right" by 1, we recognize this as equivalent to the generation of *all finite sequences of zeros and ones*, beginning with the empty sequence. A *path* in the full binary tree is a subset of T which contains the initial node and contains with each node exactly one of its successors. A path is thus represented by an *infinite sequence of zeros and ones*, and every such sequence represents a unique path. One standard set-theoretical conception of the continuum is as the set of *all* paths in T (or equivalently, as the set 2^N of all functions from N into the set $\{0, 1\}$; Feferman 2009). Given that real numbers can be represented as (infinite) sequences of ones and zero's (binary representation), the set of all paths can be understood as the set of all real numbers.

As Feferman (2009, p. 1) writes, when one speaks of the continuum in set theory, it is implicitly understood that one is paying attention only to the cardinal number that these sets have in common, i.e. ignoring differences in structure. However, if we do regard structure,[4] then the six conceptions of the continuum that we discussed so far are *not* the same. In Feferman's view, it is by a kind of "miracle of synergy" of these disctinct forms of the continuum that $\mathbb{R}$ has proved to serve together with $\mathbb{N}$ as one of the two core structures of mathematics; both pure and applied (Feferman 2009, p. 24). It might be tempting to conclude from this synergy that one version or another of the continuum is part of the natural order. Such a conclusion could then, in turn, function as a justification

[4] For example if we, like Feferman, adopt *conceptual structuralism* about mathematics, i.e. the idea that the primary objects of mathematics are structures rather than individuals such as points, lines, numbers, etcetera (for a more detailed exposition of this view, see Feferman 2009, section 2).

for the completeness axiom or property, which in all of the above accounts of the continuum is (either implicitly or explicitly) assumed.[5] However, there are some good arguments *against* such a view. Here we will discuss two.

First, even though to *some* extent the classical mathematical continuum indeed reflects common pre-theoretic intuitions about space, it certainly does not do so in all respects. For example, as we have seen the classical continuum is made up of extensionless points, but it seems very counter-intuitive to suppose that something extensive (i.e. a line) is composed only of elements that do not have an extension (i.e. points; cfr. Bell 2005, Sommer & Suppes 1997). The claim that $\mathbb{R}$ accurately reflects our conception of geometrical and physical lines has been disputed, for example, by Gödel. Gödel's argument was summed up by Putnam as follows: "at least intuitively, if you divide the geometrical line [in two] you would expect that the two halves of the line would be mirror images of each other. Yet, this is not the case if the geometrical line is isomorphic to the real numbers" (Putnam 1992, p. 38, as quoted in Moore 2007). In particular, one half will have an endpoint and the other will not (Moore 2007, p. 76). Moreover, we may wonder: do we *really* know or perceive that space is complete? As Moore (2007) points out, it is widely taken for granted that physical lines are real lines, i.e. that the arithmetical structure of the real numbers uniquely matches the geometrical structure of lines in space, however, how can we be sure that there are no infinitesimals in space? There seems to be no way to know that physical lines are real lines and not hyperreal lines.

Second, a problem with the set-theoretical conception is that it requires Platonism about mathematical objects.[6] Feferman's elaborations of Weyl's predicative theory of the continuum (see section 3.2 below), as well as the work of others (for example, the development of the calculus and geometry in *strict finitism*, see

[5] Interestingly, Dedekind himself warned against the assumption that his continuum is embeded in the real world: "If space has a real existence at all it is not necessary for it to be continuous" (1872, as quoted in Feferman 2009, p. 23 n 24). He took his continuum *not* to be a reflection of the completeness of space, but rather of *what it would be* for space to be complete (Moore 2007, p. 73).

[6] The problem with these set-theoretical conceptions of the continuum is grasping the meaning of "all" in the description of 2^N as consisting of *all* paths in the full binary tree (or *all* functions from N into the set $\{0, 1\}$), or the view of the set $S(\mathbb{N})$ of *all* subsets of natural numbers as a definite totality, so that quantification over it is well-determined and may be used to express definite properties P (Feferman 2009, sections 3.5, 3.6). On the face of it that requires a Platonist ontology, according to which the totality in question somehow exists independently of human conceptions (see section 2.2 of this thesis).

Ye 2011, van Bendegem 2002), shows however that the properties of the continuum needed for its applications in natural science do not require it to have a definite reality in the Platonistic sense (Feferman 2009, section 4.4). The set-theoretic conceptions of the continuum make only sense on the assumption of Platonism; other philosophies of mathematics led to divergent conceptions of the continuum (Feferman 2009, section 1).

It is to non-classical notions of the continuum that we will turn in the next section.

3.2 Non-classical conceptions of the continuum

An invariant among many conceptions that refer to the continuum is an *invariance of scale*: all small pieces of a continuum have the same properties as a bigger continuum (Longo 1999). This way of conceiving the continuum has a long history. As Anaxagoras put it, the essential essential character of the continuum is that (Weyl 1949, p. 41):

> "Among the small there is no smallest, but always something smaller. For what is cannot cease to be no matter how far it is being subdivided."

According to this view expressed by Anaxagoras, a continuum can be subdivided arbitrarily many times, and the parts obtained will always be continua themselves. This is not so according to the classical conceptions of the continuum that we discussed in the previous section. On these conceptions, we can perform an *actual infinity* of subdivisions, and doing so, we arrive at something *qualitatively* different: an extensionless point or real number.

It has been doubted that such classical conceptions of the mathematical continuum accurately reflect our intuitions. For instance, the great mathematician Hermann Weyl held that points are *not* part of our intuitive conception of the continuum (Longo 1999). For this reason, Weyl complains that there is a significant disagreement between theory and intuition, between the mathematical continuum and the intuitive continuum:

> "To the criticism that the intuition of the continuum in no way contains those logical principles on which we must rely for the exact

> definition of the concept "real number", we respond that the conceptual world of mathematics is so foreign to what the intuitive continuum presents to us that the demand for coincidence between the two must be dismissed as absurd" (Weyl 1918, quoted after Feferman 2009, section 4.1).

This dissatisfaction with the set-theoretical conception of the continuum, stemming from the discrepancy between theory and intuition, lay at the basis of alternative developments of the continuum within two foundational programs: *predicativism* and *intuitionism*. On both conceptions the continuum is fundamentally different from the classical continuum, in that it is not made up of an uncountable infinity of extensionless points. We will discuss these two non-classical conceptions of the continuum here in turn. Subsequently, we will discuss another conception of the continuum, which – as we will see – is in some sense somewhere inbetween the classical and the non-classical continuum: the *nonstandard* continuum.

The intuitionists and the predicativists both follow Cantor in the construction of the real numbers, namely as Cauchy (or fundamental) sequences of rational numbers; equality of real numbers is also defined the way Cantor did it (see section 3.1). In order to carry this out, in each framework one must presume the general notion of sequence of rational numbers. However, in both frameworks one does *not* accept the *totality* of such sequences. Consequently, there is no totality of real numbers, only the *concept* of what it means to be a real number. Moreover, these fundamental sequences of rational numbers are regarded *intensionally*, i.e. as described by specific formulae, not as functions determined by their values (Feferman 2009, section 4.3). The difference between predicativism and intuitionism is that the predicativists (at least those stemming from Poincaré and Weyl), in contradistinction to the intuitionist, view the natural numbers – as well as any other set which can be explicitly enumerated – as a definite totality. Consequently, for predicativists, quantification over $\mathbb{N}$ and other explicitly enumerable sets is definite and classical logic is admissible for all statements about them. Intuitionists, to the contrary, view $\mathbb{N}$ as a potential totality and reject the law of excluded middle as applied to arithmetical statements (Feferman 2009, section 4.3). As a consequence, intuitionists have always only a finite amount of information about real numbers, from which Brouwer's theorem follows to the effect that every function of real numbers on a closed interval is continuous (see section 2.3). Both the intuitionists and the predicativists reject the assumption of any completed infinite totalities of *uncountable* cardinality, and in particular the set-theoretical conceptions of $2^{\mathbb{N}}$ and $S(\mathbb{N})$ as definite totalities (ibid.).

Predicativism has its origins in the work of Russell and Poincaré on the paradoxes of early set theory. According to them, (one of) the cause(s)[7] of the typical paradoxes of early set theory – in particular, of Russell's paradox of the class of all non-self membered classes – was that there was a *vicious circle* in the purported definition, made possible by what came to be known as *impredicative definitions* (Feferman 2009, p. 2). According to one notion, a definition is impredicative if it defines an object by *reference to a totality which includes the object to be defined* (Horsten 2007, section 2.4).[8] Impredicative definitions are definitions that violate what Russell called the *Vicious Circle Principle* (VCP). In one of its formulations (Russell 1908, in Crosilla 2009, section 1.3):

> **Vicious Circle Principle**. Whatever contains an apparent variable must not be a possible value of that variable.

Thus, according to the VCP, a sound definition of a collection only refers to entities that exist independently from the defined collection. Such definitions are called *predicative* (Horsten 2007, section 2.4).[9]

Weyl had come to believe that the whole set-theoretical approach involved vicious circles to such an extent that, as he says, "every cell (so to speak) of this mighty organism is permeated by contradiction" (Bell & Korté 2009, section 3.1). In particular during the period 1918-1921 he worked on the problem of providing the mathematical continuum with what he conceived of as a *logically*

[7] Poincaré saw as a *second* cause of these paradoxes the assumption of the *actual infinite*. Russell denied this part of Poincaré's diagnosis (Feferman 2009, p. 2).

[8] As we have seen in section 2.2, the *Axiom of Separation* allows us to form a subset of a given set whose elements satisfy a given property (expressed by a formula in the language of set theory). Given a set B and a formula $\phi(X)$, separation allows us to construct a new set, the set of those elements X of B for which ϕ holds. This is usually informally represented as: $\{X \in B : \phi(X)\}$. Separation may lead to impredicativity in case the formula ϕ contains unbounded quantifiers ranging over the whole universe of sets; in fact, in defining the new set by separation we may thus refer to this very set, contradicting Russell's VCP (Crosilla 2009, section 1.3.1).

[9] As Gödel later pointed out, a Platonist would find this line of reasoning unconvincing. If mathematical collections exist independently of the act of defining, then it is not immediately clear why there could not be collections that can only be defined impredicatively (Gödel 1944, cited in Horsten 2007, section 2.4).

sound formulation. In his well-known work *Das Kontinuum* (1918), Weyl aimed to provide analysis with a *predicative foundation*. That is, he undertook the project of securing mathematical analysis through a theory of the continuum that would make no basic assumptions beyond that of the structure of natural numbers $\mathbb{N}$.[10] The concept of an *arbitrary subset* of the natural numbers was *not* taken as given; only those subsets which are determined by arithmetical (i.e. first-order) predicates are taken to be predicatively acceptable (Horsten 2007, section 2.4). Since only sequences of rational numbers which can be *explicitly described* by formulae exist according to the predicativist, and there are only countably many formulae available at each stage of construction, and moreover principles such as the power set axiom which imply the existence of uncountable totalities are not available, in predicativist mathematics the real numbers constitute only a *countable infinity*.

Building on work in generalized recursion theory, Solomon Feferman extended the predicativist project in the 1960's (see Feferman 1998, 2005). Feferman realized that Weyl's strategy could be iterated into the transfinite: those sets of numbers that can be defined by using quantification over the sets that Weyl regarded as predicatively justified, should be counted as predicatively acceptable, and so on. This process can be propagated along an ordinal path, stretching as far into the transfinite as the predicative ordinals reach, where an ordinal is predicative if it measures the length of a provable well-ordering of the natural numbers (Horsten 2007, section 2.4). Feferman then investigated how much of standard mathematical analysis can be carried out within a predicativist framework, and showed that most of twentieth century analysis is acceptable from a predicativist point of view. But it is also clear that not all of contemporary mathematics that is generally accepted by the mathematical community is acceptable from a predicativist standpoint: transfinite set theory is a case in point (Horsten 2007, section 2.4).

Until the work of Feferman, predicativism has been in a dormant state, at least partly induced by the fact that Weyl himself in the 1920's was won over

[10] Weyl agreed with Poincaré that the natural number system and the associated principle of induction constitute an irreducible minimum of theoretical mathematics, and any effort to "justify" that would implicitly involve its assumption elsewhere (Feferman 1998, p. 52). Since Weyl took the collection of natural numbers as given, his philosophical stance is in a sense intermediate between intuitionism and Platonism (Horsten 2007, section 2.4).

to, in Feferman's terms, the "most radical of the radicals": the intuitionists (Feferman 2005, p. 25). Intuitionism originates in the work of the mathematician L.E.J. Brouwer, and is just as predicativism a product of dissatisfaction with the ideal, nonconstructive methods used by most contemporaries (Bridges & Palmgren 1997, section 3.1). According to Brouwer, the paradoxes of set theory arose from illegitimate extension of the logical laws which hold for finite sets, to infinite sets. Weyl sums up Brouwer's view point as follows:

> "According to [Brouwer's] view and reading of history, classical logic was abstracted from the mathematics of finite sets and their subsets. [...] Forgetful of this limited origin, one afterwards mistook that logic for something above and prior to all mathematics, and finally applied it, without justification, to the mathematics of infinite sets. This is the Fall and original sin of set theory, for which it is justly punished by the antinomies. It is not that such contradictions showed up that is surprising, but that they showed up at such a late stage of the game" (Weyl 1946, as quoted in Bridges & Palmgren 1997, section 3.1).

Thus, compared to predicativism, intuitionism is more radical in that it sees a thread in applying classical logic to *all* infinite sets, even down to the natural numbers.

This rejection of classical logic for infinite sets is *philosophically* justified by, and deriving from, the conception of mathematics as *free creation of the human mind*. According to intuitionism, a mathematical object exists if and only if it can be (mentally) constructed by finite beings (ideal mathematicians). An ideal mathematician can never complete an infinite construction, even though she can complete arbitrarily large finite initial parts of it. In accordance with this conception, intuitionism resolutely rejects the existence of the *actual infinite*; only potentially infinite collections are given in the activity of construction (Horsten 2007, section 2.2). In reasoning about such potentially infinite collections, the intuitionist uses *intuitionistic logic* (see e.g. Moschovakis 1999). This logic was developed by Brouwer's student Arend Heyting, by abstracting from what according to intuitionists a sound mathematical proof consists in. Because these principles also underly other systems of constructive mathematics, such as

Russian recursive analysis and the constructive analysis of Errett Bishop and his followers, intuitionistic logic may be considered the logical basis of constructive mathematics (Bridges & Palmgren 1997, section 3.1).

Intuitionistic mathematics diverges from other types of constructive mathematics in its interpretation of the term "sequence". Normally, a sequence in constructive mathematics is given by a rule which determines, in advance, how to construct each of its terms; such a sequence may be said to be *lawlike* or *predeterminate* (Bridges & Palmgren 1997, section 3.1). Brouwer generalised this notion of a sequence to include *free choice sequences*: infinite sequences of numbers (or finite objects) created by the free will. The sequence could be determined by a law or algorithm, such as the sequence consisting of only zeros, or of the prime numbers in increasing order, in which case we speak of a *lawlike* sequence, or it could not be subject to any law, in which case it is called *lawless* (Iemhoff 2008, section 3.4). Lawless sequences could for example be created by the repeated throw of a coin, or by asking the creating subject to choose the successive numbers of the sequence one by one, allowing it to choose any number to its liking (ibid.). Importantly, a lawless sequence is ever unfinished, and the only available information about it at any stage in time is the initial segment of the sequence created thus far (see also section 2.3 of this thesis). This makes that the intuitionistic continuum is fundamentally different from the classical continuum; in particular, whereas the classical continuum is *discrete*, in the sense that it is composed of individual real numbers which are well-defined and can be sharply distinguished, the intuitionistic continuum is an essentially *continuous* "medium of free development" from which the real numbers are obtained by assembling them from a complex of continually changing overlapping parts (Bell & Korté 2009, section 3.2). As a consequence, as we have already discussed in section 2.3 of this thesis, the intuitionistic continuum is *not* the union of two disjoint non-empty parts; it is *indecomposable* (Bell & Korté 2009, section 3.2). As it was colourfully described by van Dalen:

> "In intuitionistic mathematics [...] the continuum has, as it were, as syrupy nature, one cannot simply take away one point. In the classical continuum one can, thanks to the principle of the excluded third, do so. To put it picturesquely, the classical continuum is the frozen intuitionistic continuum. If one removes one point from the

> intuitionistic continuum, there still are all those points for which it is unknown whether or not they belong to the remaining part" (Van Dalen 1997, quoted in Sanders 2018).

We will now shortly discuss the continuum in nonstandard analysis. As we have seen in section 2.4, the nonstandard continuum $*\mathbb{R}$ is different from the classical continuum $\mathbb{R}$, since it contains besides the standard objects also nonstandard objects: infinitely large and infinitesimal numbers. This extension with respect to standard analysis makes that the real numbers in nonstandard analysis are different from the standard real numbers and it is possible to derive new results. Notably, the nonstandard continuum – as opposed to all other conceptions of the continuum that we have been discussing – is *not scale-invariant* (Longo 1999, section 3.1). As we have seen in figure 2.1 in section 2.4, if we look at the hyperreal line from a nonstandard perspective – i.e. through an "infinitesimal microscope" – we see that every real number a is contained in a *monad* consisting of all hyperreal numbers infinitely close to a. Within this monad the Archimedean property fails, which implies that – from the nonstandard perspective – the nonstandard continuum is *incomplete*: it has "gaps". Both the monad and the gaps, i.e. the incompleteness, cannot be perceived from a standard perspective, and in that manner the scale-invariance fails in the nonstandard continuum.

Furthermore, as was shown by Sanders (2018, p. 41), the continuum in nonstandard analysis is *indecomposable* from the viewpoint of the *standard* universe, and *decomposable* from the viewpoint of the *nonstandard* universe. This is because, as he explains, there is a decidable decomposition of the nonstandard continuum, but this decomposition is *essentially* based on *nonstandard methods*, and should therefore from a standard perspective be disqualified (ibid.). This means, as Sanders observes, that in nonstandard analysis the standard continuum *changes the associated logic*: the *Law of the Excluded Middle* is not available in nonstandard analysis for reasoning about the standard continuum (Sanders 2018, pp. 49-50). Interestingly, this confirms Brouwer's claim that *logic is dependent on mathematics*: if one changes the mathematics (by extending the universe with nonstandard elements) the logic changes with it (ibid.).

The continuum in nonstandard analysis can thus be said to be somewhere inbetween the continuous and the discrete, as well as inbetween the classical and the non-classical. Note that many results of classical analysis hold in the nonstandard universe, but the continuum in the nonstandard universe is *not* complete. This means that these results do *not* depend on what was taken to be the essential property of the continuum: completeness. Thus, as I see it, nonstandard analysis can help us to think about what the continuum actually *is*.

3.3 Against using the classical continuum in physical modeling

Thusfar, we have seen six classical, and three non-classical (or nonstandard) conceptions of the continuum. On most, but not all of these conceptions, the continuum is *scale-invariant*; on most, but not all of these conceptions, the continuum is "gapless" or *complete*; on most, but not all of these conceptions, the continuum consists of an *uncountable* infinity of *extensionless* points or real numbers. So far, we have been comparing these different conceptions of the continuum in terms of their mathematical properties, as well as their connection to diverging views on the nature of mathematics and its objects. But what effect would the differences between those conceptions of the continuum have in physics? We will compare different conceptions of the continuum with regard to the specific context of the case-studies in chapters 4 and 5. In this section, we will discuss in more general terms an argument – of Michael Dummett – to the effect that the classical continuum is *not* suitable to model physical quantities, and we will discuss some arguments for disposing of the classical continuum in modeling physical quantities, based on the appeal to empirical equivalence.

According to Dummett (2000, 2005), the classical continuum is not suitable for modeling physical quantities, because when modeling physical quantities are modeled by a classical continuum, properties are attributed to those quantities which *contradict our pre-theoretic intuitions* regarding those quantities. Naturally, Dummett wants to exclude physical quantities derived as derivative (e.g. *velocity*) as well as physical quantities whose magnitudes are quantized from his discourse and therefore restricts his discussion to modeling *time* as a classical continuum – in his words, to the classical model of time. The classical model, Dummett explains, models time in the following way. As we have seen in the previous section, the classical continuum is composed of an uncountable infinity of points or real numbers; each real number represents a determinate position on the rational line: a point on that line if it is rational, an dimensionless gap in that line if it is irrational. In other words, the ordering of real numbers by magnitude is a dense, complete linear ordering. Consequently, according to the classical model of time – i.e. when conceiving of time by analogy with the classical continuum of real numbers –, time looks as follows. Time is composed of *durationless instants*, arranged in a dense linear ordering; since instants are durationless, no change or motion takes place within any instant. The ordering being dense, there is *between any two instants another instant*, and there is therefore no such thing as "the next" instant after a given one. Moreover, since time is continuous (complete) there are *no gaps* in the sequence of instants (2000, p. 499).

Dummett gives three examples to show that the classical model of time leads to models of physical quantities which contradict our intuitions. All three his examples involve discontinuities. The first example concerns a model of a surface illuminated only by a candle, which at a certain moment "goes out like a light", i.e. the illumination of the surface abruptly vanishes as the candle is extinguished (Dummett 2000, pp. 502-3). It involves a *jump discontinuity* as the intensity of illumination goes instantaneously from some positive value to 0, where it remains for some further time. In the classical model, the illumination is naturally represented by a function $f(t)$ giving the intensity of illumination at each time t within the appropriate interval, and having a jump discontinuity at $t = t_0$, for some t_0 within that interval. The problem, according to Dummett, is that according to the classical model there are *two* distinct such functions f and g with a jump discontinuity at $t = t_0$, with $f(t_0) = 0$ but $g(t_0)$ still positive, while f and g agree for every other value of t within that interval. In other words, the classical model represents the abrupt change as being one of two physically distinct events: one in which the illumination vanishes *at* the instant of change, the surface having positive illumination at every instant *before* the change (which is the situation as described by f); and the other in which the surface continues to have a positive illumination at the instant of change, but 0 illumination for every instant in some interval *after* that instant (which is the situation as described by g). But, as Dummett sees it, there are no two such distinct physical possibilities: nothing could determine whether the surface had zero or positive illumination at the precise instant of change, and we cannot conceive of there being any genuine distiction between the two cases. Thus, in this case the classical model provides a means of differentiating between two physically different states of affairs which cannot possibly correspond to any distinction in physical reality. In sum, according to Dummett, the classical mathematical framework *differentiates too much*.

The second example concerns *removable discontinuities*, i.e. discontinuities that could be removed by changing the value of the physical quantity in question at a single instant (Dummett 2000, p. 503). Imagine a lamp which is always on except for the one instant $t = 1$ at which it is off.[11] According to Dummett, "our conception of physical quantities is plainly such that this supposition makes no sense". That is, according to our intuitions the lamp would be off during some *interval*, and *not*, as discribed in this model, during some *durationless* instant. Accordingly, in Dummett's view, in admitting for such removable discontinuities, "[t]he classical model supplies descriptions for states of affairs which, being conceptually impossible, should admit no description" (ibid.).

The third example concerns an infinite sequence of events which happen in finite time. A third kind of discontinuity is exemplified by the example of a body which oscillates with increasing rapidity in a plane between a position R 1 cm to the right of a point M, and a position L 1 cm to its left (Dummett 2000, pp. 503-504). It begins by swinging from M to L in $\frac{1}{3}$ minute, then from L to M in $\frac{1}{6}$ min, then from M to R in $\frac{1}{10}$ min, the n-th swing taking $\frac{2}{(n+1)(n+2)}$ min. As Dummett explains, the sum of the first n terms of the series $\frac{1}{3} + \frac{1}{6} + \frac{1}{10} + \ldots$ is $\frac{n}{(n+2)}$, so the series converges to 1; hence by 1 min after the start the body will have made infinitely many swings. The problem with this model is in Dummett's view the following. Wherever we suppose the body to be 1 minute after the start, there will be a discontinuity in its spatial position at that instant. For if $f(t)$ gives its position in cm to the right of M t min after the start, f does not approach any limit as t approaches 1 from the left. The classical model of time allows us to define such a model, but in Dummett's view it is conceptually abhorrent that what happens to the body for $t < 1$ does not tell us at all where it will be at the instant $t = 1$. Its position at that instant is *completely indeterminate*: it might be anywhere, so far as its past history goes. As Dummett sees it, we do not suppose that events are as loose and separate as this (ibid.).

In sum, there are in Dummett's view three problems with the classical model of time. First, it differentiates too much, in that it allows for different descriptions

[11] I follow Meyer (2005) in giving this physical interpretation to Dummett's argument.

of events which in his view are physically indistinguishable. Second, it supplies descriptions for states of affairs which obtain only during a durationless instant, which in his view is physically impossible. Third, it allows for descriptions of an infinite sequence of events happening in finite time, which implies that the state of the system at the first instant after this infinite sequence is independent of the instant before, and this according to him contradicts how we conceive of the evolution of physical systems in time. The *right* model of time, according to Dummett, rules out this kind of models and therefore must have some kind of continuity requirement built-in. Dummett therefore proposes to model time in constructive mathematics (cfr. section 2.3 of this thesis), so that the possibility to formulate philosophically problematic physics such as in these three examples is *excluded* on mathematical grounds.

Another argument that Dummett offers to the effect that the classical model is unsuitable (or undesirable) for modeling physical quantities, is that it embodies, as he calls it, a *super-realist* metaphysics (Dummett 2000, pp. 497-498; Dummett 2005, p. 141). Super-realism, as Dummett puts it, postulates states of affairs that subsist *independently of even the theoretical possibility of our knowing of them* (Dummett 2000, p. 497), or that there are true propositions which we understand, but which we cannot ever in principle come to know (Dummett 2005, p. 141). For example, in his view, in physics the classical model embodies the assumption that every physical quantity has, in reality, an absolutely determinate magnitude which is accurately represented, relative to a given unit, by a classical real number. This assumption cannot be tested: we can never, by measurement, identify a specific real number as the magnitude of a physical quantity – to do so would require our measurements to be *infinitely precise*. According to Dummett, three further assumptions are made to uphold the assumption that every physical quantity has in reality a determinate magnitude, namely that (1) there is a determinate answer to hypothetical questions about what the results of ever more precise measurements would be; (2) an infinite process will yield a determinate outcome, namely, the limit of an infinite monotonic sequence extending into the future; and (3) such a sequence converges to a classical real number (rather than to an interval on the real line; Dummett 2000, p. 498).

If we model physical quantities by constructive real numbers instead, we do *not* make such super-realist assumptions. First, in constructive mathematics *there* *are no* true propositions that exist independently of the possibility of us knowing them; in fact, in constructive mathematics a proposition counts as true only if its truth is constructively proven (and not, for example, if merely its negation is disproved). Second, by modeling a physical quantity by a constructive real number – given that, as we have seen in section 2.3, constructive real numbers are only *finitely precise* – we do not assume that this quantity has in reality an infinitely precise magnitude, nor that a sequence of ever more precise measurements converges to an infinitely precise real number. Moreover, in virtue of this finite precision – in particular, in virtue of the failure of *trichotomy* for constructive real numbers – constructivism fits well with the practice of measurement in physics (cfr. Bauer 2012). For example, consider a real-valued measured quantity t and some t_0. Whatever experiment we perform, there will always be a small *measurement uncertainty* Δt. We can decide that $t < t_0$ or $t > t_0$ holds, if by luck $t + \Delta t < t_0$ or $t - \Delta t > t_0$. Otherwise, we might perform a more precise measurement, but there is no guarantee that we can determine the matter in finite time. Thus, we can say that constructive real numbers reflect our *imperfect methods* of determining the values of physical quantities.

As we have seen in section 1.2.1, an important reason for the physicist's ease with infinities are *limits to experimental accuracy*. For example, physicists justify the use of infinite component limits by appealing to the fact that it is virtually impossible to detect observable differences between the behavior of infinite systems and systems consisting of a really big number of components, i.e. of the order of 10^{23} or larger: the difference between the relevant values for infinite systems and those for very large finite systems is usually smaller than the measurement error. It seems that the same argument is available to justify the use of constructive, instead of classical models, in certain physical contexts. Given the fact that measurement techniques are only finitely precise, no measurement will ever be able to distinguish between a finitely precise constructive real number and an infinitely precise classical real number as the value of a physical quantity. In that sense, we can say that the constructive and classical model of a certain physical system will be *empirically equivalent*. Furthermore, as we have seen in section 2.3, constructive total functions are continuous. Since datasets are necessarily discrete, no finite collection of measurements can ever determine whether a physical quantity varies discontinuously, or rather rapidly but continuously. Thus, also in this respect a constructive model will be empirically equivalent to its classical counterpart.

The argument of empirical equivalence is also used by Sommer & Suppes (1997). They developed a constructive system of nonstandard analysis called *elementary recursive nonstandard analysis* (ERNA) to show that we can "dispense with the continuum" in physics. ERNA is an axiomatic approach to nonstandard analysis, which has models in which the completeness axiom is replaced by an axiom stating the existence of infinitesimals (see section 2.4 above). That is, ERNA can be used to construct the *hyperrationals* $*\mathbb{Q}$, i.e. an extension of the rationals $\mathbb{Q}$ which includes all the fractions of (possibly nonstandard) integers with nonzero denominators (Sommer & Suppes 1997, pp. 3,4). As Sommer & Suppes argue, the hyperrationals are rich enough to carry out (large parts of) the mathematics used for scientific purposes, which implies that *completeness is not necessary* in the empirical sciences at all (Sommer & Suppes 1997, p. 4; cfr. Impens & Sanders 2007). Importantly, as they suggest, these nonstandard models and models based on the standard reals are *empirically indistinguishable* (meaning that the difference between them is infinitesimal), and thus there seems to be no reason to prefer the standard models on empirical grounds (Sommer & Suppes 1997, p. 5). In sum, since proofs in ERNA do not rely on the completeness property, and (a large part of) the mathematics used in the physical sciences can be carried out in ERNA, Sommer & Suppes conclude that a significant portion of the mathematics that is used in the physical sciences does not rely on the special properties of the continuum.

Conclusion to Part II

> *[A]s long as science takes the real number system for granted, its philosophers must eventually engage the basic foundational question of modern mathematics: "What are the real numbers, really?"*
> - Solomon Feferman (1998)

Given the differences in nature between mathematics (abstract) and the physical world (concrete), it might be considered *hardly suprising* that there are philosophical problems concerning the applicability of mathematics to the physical world. What might, to the contrary, come as a surprise is that there are many philosophical problems in *pure* mathematics. For instance, one might think that all that matters in pure mathematics is whether or not the theorems are correct – and it is up to mathematicians, and not philosophers, to establish that. What does a philosopher have to say about, e.g., such a well-established mathematical theory as set theory?

Such a view misses something important, namely, how *applied* pure mathematics actually is. The chapters in this second part have shown that there is very much to say about mathematical systems besides whether they are mathematically correct or not. In relation to what we have seen in these chapters, we will here shortly discuss three arguments to the effect that it is a mistake to hold that pure mathematics is merely about mathematical correctness, and that consequently philosophers have nothing relevant to say about it.

First of all, mathematical theories aim to give *precise meaning to terms with whose ordinary meaning we are already familiar*. For example, we have seen that set theory, as well as its alternatives, aims to give precise meaning to terms like *set*, *cardinality*, *larger*, *smaller* (cfr. Potter 2004, p. 205): in application to finite sets we take the theory to formalize the practice of counting which we have practised since childhood, and in application to infinite sets we take the theory to be a natural conceptual extension of the finite case. For example, we have seen that $2^{\aleph_0} > \aleph_0$, but is this true in the same sense that $4 > 2$? Set theory affirms this (for both are instances of Cantor's theorem), but in answering such we are committed to the idea that we should think about infinite sets in the particular manner that set theory prescribes. Similarly, theories of the continuum or real numbers aim to make precise our pre-theoretical understanding of the *completeness* of *space*. Dedekind was convinced that his notion of completeness was the only way of making precise our pre-theoretical understanding of what it would be for space to be complete, and that – although he was "utterly unable to adduce any proof of [his account's] correctness, nor has anyone the power" – he was convinced that he did not "err in assuming that every one will at once grant [its] truth" (1872, as quoted in Moore 2007, p. 73). However, as we have seen, in fact *not* every one agrees that Dedekind accurately captured our intuitions, and alternative accounts of the continuum – mathematically equally coherent – have been developed.

Second, set theory might *seem* to be the quintessential example of an abstract, purely mathematical theory, originating exclusively from intra-mathematical questions, but in fact extra-mathematical, and indeed *physical* and *biological* motives seem to have played an important role in its development (Ferreirós 2004). Cantor was an opponent of the atomic theory of matter, and seems to have developed his set theory with the aim of *application* in natural science, linking his theory with the constitution of matter. In his view, any satisfactory theory of nature must begin with the assumption

> "That the ultimate, properly *simple* elements of matter are given in an *actually infinite number*, and regarding the *spatial* they must be considered as *totally unextended and rigorously punctual*" (Cantor 1885, quoted in Ferreirós 2004, p. 27).

Cantor called the simple elements "monads" or "units", in explicit reference to Leibniz' *Monadology* (Ferreirós 2004, p. 27). He applied his point-set theory to give mathematical rigor to what he called the "organic explanation of Nature", that is, to offer a precise model of the world of monads. His *first physical hypothesis* was that the set of corporeal monads is of cardinality $\aleph_0$, while the set of etheral monads has cardinality $\aleph_1$, i.e. the power of the continuum (Ferreirós 2004, p. 27).

Also Bolzano – which is often praised, also by Cantor himself, for being one of the first mathematicians to embrace the actual infinite – believed that the physical world was made up of infinitely many monads, which he called "atoms".[12] As is explained by Simons (2015), according to Bolzano atoms have zero volume, and since material bodies have finite volume, every material body is made up of infinitely many atoms. Bolzano's physical world is a plenum: there is no space that is not completely filled by atoms, not even a single point. But Bolzano holds that motion is possible even in a plenum, because despite completely filling space, the atoms in it can be compressed together without any of them coinciding and expanded from one another without leaving holes (Bolzano 1978, sections 15, 16). To motivate this thought, Bolzano gives the mathematical argument that there is a one-to-one correspondence between the points on any two line segments, including segments of different sizes (idem, section 16).[13]

Thus, Cantor as well as Bolzano – both key figures in the history of infinitary mathematics – held that their actually infinite sets were a direct reflection of the nature of physical reality, in the sense that sets are composed of infinitely many extensionless points in the same way as reality is made up of infinitely many extensionless monads. They both believed that their mathematical results about such infinite sets revealed truths about physical reality. Whereas this might sound not so shocking – for also nowadays we believe that our mathematical results reveal truths about physical reality – one might wonder whether the acceptance of the atomic theory of matter, i.e. the theory asserting that concrete systems are made up of *finitely* many *extensive* atoms, did not make the applicability of infinitary mathematics to the physical world is *much more complicated* than it was set out to be. That is, as I see it, there is something very intuitive in the idea that *if* physical systems were indeed made up of uncountably many monads as Cantor and Bolzano presumed, *then* this mathematics would apply better or more unproblematically to them, than when – as turns out to be the case – they

[12] Bolzano's (1851) *Paradoxes of the Infinite* is a defence of the notion of infinity against claims that it is paradoxical, absurd, or self-contradictory. Cantor spoke highly of this work in his 1883.

[13] Simons writes instead that Bolzano motivates this thought by noting that there are *just as many* points on a line of given length as there are on a line of greater or lesser length (Simons 2015, p. 1080). This is not true: in the references that Simons gives, nor anywhere else to the best of my knowledge, Bolzano writes such a thing. For Bolzano, the fact that two sets can be put in one-to-one correspondence is *not* sufficient for these sets to be of the same *size* (see e.g. Bolzano 1851, section 20, cfr. van Wierst 2016, section 5).

are made up of finitely many atoms. For example, it seems that we have much less reason to assume that real-valued functions accurately reflect the behavior or constitution of systems made up of finitely many atoms, than of systems made up of an uncountable infinity of atoms. Given the finitude of physical systems at the microlevel, it seems that the fact that our mathematics is *only an approximation* to the physical world is already built-in from the start, given the mismatch between, e.g., the infinitary nature of the mathematical continuum and the finite nature of concrete objects in the physical world.

A third argument to the effect that pure mathematics is not only about mathematical correctness, is that the debate on the foundations of mathematics proves itself to be no objective, but a *subjective and also very social* affair. To start, the majority of mathematicians is only familiar with classical mathematics, because this is what is taught in high-school and regular university mathematics classes. Objections against classical mathematics are often not known and rarely well-understood by regular mathematicians (Waaldijk 2005). Further, mathematicians often take their views on the foundations of mathematics highly personal, and discussion about them is almost a matter of politics (Sanders 2018). To name one famous example, Hilbert proclaimed that giving up set theory is a "betrayal against our science" and vowed that "from the paradise, that Cantor created for us, no-one can expel us" (Hilbert 1926).

Curiously, Einstein was apparently disturbed by the fierce struggle between competing views on the foundations of mathematics[14] and at a certain occasion he exclaimed (Sanders 2018):

> "What is this frog and mouse battle among the mathematicians?"

In this context, one might wonder *how relevant* the debate about the foundations of mathematics actually is for physics. Much of the mathematics used in physics – e.g. the calculus, differential equations, real and complex analysis, countable algebra, the topology of complete separable metric spaces, etcetera – is prior to, and thus – at least in some sense – independent of the introduction of abstract set-theoretic concepts (cfr. Simpson 2009, quoted in Sanders 2017). With this in mind, one might argue that physics is not based on set theory at all, and thus that the debate on mathematical foundations is simply *irrelevant* for physics.

[14] Precisely, Einstein was disturbed by the so-called *Grundlagenstreit* between Hilbert's *formalism* and Brouwer's *intuitionism* (Sanders 2018).

However, I contend, to better understand the relation between mathematics and the physical world to which we apply it, one should think deeper about the nature of the mathematics that is used in physics: what justifies its concepts and methods? In other words: what is its *foundation*? This is particularly relevant, I think, when we are concerned about philosophical problems raised by infinite idealizations. If we are talking about infinite systems, do we mean *actually* infinite? If we idealize a concrete system to be a continuum, do we mean it consists of an *uncountable* infinity of components? Answers to these questions might be crucial for understanding, and dissolving, these paradoxes.

In Part III of this thesis we will start this undertaking.

Part III

Two paradoxes of the applied infinite

Introduction to Part III

In this third part of the thesis we will discuss two paradoxes of the applied infinite. That is to say, we will discuss two cases of physical theories that employ infinite idealizations which raise philosophical problems in their interpretation. The paradoxes that we will discuss are (1) the paradox of phase transitions, which regards the use of the so-called *thermodynamic limit* in statistical-mechanical theories of phase transitions (chapter 4), and (2) the paradox of Norton's Dome, which is a case of indeterminism in classical or Newtonian mechanics (chapter 5). Both chapters start with a short summary of the paradox, which is followed by, first, an exposition of the relevant physics, second, a discussion of the philosophical debate on the paradox in the literature (to the extent that it is pertinent to this thesis), and third, my analysis of the paradox, with a special focus on the mathematics used in the physics. The aim of these chapters is to answer, specifically for these case studies, the central question of this thesis: are the philosophical issues with the infinite idealizations in these case studies relative to classical mathematics and its notion of infinity?

As mentioned in the introduction to this thesis, in general the philosophical problems surrounding infinite idealizations connect to other issues in philosophy of science. Here, a word on the relevant issues with regard to our case studies. Paradox (1) statistical mechanical theories of phase transitions is often discussed in connection to the issue of *inter-theory relations*, in particular, with respect to the question whether thermodynamics can be *reduced to* statistical mechanics (or vice versa, depending on the notion of *reduction* one assumes). Phase transitions are often taken as evidence for claims of irreducibility: the need for the thermodynamic limit supposedly shows that thermodynamics is irreducible to statistical mechanics. Statistical mechanical theories of phase transitions also relate to the issue of *scientific explanation* – specifically, explaining the success of one theory in terms of another: the need for the thermodynamic limit seems to be an obstacle in explaining the macrolevel phenomenon of a phase transition in terms of the system's behavior at the microlevel.

Paradox (2), Norton's Dome was introduced by Norton to support his claim that *causation* is a notion belonging to folk science, rather than a fundamental principle underlying all natural processes which unifies all domains of science at some deeper level. According to him, the Dome is an undeterministic Newtonian system, and thus shows that not every effect is produced by a lawful necessity from some cause. However, the Dome raises questions regarding the *legitimacy of idealizations*: it is argued in the literature that in virtue of the idealizations employed in it, the Dome falls outside the domain of Newtonian mechanics, that is, that contrary to Norton's claims it is *not* a proper Newtonian system. As such, the Dome raises the question: how far and in which manner may we idealize, so that we still have a legitimate idealization? It seems natural to relate the question of legitimacy in this context to the notion of *physicality*; the distinction between physical and unphysical systems is fundamental to the intuitions of physicist, but it is never made explicit what physicality actually amounts to. In fact, it seems that physicists remain unimpressed by the consequences of Norton's Dome for Newtonian mechanics, simply because they regard the Dome as an *unphysical* system. Norton's Dome thus brings up the question whether in philosophy we may press physical theories towards their borders, or instead should focus on the restricted domain of those systems that physicists consider to be physical, and thus relevant.

Chapter 4

Paradox I: Phase transitions

Phase transitions are successfully analyzed and predicted within statistical mechanics, the area of physics which analyzes macroscopic properties of systems in terms of the behavior of their microscopic constituents. These theories require the so-called 'thermodynamic limit' to obtain phase transitions. But systems in the thermodynamic limit have infinite component-number and volume, whereas the component-number and volume of concrete systems are finite. Do our best theories of phase transitions tell us that no phase transitions can occur in concrete systems?

4.1 The physics

To start, let us consider how common statistical-mechanical theories of phase transitions give rise to the paradox. Statistical mechanics is the area of physics which aims to explain macroscopic (thermodynamic) properties of physical systems in terms of their microscopic constituents. The macroscopic properties of a statistical-mechanical system are functions of the number of its components, and since the number of components of these systems is usually so large that any further increase hardly makes a difference in the values of these functions, it is standard practice in statistical mechanics to take these functions to their limit and consider the component number of the system to be infinite. In most cases in statistical mechanics limits are taken for mere convenience – namely, to simplify the calculations. However, as is widely discussed in the philosophical debate on infinite idealizations in science, this seems to be different in the case of phase transitions: the most successful statistical-mechanical theories of phase transitions *require* taking the so-called *thermodynamic limit*, that is to say, these theories require systems in which the component number and volume are taken to infinity (in such a manner that the ratio between the two remains fixed).[1] Without taking the thermodynamic limit, namely, within these theories no phase transitions occur, and thus it is – in some sense to be spelled out below – *necessary* to consider an infinite system in order to obtain a phase transition. As it was famously put by Callender (2001), it seems that until we say more, these theories give us a *mathematical proof* that the finite systems in the real world *cannot undergo* phase transitions. Thus, the much debated paradox of phase transitions consists in the fact that, on the one hand, concrete systems in the real world display phase transitions and – given the atomic theory of matter – are finite, while, on the other hand, our best theories tell us that phase transitions occur *only* in infinite systems.[2]

Why is it taken to be necessary to take the thermodynamic limit in statistical-mechanical theories of phase transitions in order to show that a system displays a phase transition? In short: it is considered necessary to take the thermodynamic limit in these theories, because without doing so the mathematical entities which are taken to represent phase transitions (namely, discontinuities in the functions describing the thermodynamic properties of the system) do not appear. Let us consider in somewhat more detail why this is the case. As is well-known, the occurrence of a phase transition goes hand-in-hand with sudden and large changes in the system's macroscopic properties. For example, when vapor condenses into liquid water there is a large and sudden decrease in the volume. The phase

[1] The phase transitions with which we will be concerned are the so-called 'first-order' phase transitions. These are phase transitions where the system goes from one phase to another crossing a so-called 'coexistence line' or 'phase boundary', which corresponds to a discontinuity in a macroscopic (thermodynamic) property, which in turn corresponds to a discontinuity in the first derivative of the free energy of the system (see below). There are also 'continuous' phase transitions, in which a system changes phase without crossing a phase boundary (for details, see Goldenfeld 1992). Both kinds of phase transitions are philosophically interesting, but given that the physics of continuous phase transitions is much more complicated than that of the first-order ones, we will focus on first-order phase transitions here.

[2] I agree with Shech (2013, 2015), that no paradox can be found in these facts *per se*, and that, in order to obtain a paradox, assumptions have to be made concerning the representative relationship between the infinite (abstract) and the finite (concrete) systems. However, this issue will not concern us here.

transition from vapor to liquid water can accordingly be represented graphically as in figure 4.1 (read from left to right).

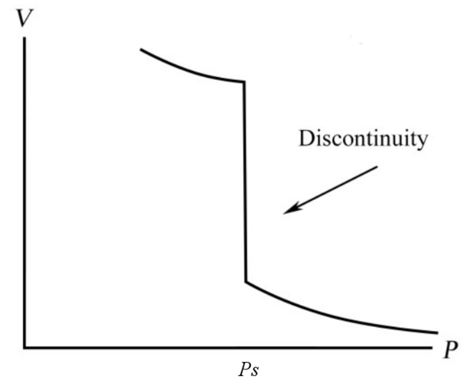

Figure 4.1: Phase transition in terms of volume (V) and pressure (P) (after Stanley 1971)

The abrupt change or 'jump' – in mathematical terms: the *discontinuity* – in the graph of the volume V in figure 4.1, which occurs when the pressure reaches the saturated vapor pressure P_s, represents in this example the phase transition from vapor to liquid water. Now, macroscopic properties of a system – volume and pressure included – can be obtained from appropriate differentiations of the system's *free energy*.[3] Accordingly, the large and sudden changes in the macroscopic properties which characterize phase transitions, occur when the free energy F itself has a *singularity* (meaning in this case a point where F is not differentiable). The free energy F of a system is a macrolevel (thermodynamic) property, and it is related to microlevel (statistical-mechanical) properties of the system via the *partition function* Z:

$$F = -k_B T ln Z \tag{4.1}$$

(where k_B is Boltzmann's constant and T the temperature).[4] Given that no

[3] There are various kinds of free energy, each appropriate in different contexts. We will be concerned with *Helmholz* free energy $F = U - TS$ (where U is the internal energy of the system, T the absolute temperature, and S the entropy), which is the maximum amount of work that a system can do at a constant volume and temperature.

[4] For more details, see e.g. McComb (2004).

singularities can obtain in k_B and T, we see from equation 4.1 that if F has a singularity, then Z must have a singularity too. And this is problematic. The partition function Z, namely, is a function of the temperature of the system and the sum of all particular total energies E_r that the system can have:

$$Z = \sum_r e^{-\frac{E_r}{k_B T}} \tag{4.2}$$

Thus, for any *finite* amount of particles, Z is a finite sum of analytic functions and hence cannot feature any singularities. This is the reason why physicists take the thermodynamic limit, i.e. let the particle number and volume of the system go to infinity while keeping the system's intrinsic properties, such as its density, constant: systems in the thermodynamic limit Z *can* feature singularities. For these reasons, taking the thermodynamic limit is necessary in order to obtain the discontinuities which are taken to represent phase transitions.

Note that the fact that it is considered *necessary* to idealize the system to be infinite makes that the idealization in theories of phase transitions problematic in a way that other idealizations are not: usually, it is not too problematic to de-idealize idealized systems to resemble more the concrete systems they are meant to represent (for example, we can introduce some friction in frictionless planes). But – since phase transitions are represented by discontinuities – in the case of phase transitions such de-idealization seems impossible: whether or not discontinuities in Z obtain is an *all-or-nothing matter* and for *any* finite number of particles, no matter how big, Z is analytic and no discontinuities obtain. Thus, discontinuities – and hence the phase transitions which they are taken to represent – occur only for infinite N. In this manner, our best statistical-mechanical theories of phase transitions seem to tell us that phase transitions occur *only* in infinite systems.

In sum, the paradox of phase transitions consists in the fact that, on the one hand, the finite concrete systems in the real world display phase transitions, while, on the other hand, our best theories tell us that phase transitions occur only in infinite systems. A crucial component in this paradox is the fact that phase transitions are taken to be represented by mathematical discontinuities: it is in order to obtain these discontinuities that taking the thermodynamic limit, and thus the assumption of an infinite system, is necessary. We will discuss the philosophical problems which the paradox of phase transitions raises in more detail in the next section.

4.2 The philosophical debate

As a result of the apparently necessary appeal to the thermodynamic limit in statistical-mechanical accounts of phase transitions which we discussed in the previous section, we get two – related – philosophical problems. First, how to interpret the *infinite systems* obtained by taking the thermodynamic limit: how, exactly, do they relate to the finite systems in the real world? Second, how to interpret the *discontinuities* in the functions describing the microscopic properties of these infinite systems, which are taken to represent phase transitions in statistical-mechanical theories: finite systems cannot display discontinuities in their microscopic properties. In this section, we will discuss several takes on these two conceptual problems with the thermodynamic limit which have been put forward in the philosophical literature on infinite idealizations in science.

The claim that idealizing systems as infinite is necessary in statistical-mechanical theories of phase transitions is philosophically defended by Batterman (2005). Batterman's view that this infinite idealization is ineliminable is motivated by his take on the issue of how to interpret the discontinuities which denote phase transitions in these theories. Famously, Batterman argues that these mathematical discontinuities have genuine physical significance: they represent *physical discontinuities* (cfr. Batterman 2005, p. 233). What are physical discontinuities? Batterman writes:

> "As an instance of a physical discontinuity we can take the phenomenon with which we are concerned, namely, the observed *qualitative* distinctions between [...] the phases of a "fluid" – gaseous, liquid, and solid" (Batterman 2005, p. 233; emphasis in the original).

Thus, according to Batterman, physical discontinuities at phase transitions are the observed qualitative distinctions between phases. Statistical mechanics should thus in his view provide a *qualitative* characterization of phase transitions (cfr. Batterman 2005, p. 230). Furthermore, these physical discontinuities are in Batterman's view *rightly represented* by mathematical discontinuities: "their faithful representation demands curves with kinks" (Batterman 2005, p. 235). In other words, Batterman's understanding of the infinite idealization in statistical-mechanical theories of phase transitions as ineliminable follows from his views that, first, statistical mechanics should provide a qualitative characterization of phase transitions – irrespective of whether such a characterization is consistent with microscopic theories -, and second, phase transitions understood as qualitative distinctions ('physical discontinuities') between phases are rightly represented by mathematical discontinuities in the functions describing the system's macroscopic properties in statistical-mechanical theories.

This second view is puzzling: why would the qualitative distinctions at phase transitions be rightly represented by mathematical discontinuities? Or, to put the matter differently: in which sense are these 'physical discontinuities' *discontinuous*? Unfortunately, Batterman does not clarify, so let us try to shed some light on the matter. An argument to the effect that phase transitions understood as qualitative changes are discontinuous could go as follows. One could argue that systems at phase transitions 'jump' from one qualitative state into another, in the sense that they go from one qualitative state (i.e. phase) into another, without going through any intermediate qualitative state (phase). The problem with this argument is that the qualitative changes at phase transitions in concrete systems are not discontinuous in this sense: concrete systems do not immediately switch from one state to another, but rather transition *gradually* into another phase. This is because, first, driving parameters behind the phase transition – such as temperature and pressure – are unevenly distributed in concrete, extended bodies of matter, which makes that some regions of the system transition before others. For example, think of some boiling water, where the upper molecules turn into vapor before those at the bottom. Second, if – as it is usually held – macrolevel phenomena in concrete systems can be explained in terms of the behavior of the system at the microlevel, then actual discontinuities at the macrolevel cannot occur: microlevel theories tell us that the particles making up the system may change their configuration rapidly, but always continuously.[5] For these reasons, in concrete systems there often is a gray area between phases in which it is not

[5] Cfr. e.g. Norton (2012, 2014), Shech (2013). If I understand him correctly, then Batterman takes his 'physical discontinuities' (at phase transitions and in other cases) to show that macrolevel phenomena in concrete systems *cannot* be explained in terms of the behavior of the system at the microlevel (2005). However, as I see it, this argument is unconvincing as long as it is not clear exactly in which sense these 'physical discontinuities' are discontinuous.

clear in which phase – if any – the system is. This gray area between phases is 'idealized away' in the thermodynamic limit, where systems can be said to be always in one phase or another. Consequently, *only* phase transitions in *infinite* statistical-mechanical systems – and *not* those in concrete systems – can be discontinuous qualitative changes in the sense suggested here. Accordingly, this argument does not justify the representation of phase transitions *in concrete systems* by mathematical discontinuities.

Another way to argue that phase transitions understood as qualitative changes are rightly represented by mathematical discontinuities, might be to contrast phase transitions to another kind of changes that these systems can display, which do not involve observed qualitative distinctions.[6] Namely, one could point out that systems can undergo *two kinds of changes* in their thermal properties: (1) changes *within* a phase (e.g. from liquid water at some temperature and pressure to liquid water of a higher temperature and pressure) and (2) changes *between* phases, i.e. phase transitions (e.g. from liquid water to vapor), and argue that one needs functions which are discontinuous at some points and continuous elsewhere to represent these different kinds of changes mathematically. The qualitative changes ('physical discontinuities') at phase transitions can then be said to be discontinuous in the sense that phase transitions involve the discontinuation of a certain qualitative state, but it is not clear why discontinuity in this sense demands representation by mathematical discontinuities. Certainly, using functions which are discontinuous at some points and continuous elsewhere is *a* way to represent mathematically the distinction between phase transitions and changes within a phase, but it is definitely *not the only way* to do so. It is not clear why other mathematical signatures for phase transitions – such as for example sufficiently steep gradients of the relevant functions – would do worse. In fact, contrasting phase transitions to changes within a phase and representing them by discontinuous versus continuous changes in the value of the relevant functions, presupposes that there is a clear-cut distinction between phase transitions and changes within a phase, which is *not* the case in *concrete* systems. As previously discussed, in concrete systems there often is a gray area between phases in which

[6] This argument presupposes that it is possible to define 'observed qualitative distinctions' in such a way that *only* (2) phase transitions – and *no* (1) changes within a phase – involve observed qualitative distinctions. It is not clear to me that this can be done, but we will ignore this issue here.

it is not clear in which phase – if any – the system is, which makes that whether or not a concrete system displays a phase transition is not an all-or-nothing matter. A clear-cut distinction between phase transitions and changes within a phase is present *only* in *infinite* statistical-mechanical systems. Accordingly, also this argument does not justify the representation of phase transitions *in concrete systems* by mathematical discontinuities.

The upshot of the above is that even if phase transitions are understood as qualitative distinctions – 'physical discontinuities' – between phases in statistical mechanics, it does not follow that they should be represented by mathematical discontinuities: we did not manage to identify a sense in which these physical discontinuities are discontinuous in such a manner that, as Batterman argues, they 'demand curves with kinks'. Both arguments we presented to justify the representation of phase transitions by mathematical discontinuities already *presumed* an infinite system, and therefore do not justify the appeal to infinite systems in order to represent phase transitions in concrete systems. Since the alleged need for discontinuities implies the need for an infinite idealization – the thermodynamic limit – in statistical mechanics, it seems not true that, as Batterman argues, we need to idealize concrete systems as infinite in statistical mechanics in order to obtain a qualitative characterization of phase transitions. In sum, it is on Batterman's account thus clear how to interpret the discontinuities in the functions describing the thermal properties of systems in the thermodynamic limit: these mathematical discontinuities denote physical discontinuities which occur at phase transitions in concrete systems – whatever they might be. How to interpret the infinite systems obtained in the thermodynamic limit is on this account however not clear – in fact, on this account it remains an open question why we need to idealize concrete systems as infinite in statistical mechanics in order to represent their thermal behavior at phase transitions.

There are alternative philosophical accounts of the use of the thermodynamic limit in statistical-mechanical theories of phase transitions which do not run into the same problems. One route taken by several authors in the literature is arguing that in statistical-mechanical theories of phase transitions *we do not* – or not *necessarily* – *consider infinite systems*. That is to say, several authors point out that taking the thermodynamic limit (and other similar limits) does not *in itself* amount to considering an infinite system. We find two different arguments to this effect in the literature on infinite idealizations in physics. One of these arguments presses on a distinction between *two ways of taking limits*. This distinction was the topic of two papers of Norton (2012, 2014), but can already be found in earlier literature.[7] The other argument is from Pincock (2014), which argues that when we take limits we are, in fact, *decoupling* parts of the representation from their original interpretation, and as a consequence we obtain a representation with *schematic content*. We have discussed these views in the more general context of infinite idealization in science in sections 1.1.1 and 1.2.2; here we will apply that discussion to the specific context of the statistical mechanics of phase transitions.

As we have seen in section 1.2.2, in Norton's view there are two different ways in which limits are used in physical practice: (1) as *approximations*, which means that one tracks the behavior of a system's properties as the number of components becomes arbitrarily large, or (2) as *idealizations*, which means that one considers an actually infinite system as a surrogate for large systems of finitely many components (Norton 2014, p. 205). Importantly, according to Norton, when one uses a limit as (1) an approximation, one takes this limit in order to obtain *propositions*, which can be used to analyze finite systems because these propositions are *approximately true* of (large enough) finite systems, in the sense that the difference between the limit-value and the actual value of the function describing the properties of the system is small – in general much smaller than the measurement error; when one uses a limit as (2) an idealization, to the contrary, one takes this limit in order to obtain an actually infinite *system*, which – one hopes – can be analyzed instead of the finite system, because some of the properties of the infinite system approximate some of the properties of the target system (Norton 2012, section 2.1; see section 1.2.2 of this thesis).

As Norton shows however, things often go wrong when using limits as idealizations: the properties of limit systems might not agree with the limit of the properties of finite systems, and one obtains either an unsuitable idealization or no idealization at all (Norton 2012, section 3; section 1.2.2 of this thesis). For this reason, he argues that it is often better to *dispense* with idealizations and consider just approximations. In particular in statistical-mechanical accounts of phase transitions, in Norton's view, we do not need to employ idealizations: approximations are sufficient. That is to say, in his view we can simply consider

[7] For an overview, see Mainwood (2005).

the propositions obtained by taking the thermodynamic limit (i.e. the approximations) and there is no need to consider an infinite system (i.e. the idealization): "In this particular case of phase transitions [...], if infinite idealizations are employed, far from being ineliminable, the infinite idealizations can be and should be eliminated" (Norton 2012, p. 223).

The key point of Norton's views for our purposes is that when we take the thermodynamic limit to generate an idealization, we consider, according to him, a system consisting of an *actual infinity of components* (cfr. Norton 2012, p. 215, Norton 2014, p. 197). When, instead, we use the thermodynamic limit only to generate approximations, in his view we consider a system's properties as the *number of components becomes arbitrarily large* (Norton 2012, p. 225). Disposing of the infinite idealizations thus comes down to disposing of the actually infinite systems in favor of systems of any large enough but finite size. Thus, in other words, in Norton's view we use the thermodynamic limit in the 'good' way when we do *not* take it to constitute an actually infinite system, but when we consider systems of any (large enough) *finite* number of components instead.

On Pincock's account, what Norton calls 'idealizations' do not even occur in physical practice. According to Pincock, as we have seen in section 1.1.1, when we take infinite limits in physics *we do not ever assume* that these quantities are infinite (Pincock 2014). This is because, as Pincock puts it, "false assumptions take us from an interpreted part of a scientific representation to an idealized representation where the particular part is no longer interpreted" (Pincock 2014, p. 2963). Thus, in Pincock's view, the false assumption of an infinite component number N in the thermodynamic limit, *changes* the representation in which N denotes the components number into one in which N is a place holder. As he puts it himself, the infinite idealization transforms the representation so that it has merely schematic content (ibid.). This way, according to Pincock, we obtain a representation which can be *accurate no matter what the particle number is* (as long as it is large enough; Pincock 2014, pp. 2963, 2969).

It might be argued that Pincock's conception does not accurately describe physical practice when it comes to the statistical mechanics of phase transitions, for at least some physicists working in this field clearly consider themselves studying actually infinite systems (e.g. Lanford 1975 and Ruelle 1999, 2004, see Norton 2014). But leaving aside the question whether or not Pincock's account is *descriptively* accurate, I propose to read him *normatively.* Read in this manner, on Pincock's account we should interpret the mathematical representation of a system in the thermodynamic limit *not* as referring to an actually infinite (ideal) system, but instead we should acknowledge that this mathematical representation does not tell us anything about the system's size (cfr. Pincock 2014, pp. 2963, 2969-2970). Thus, Pincock agrees with Norton that we use the thermodynamic limit in the 'good' way when we do *not* take it to constitute an actually infinite system, but when we consider systems of any finite (large enough) number of components instead, and analyze their properties when we ignore the exact size of these systems.

How understanding 'infinite' as 'large but finite' – instead of 'actually infinite' – solves the paradox of phase transitions, can be seen when we take into account Pincock's view that the infinite idealization transforms the representation so that it has merely schematic content. As we have seen, according to Pincock the false assumption of an infinite component number N that we make when taking the thermodynamic limit, changes the representation in which N denotes the component number to one in which N is a place holder. This means that – instead of telling us that for infinite N, the system will display a phase transition – statistical-mechanical theories of phase transitions which appeal to the thermodynamic limit will tell us that for any (large enough) number N, a system of N components will display a phase transition.

In sum, in the views of Norton and Pincock, the appropriate use of the thermodynamic limit does *not* amount to considering actually infinite systems, but rather gives us propositions which are approximately true of (Norton's terminology) or representations which are accurate for (Pincock's terminology) systems of any large enough but *finite* size: on Pincock's account, we may consider the limit system, but we should disregard its extensive properties – that is, we should take the particle number N and the volume V to be merely place holders -, and thus we should *not* consider such a system to be infinite; on Norton's account we should disregard the limit system entirely, and just consider the propositions about the limit properties. Accordingly, on their accounts it is no problem how to interpret infinite systems which can be obtained by taking the thermodynamic limit: they are idealizations which can and should be disregarded. Thus, on Norton's and Pincock's account, statistical-mechanical theories of phase transitions which use the thermodynamic limit do *not* tell us that phase transitions occur only in infinite systems. Furthermore, on their accounts the mathematical discontinuities need not to be interpreted as literally true of concrete systems (denoting something like Batterman's 'physical discontinuities'): the idealized, discontinuous phase transitions which occur in infinite systems are taken to be only approximate representations of the continuous phase transitions which occur in concrete systems. This can be considered a benefit, because on this interpretation the *consistency* of statistical-mechanical theories of phase transitions which employ the thermodynamic limit with atomism is maintained. Since it is not clear what Batterman's physical discontinuities actually *are*, in particular, in which sense they are *discontinuous*, there seems to be no reason why the mathematical discontinuities denoting phase transitions should be considered physically real.

Thus, the views of Norton and Pincock provide a better response to the conceptual problems with the thermodynamic limit than Batterman's view does. In particular, the paradox of phase transitions seems solved on their accounts. For this reason, I will reject Batterman's defense of the need for infinite systems, and embrace Norton's and Pincock's view that the appropriate way to use the thermodynamic limit is to obtain propositions which are approximately true of or representations which are accurate for systems of any large enough but finite size, and thus to consider, instead of actually infinite systems, systems of large but finite size.

4.3 Analysis

In the previous section we have seen that the solution of Norton and Pincock to the paradox of phase transitions – which we embraced – is to interpret the thermodynamic limit as referring to systems of any large finite size and *not* to actually infinite systems. Accordingly, it seems that where statistical-mechanical theories of phase transitions which appeal to the thermodynamic limit talk of infinite systems, this *should be taken to mean* systems of large but finite size. In other words, it seems that the *correct interpretation* of the term 'infinite' in these theories, is *not* 'actually infinite', but rather 'large but finite'.

In my (2017) paper on the paradox of phase transitions, I took the solution of Norton and Pincock to suggest that philosophical problems with statistical-mechanical theories of phase transitions might be solved by developing these theories in *constructive mathematics*. For one might wonder whether, by embracing the solution to the paradox proposed by Norton and Pincock, we have gotten to the bottom of the problem: if it is philosophically pertinent to interpret infinity as 'large but finite' instead of 'actually infinite' in statistical-mechanical theories of phase transitions when talking about infinite *systems*, might it then not be the case that it is philosophically appropriate to interpret infinity *in general* as 'large but finite' in this context? That is, in other words, I proposed in that paper to take the solutions of Norton and Pincock a step further and interpret infinity *consistently* as 'large but finite', instead of only when talking about systems – that is, to develop a statistical-mechanical theory of phase transitions in constructive, instead of classical, mathematics. I worked out the potential philosophical benefits of a constructive statistical-mechanical theory of phase transitions in that paper; we will here summarize the argument, and use it for further discussion.

As we have seen in section 2.3, constructive mathematics rejects the notion of actual infinity and replaces it with *potential infinity*. According to this notion, a collection is infinite if it can be indefinitely extended, and hence the notion refers to arbitrarily large but finite quantities (cfr. Fletcher 2007). As we have seen in section 2.3, constructive mathematics differs from classical mathematics – in virtue of its different conception of infinity – principally with respect to the conception of the *continuum* and of *functions*. These differences are relevant for both philosophically problematic aspects of thermodynamic limit discussed in section 4.2: that systems are idealized to have an actual infinity of components, and that the changes in thermal properties that concrete systems display at phase transitions are idealized to be discontinuous.

Concerning the first problem, i.e. how to interpret the infinite systems in the thermodynamic limit, we note that, since infinity is understood as potential instead of actual infinity, from a constructive viewpoint talk of actually infinite systems is meaningless. Thus, in a theory of phase transitions formulated within constructive mathematics, we cannot idealize a system to consist of an actual infinity of components. Accordingly, we seem to be forced to use the thermodynamic limit in what we identified in the previous section as the appropriate manner: to obtain propositions which are approximately true of or representations which are accurate for systems of any large enough but finite size (see section

4.2). Concerning the second problem, i.e. how to interpret the discontinuities that represent phase transitions, we note that in constructive mathematics all functions defined on every real number are continuous (see section 2.3). It remains to be seen how to construct functions describing the thermodynamic properties of systems at phase transitions, given that the discontinuous functions obtained by taking the thermodynamic limit in the classical theories do not even qualify as functions within constructive mathematics, but for now we will assume that this can be done in a satifactory way.

As I explained in my (2017) paper, a constructive statistical-mechanical theory of phase transitions can be considered a *de-idealization* of classical statistical-mechanical theories. Reasons for this are the following. As we have seen in section 2.3, constructive real numbers are only finitely precise. That means, first, that physical quantities and their measurement are, when denoted by constructive real numbers, *not* idealized to be infinitely precise, as is the case with classical real numbers. Second, this means that functions defined on them – and thus in particular the functions representing the thermal behavior of the systems under consideration – are continuous.

These continuous functions, in turn, can be considered a de-idealization with respect to the classical discontinuous functions for two reasons. First, they may be taken to represent the system's microlevel components (i.e. particles) to come apart rapidly at phase transitions, instead of making a discontinuous 'jump'. In contrast to the classical discontinuous functions, these constructive continuous functions can thus be said to be *consistent* with microlevel theories. Batterman might object to this argument that this is a de-idealization *only* with respect to what he calls *P*hysics, i.e. the – according to him *mistaken* – hierarchical picture of physics in which it is presumed that macrolevel theories can be reduced to microlevel theories (see subsection 1.1.2). That is, in other words, Batterman might object that without the assumption that our statistical-mechanical theory of phase transitions can be reduced to, or is consistent with, microlevel theories, there is no reason to presume that these discontinuous functions are *idealizations*. However, as we have discussed in section 4.2 above, *even if* we want a *qualitative* representation of phase transitions – and don't require consistency with microlevel theories – and understand the qualitative distinctions at phase transitions, following Batterman, as physical discontinuities, *then still* it does not seem to be necessary to denote these physical discontinuities by mathematical discontinuities, for it remains unclear *in which sense* these qualitative distinctions – 'physical discontinuities' – are *discontinuous*. Thus, Batterman presumes without argument that qualitative changes between phases are discontinuous, which can be regarded an idealization.

Second, continuous functions can be considered a de-idealization with respect to the classical discontinuous functions, because, as we have seen in section 4.2 above, representing phase transitions by discontinuities idealizes both the distinction between different phases, and the distinction between changes within a phase and changes between phases (i.e. phase transitions) to be an all-or-nothing matter. Continuous changes instead reflect the fact that in reality both these distinctions are gradual. For example, whereas discontinuous functions idealize systems to be always *clearly* in one phase or another, continuous functions do justice to the fact that although there are values of – e.g. – pressure and volume at which we happily say that a certain system is in its liquid or in its gaseous phase, there are also values at which we would say that it is in no particular phase at all.[8]

In sum, classical statistical-mechanical theories of phase transitions which use the thermodynamic limit idealize systems to have an actually infinite number of components and idealize phase transitions to occur discontinuously. As a consequence, systems are idealized to be always clearly in one phase or another, and the distinction between phase transitions and changes in thermal properties within a phase are idealized to be clear-cut. All these idealizations in such classical theories are made possible by the classical conception of real numbers as infinite objects, and thus the idealization of physical quantities and our measurement of them as infinitely precise. In constructive mathematics, real numbers are considered to be only finitely precise. As a consequence of the constructive conception of real numbers, talk about actually infinite systems is meaningless and functions which are defined on all real numbers are continuous. In classical statistical-mechanical theories two philosophical problems arise: how do these infinite statistical-mechanical systems relate to finite concrete systems, and how do the discontinuous phase transitions in statistical mechanics relate to the continuous phase transitions in concrete systems? No such problems would arise in constructive theories.

One might argue that we do not need to switch to constructive mathematics

[8] This point is also made in Mainwood (2005) and endorsed by Butterfield (2011).

to obtain this de-idealization: one can simply do classical statistical mechanics without the thermodynamic limit. However, as I see it, de-idealizing statistical-mechanical theories of phase transitions by means of switching to constructive mathematics is, so to say, a more *profound* de-idealization than merely doing statistical mechanics without the thermodynamic limit, because constructive mathematics does not idealize real numbers to be infinitely precise, and thus the constructive conception of real numbers – arguably – fits better with measurement practice in science: constructive real numbers reflect the fact that we always have a merely finite amount of information about physical quantities. Functions (which are defined on all reals) in the constructive framework are continuous as a consequence of the constructive conception of the reals. Thus, whereas the classical framework allows for philosophically problematic discontinuous changes, the constructive framework can be said to *rule* these *out* on conceptual grounds.

Importantly, however, it seems *not* to be the case that modeling phase transitions in statistical mechanics by mathematical discontinuities is *per se* philosophically problematic.[9] For, as we have seen in section 2.3, in constructive mathematics we have the option to formulate discontinuous *partial functions*. As explained in that section, we can constructively define a function $f : S \to \mathbb{R}$, where $S = \{x \in \mathbb{R} \mid x < 0 \text{ or } x = 0 \text{ or } x > 0\}$, i.e. an element $x \in S$ is a real number *together with* the information whether x is negative, zero, or positive. In other words, we *restrict* the domain of the function to these values of which we are sure – on the basis of the finite information that we possess – whether they are negative, zero, or positive.

Modeling phase transitions by discontinuous constructive partial functions seems *not* to come with the philosophical problems that come with modeling them by discontinuous classical total functions. For, as we have seen in section 2.3, constructive real numbers, as opposed to classical real numbers, are always only *finitely precise*. So, even though these partial functions are *discontinuous*, these discontinuities *mean something different* than their total counterparts in classical mathematics. Given that constructive real numbers are understood to have only finite precision, we do not commit ourselves to what happens *exactly* at the moment of phase transition. In particular, the discontinuity in the constructive partial function does *not* imply that the system "jumps" from one state into

[9] From here on I will be going beyond what I wrote in van Wierst 2017.

another.

Thus, modeling phase transitions by constructive partial functions seems attractive, given that we do not commit ourselves to what happens exactly at the moment of phase transition. As Bauer puts it, the domain S of constructive real numbers together with the information whether they are negative, zero, or positive, encodes moments in time or places in space where *special things happen*, such as a sudden change of density (personal communication, see also Bauer 2012). Personally, I am not in a position to judge whether modeling phase transitions by such constructive partial functions is feasible at all (it seems more complicated to me than simply developing a constructive – especially, whether they can be built-in the theory from the start, rather than as an afterthought –, but it might be interesting to look into this possibility. Another option that maybe can be embedded more unproblematically, and seems to have similar philosophical benefits, is the following.

A further way not to commit ourselves to what exactly happens at the moment of phase transition, is modeling phase transitions by using *interval arithmetic* – either in classical or in constructive mathematics.[10] In this case, rather than restricting the domain as is done in case of constructive partial functions, we *extend* the codomain to *ranges* of possible values. Thus, in interval arithmetic a function f returns instead of a real x, an interval $[a, b]$ containing x. The set of all intervals $[a, b]$ for the closed interval with endpoints a and b is known as the *interval domain* and is denoted $\mathbb{IR}$:

$$\mathbb{IR} = \{\ [a, b] \mid a, b\ \in\ \mathbb{R} \text{ and } a \leq b\}$$

The ordinary real numbers are embedded in $\mathbb{R}$: every real $x \in \mathbb{R}$ is represented by the interval $[x, x]$ of zero width.

Interestingly, discontinuous functions $f\ :\ \mathbb{R} \to \mathbb{R}$ can be made continuous by letting them map to $\mathbb{IR}$ instead. Consider, for instance, a function $f\ :\ \mathbb{R} \to \mathbb{R}$ that has value -1 for negative numbers and 1 for positive numbers. Now let

[10] I thank Andrej Bauer for making me aware of interval arithmetic. I follow here his explanation (in personal communication).

$f : \mathbb{R} \to \mathbb{IR}$ such that

$$f(x) = \begin{cases} [-1,-1] & \text{for } x < 0 \\ [-1,1] & \text{if } x = 0 \\ [1,1] & \text{for } x > 0 \end{cases}$$

We will not go into the details here, but such an f is continuous. As Bauer explains (personal communication), it is quite natural to take $f(0) = [-1,1]$ as a sort of "non-determinate" value beteween the two phases (negative and positive). Another way to understand the interval value $[-1,1]$ is as "many results are possible, the theory does not predict which one will occur".

Representing the values of f as an interval $[a,b]$ containing x, can be seen as reflecting uncertainty in the knowledge of the exact value of physical quantities arising from *measurement error*, or, as Bauer suggests, from *history-dependent evolution* of systems. For these (and perhaps other) reasons, physicists tend *not to commit* themselves to what exactly happens at the moment of phase transition, and therefore the mathematical model that allows many possible answers at once may appeal to physicists, for it seems to make their intuitions precise.

To recapitulate: discontinuities in the classical theories are problematic in the sense that the mathematics seems to assert something specific and unrealistic about the exact moment of phase transition – and thus about phase transitions and phases themselves. Constructive partial functions do not make this commitment, neither do (classical or constructive) functions in interval arithmetic.

Another issue to consider is the following. From the philosophical literature on phase transitions, one might get the impression that the two problems with statistical-mechanical theories of phase transitions that we distinguished in section 4.2 – i.e. how to interpret the *infinite systems* obtained by taking the thermodynamic limit, and how to interpret the *discontinuities* which are taken to represent phase transitions – are intrinsically connected, or even stronger, that they amount to the same problem. For example, as we have seen in chapter 1, in the philosophical literature on the paradox of phase transitions, it is argued – as a *reductio* – that if the discontinuous changes in the partition function were literally true of concrete systems, that is, if concrete systems indeed display discontinuous changes in their thermodynamic properties at phase transitions, then the atomic theory of matter would be false. For example Norton writes:

> "If the atomic theory of matter is true, then ordinary thermal systems of finitely many components cannot display discontinuous changes in their thermodynamic properties. The changes they manifest are merely so rapid as to be observationally indistinguishable from discontinuous behavior. Indeed, if we could establish that the phase transitions of real substances exhibit these discontinuities, we would have refuted the atomic theory of matter, which holds that ordinary thermal systems are composed of finitely many atoms, molecules or components" (Norton 2012).

In other words, Norton holds that if concrete systems display discontinuous changes in their thermodynamic properties, then they are not composed of finitely many particles (atoms, molecules, components). This, of course, means that these concrete systems would be composed of *infinitely* many particles. Shech makes a similar claim:

> "[I]f systems are composed of finitely many particles [...], then it makes no sense to talk of concrete discontinuities. The notion of concrete discontinuity presupposes that matter is a continuum so that there can be an actual discontinuity." (Shech 2013).

Shech argues that if concrete systems display discontinuous changes, then matter must be a continuum. Note that Shech's claim is, or at least appears, stronger than Norton's: Shech argues that concrete discontinuities would not only imply that the particle number is *infinite*, but that it is a continuum, and thus we should add that it is *uncountably* infinite. Unfortunately, neither Norton nor Shech explains *why*, according to them, discontinuous changes in the thermal properties of concrete systems would presuppose that these systems are composed of infinitely many – or *a fortiori* a continuum – of particles.[11]

This raises the question: What exactly do the infinite systems and the discontinuities have to do with each other? Do discontinuous changes in thermal

[11] Of course, both argue from the reductionist perspective that thermal phenomena can be explained in terms of, or reduced to, the microscopic behavior of these systems (see section 1.1.2), but this assumption *alone* does not suffice to justify the view that discontinuous changes in the thermal properties of concrete systems would presuppose that these systems are composed of infinitely many particles.

properties, or concrete discontinuities, indeed presuppose that these systems are made up of a (countable or uncountable) infinite number of particles?

In order to answer this question, we should first clarify: *in which sense* is a system in the thermodynamic limit *infinite*? In particular, do we obtain a *continuum* of matter by taking the thermodynamic limit?

As we have seen in section 4.1, the thermodynamic limit is taken by letting the particle number N and the volume V of a (finite) statistical-mechanical system "go to infinity", in such a manner that the density $\frac{N}{V}$ remains fixed. From this description, it seems that taking the thermodynamic limits amounts to endlessly adding elements to a finite set (while letting it proportionally expand in volume so that the density remains constant). Moreover, by merely endlessly adding elements to a finite set, we will *not* obtain a continuum, for any countable union of countable sets is countable, and thus strictly smaller than a continuum.[12] So, from the description of the thermodynamic limit as "letting the particle number and volume go to infinity", it seems, conceptually speaking, that we do *not* obtain a set of particles of continuum size.

What can we say about this mathematically? It was shown by Compagner (1989) that the thermodynamic limit is in many contexts mathematically equivalent to – meaning that in calculations it gives the same results as – what he calls the *continuum limit*. In the continuum limit, the particle number N of the system increases to infinity whereas its volume V – and other quantities that refer to the system as a whole, such as the entropy S, the temperature T, the total energy E – remains constant (and thus finite). The size of the individual particles d goes to zero as their number N goes to infinity. Thus, what we obtain in the continuum limit, is an infinite number of infinitely small particles spread over a finite

[12] It is interesting to note that Cantor himself proved about what could be considered a similar physical situation that such a splitting does *not* result in a continuum, i.e. a set of cardinality $\aleph_1$. He wrote in a letter to the psychologist and philosopher Wilhelm Wundt (16 Oct. 1883):

> "If we consider the collection of all organic cells at a given time in our cosmos, which expands itself infinitely in all directions, it is certain that this collection consists of infinitely many individuals; one can therefore state the question regarding the "power" [cardinality, aleph] of this set, and I can prove rigorously that the power in question is the *first* [$\aleph_0$], i.e., it not a greater one."

Cells can be regarded as three-dimensional continuous subdomains of Euclidean space, separated from each other so that they can at most touch each other along their borders; and Cantor proved in his 1882 paper that a set of infinitely many subdomains, having such properties, must be denumerable, i.e., of power $\aleph_0$ (Ferreirós 2004, p. 14).

volume; in the thermodynamic limit, to the contrary, both the number of (finitely sized) particles and the volume they occupy are infinite. As Compagner puts it himself, the continuum limit works on the particles, whereas the thermodynamic limit works on the container (1989, pp. 109-110).

What does a system in the continuum limit look like? Does matter in that limit constitute a continuum? Norton has an argument to the effect that this is *not* the case (2012, pp. 217-218). As he points out, one of the quantities remaining constant in the continuum limit is the volume occupied by matter Nd^3. This means that in the limit, like at any stage approaching it, the system will consist of portions of space occupied by matter surrounded by emptiness. As the number of particles gets bigger, their size as well as the size of the regions of emptiness surrounding them will get smaller and smaller, but at every stage the system consists of regions with matter density unity and regions with matter density zero. The argument is nicely illustrated with a simplified example (after Norton 2012):

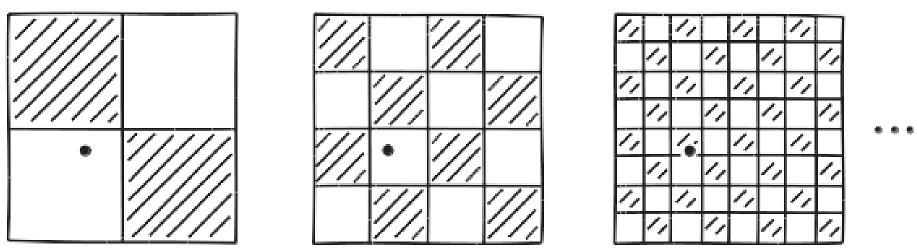

Figure 4.2: Sequence in halftone printing (Norton 2012, p. 218)

Represented in figure 4.2 is a unit square that is divided into half, quarter, eighth squares, and so on, in stages 1, 2, 3, and so on. At each stage, half the squares are occupied by matter – represented by shading -, and half are not. The sequence resembles halftone printing, i.e. a printer simulating grey scales by assigning either black or white to each point on the paper. Points on the diagonal, such as $(\frac{1}{3}, \frac{1}{3})$ will have the state "white" at all stages and thus have "white" as a limiting value. Others will oscillate between black and white indefinitely. For example, the states at the point $(\frac{1}{3}, \frac{2}{5})$ shown in the figure will oscillate indefinitely as white, white, black, black, white, white, black, black, and so on, and therefore do not have a limiting value.[13] Hence, in the limit there will be white points, black points, and

[13] As Norton explains, the rule for computing this series requires that the coordinates be

points with no color. The halftone printing is thus similar to the continuum limit in that in the limit of halftone printing there will be points which are neither black nor white, while in the continuum limit there are points which are neither occupied by matter nor empty. As Norton argues, the halftone printing fails to yield a unit square uniformly covered in a 50% grey tone; the continuum limit fails to return a system with a uniform matter distribution of density $\frac{1}{2}$.

Does Norton's argument show that systems in the continuum limit do not consist of a continuum of matter? Well, that depends on what we take a "continuum of matter" to be. Norton, if I understand him correctly, seems to take a continuum of matter to be a system with a *uniform matter distribution of density* $\frac{1}{2}$ (see above, cfr. Norton 2012, pp. 217-218). On this conception, indeed, Norton's argument shows that in the continuum limit we do *not* obtain a continuum of matter. However, Compagner seems to think about a continuum of matter in a different way. He seems to regard a system in the continuum limit to be a continuum of matter, because *microscopic fluctations disappear* in this limit (cfr. Compagner 1989, p. 108). That is, on Compagner's conception a system in the continuum limit is a continuum because its microscopic discreteness is ignored.[14] Thus, on Compagner's conception the continuum limit *does* result in a continuum of matter. The same seems to hold for systems in the thermodynamic limit: since fluctations disappear and the microscopic structure is neglected, a system in the thermodynamic limit qualifies as a "continuum of matter" in Compagner's sense, but – given that the density in the thermodynamic limit remains constant – in Norton's sense it does not.

Let us try to answer our question: *in which sense* are systems in the thermodynamic limit *infinite*? For Norton a system in the thermodynamic limit is infinite in the sense that it consists of infinitely many particles, but it does *not* form a continuum – where "continuum" is understood as uniform distribution of density. For Compagner, a system in the thermodynamic limit is infinite in the sense that it is mathematically equivalent to the continuum limit in which a continuum of matter is obtained, but Compagner understands a "continuum of matter" in a

expanded as binary numbers: $\frac{1}{3} = 0.010101010...^2$, and $\frac{2}{5} = 0.011001100...^2$. The point is white at the n^{th} stage if the n^{th} digits of the two numbers agree, and black if they disagree.

[14] Cfr.: "The continuum limit emphasizes that macroscopic or classical thermodynamics is only an approximation, which breaks down when the discrete structure of matter is important" (Compagner 1989, p. 111).

different way, namely, in the sense that the microscopic discreteness of the system is ignored. It is clear, however, that neither systems in the thermodynamic limit, nor in the continuum limit, can be considered as a "continuum of matter" as analogous to a mathematical continuum: Norton's argument shows that the matter distribution of systems in the continuum limit does not have the *completeness* property which is essential for a mathematical continuum (see section 3.1).

So let us go on and try to answer our other question: what exactly do the infinite systems and the discontinuities have to do with each other? In particular, would, as Norton and Shech argue in the above quotes, discontinuous changes in the thermal properties of concrete systems, or concrete discontinuities, presuppose that these systems are composed of infinitely many – or *a fortiori* a continuum – of particles?

To address the last issue first: I do not see why it would be the case that discontinuous changes in the thermal properties of concrete systems presuppose that these systems are composed of (countably or uncountably) infinitely many particles. That is, I do not see why discontinuous changes in the configuration of particles would be possible in an infinite system. As I see it, discontinuous jumps in the configuration of particles are a *conceptual impossibility*, in something like the sense of Dummett (see section 3.3), and whether the system consists of finitely or infinitely many particles does not change anything about that.

Of course, it might be objected that there is no ground to hold that discontinuous jumps in the configuration of particles are impossible in infinite systems. After all, *there are no infinite systems* in the real world, so we have no experience with them that could justify such an assertion. We may argue that since there are no infinite systems, we may attribute to them whatever properties we like. And in fact, the mathematics tells us that infinite systems *can* display such discontinuities. So why not just rely on what the mathematics tells us?

The problem with relying on what the mathematics tells us about infinite systems, as I see it, is two-fold. First, it should be clarified *what do we mean* by this mathematics? Which parts of it do we take to represent features of physical systems? To which extent do we take physical systems and the mathematics that we use to represent them to be analogous? For example, what is a "continuum of matter"? In particular, do we take it to mean that matter is so distributed that it is *complete*? Is matter idealized to consist of an *uncountable* infinity of components, just as a (classical) mathematical continuum consists of an uncountable infinity of points? Or do we simply mean that the microlevel discreteness of the system is *ignored*? And another example: in which sense are "physical discontinuities" *discontinuous*? Do they involve instantaneous "jumps" like in the mathematical case? What exactly is it that makes that, as Batterman puts it, they "demand curves with kinks"?

Second, it should be established *which* mathematics exactly are we relying on? For as we have seen, the meaning of the mathematics, and consequently the extent and form in which philosophical problems arise in interpreting it, differs between mathematical systems and methods. For example, the meaning of "discontinuity" differs between discontinuous classical total functions and discontinuous constructive partial functions. Similarly, even within classical mathematics, what "continuity" means depends on whether we use regular methods and have $\mathbb{R}$ as a codomain, or rather we use interval arithmetic and take $\mathbb{IR}$ as a codomain. Depending on the mathematical framework and methods we use, we do or do not idealize systems as actually infinite; we do or do not idealize phase transitions as infinitely sharp; we do or do not commit ourselves to what happens at the exact moment of transition.

In answer to our question what exactly the relationship is between infinite systems and discontinuities, we note the following. Whereas we have infinite systems and discontinuous changes in classical statistical-mechanical theories with the thermodynamic limit, and no infinite systems and discontinuous changes in constructive theories which use total functions, if we use partial functions in constructive mathematics we have discontinuities but no infinite systems, whereas when we use interval arithmetic in classical mathematics we may have infinite systems but no discontinuous functions. In sum, the relationship between infinite systems and discontinuities is *relative to the mathematics in which the physics is formulated.*

For these reasons, it seems to me that we should *not* turn to the mathematics in order to judge, e.g., whether discontinuous changes in thermal properties presuppose that matter is a continuum.

4.4 Conclusion

So: *do* our best theories of phase transitions tell us that no phase transitions can occur in finite systems? Yes and no. If by "phase transition" we mean the idealized concept according to which systems "jump" from one phase to another, according to which the distinction between phase transitions and changes within a phase is clear-cut, and according to which systems are always clearly in one phase or another, then statistical-mechanical theories show that no finite, and thus no concrete, system can display phase transitions. If, to the contrary, by "phase transition" we mean the phenomenon with which we are so familiar, and according to which systems often change gradually from one phase to another, according to which we often cannot tell whether or not a system displays a phase transition, or in which phase – if any – the system is, then statistical-mechanical theories do *not* show that no finite, and thus not concrete, system can display phase transitions. In other words, the answer to this question depends on what we mean by "phase transition".

And: *are* the philosophical problems with statistical-mechanical theories of phase transitions relative to classical mathematics and its notion of actual infinity? Yes and no. Classical mathematics can be seen as the cause of the paradox of phase transitions, in the following sense. First, the paradox arises from certain idealizations which are *made possible* in classical mathematics: infinite systems and discontinuous functions. Second, the problem arises from the classical *meaning* of the mathematics: an "infinite" system means an *actually* infinite system, and – given that real numbers are understood to have infinite precision – discontinuities represent actual "jumps" in the physical quantities. In constructive mathematics, first, there is no possibility to formulate this mathematics and thus to idealize systems in this philosophically problematic way, and second, the meaning of the mathematics – such as e.g. the concept of real number – is less problematic, in that it fits more naturally with scientific practice – constructive real numbers can be understood to reflect the fact that there is always some measurement error. However, modeling phase transitions in classical mathematics is not *necessarily* philosophically problematic. First, there is of course the option of taking "phase transition" in the second meaning in the paragraph above – i.e. as the vague concept of a phase transition such as we see it exhibited in the real world – and modeling them in statistical mechanics without the thermodynamic limit. But there seem to be other options as well, such as using interval arithmetic. In both cases, however, as well as when we opt for adopting (total functions in) constructive mathematics, this means that we have to let go of the idealization of systems as displaying clear-cut, infinitely sharp phase transitions. And this might be too high a price to pay for physicists.

Further, we can take away the following from this chapter. The paradox of phase transitions clearly relates to the issue of *consistency* (see chapter 1): the thermodynamic limit introduces an *internal inconsistency* in order to eliminate an *external inconsistency*. The internal inconsistency introduced by the use of the thermodynamic limit in statistical mechanics, is that systems in the thermodynamic limit are *infinitely large* in particle number and volume, whereas statistical mechanics is (or is supposed to be) a theory of systems which are *finite* in particle number and size (namely, concrete systems). This idealization is introduced in order to undo the external inconsistency between statistical mechanics and thermodynamics: whereas systems in thermodynamics at certain values of, e.g., pressure and volume display discontinuities, systems under the same circumstances in statistical mechanics without the thermodynamic limit do not.

The view that this external inconsistency between thermodynamics and statistical mechanics is *problematic* at all, rests on two assumptions: first, that thermal behavior of systems can be explained in terms of their behavior at the microlevel – or, in other words, that thermodynamics can be reduced to statistical mechanics –, and second, that statistical mechanics, like thermodynamics should denote phase transitions by mathematical discontinuities. The first assumption comes from the view of physics as what Batterman called *P*hysics (see section 1.1.2), according to which microlevel theories such as statistical mechanics are in some sense "more fundamental" than, and accordingly privileged over, macrolevel theories such as thermodynamics. This view is held by the majority of philosophers of science, and can be said to be the main cause of the widespread bewilderment regarding the need for the thermodynamic limit in statistical-mechanical theories of phase transitions.

The second assumption to the effect that statistical mechanics, like thermodynamics, should represent phase transitions by discontinuities is sometimes disputed in the literature (e.g. in Callender 2001). Notably, Batterman argues in favor of this assumption, because in his view statistical mechanics should give a *qualitative* account of phase transitions. Qualitatively understood, Batterman argues, concrete systems display "physical discontinuities" at phase transitions which are in his opinion *rightly represented* by mathematical discontinuities. But our analysis in section 4.2 revealed that *even if* we want a qualitative representation of phase transitions and we accept Batterman's claim that systems display "physical discontinuities" at phase transitions, *then still* it does not seem to be necessary to denote these physical discontinuities by mathematical discontinuities. In particular, these "physical discontinuities" do not seem to be *discontinuous* in any sense of the term that makes that, as Batterman argues, they "demand curves with kinks".

The proposal from my (van Wierst 2017) paper to develop a constructive account of phase transitions should be understood as a more rigorous version of the solutions to the paradox of phase transitions offered by Norton and Pincock: whereas the latter argue to interpret "infinity" as "large but finite" instead of "actually infinite" *merely* when talking about the size of systems, I argue that we should *consistently* interpret it in this manner. That is, I propose to interpret the infinite everywhere in the mathematics as the constructivist's *potential infinity* – not only when talking about systems, but already in the definition of real numbers. Doing so, as we have seen in section 4.3, not only we avoid both conceptual problems with the current theories – that systems are idealized to have an actual infinity of components, whereas the concrete systems they represent have only finitely many components, and that the changes in thermal properties at phase transitions are idealized to be discontinuous, whereas concrete systems can only display continuous changes – but also we obtain a continuum which seems better to reflect the ontology of bounded statistical-mechanical systems: discrete at the microlevel, but in principle without a bound to their size.

Two remarks regarding this proposal are pertinent. First, as already discussed earlier in this conclusion, adopting constructive mathematics is *not the only way* to avoid the philosophical problems with statistical-mechanical theories of phase transitions; there are other options as well. Avoiding the philosophical problems, however, seems to require giving up idealizing systems as displaying clear-cut, infinitely precise phase transitions. Thus, the real problem seems to come not from the classical mathematical framework *per se*, but rather from the particular way in which phase transitions are idealized. In contrast to constructive mathematics, however, classical mathematics does *permit* this kind of idealization. We can even say, I think, that it is very *natural* to idealize in this way in classical mathematics, for real numbers themselves are already idealized as infinitely precise.

Second, adopting constructive mathematics will certainly *not solve all the problems* surrounding statistical-mechanical theories of phase transitions. For first, the paradox of phase transitions relates to much bigger issues such as intertheory relations and the hierarchical picture of science. And second, prior to developing alternative theories it should be established: *what do we mean* by the concepts involved in the paradox, and in particular, to what extent do we take these physical concepts to be analogous to the mathematical concepts we use to represent them?

Chapter 5

Paradox II: Norton's Dome

Newtonian (or: classical) mechanics is often taken to be the paradigm of a deterministic theory: Newton's equations of motion uniquely determine the motions of objects given their initial positions and velocities. However, when a point-mass is placed at the apex of Norton's Dome – a radially symmetric, infinitely rigid surface with a particular shape in a gravitational field -, according to Newtonian mechanics it can remain forever motionless, or after an arbitrary time roll of the Dome in an arbitrary direction. Is Newtonian mechanics not deterministic after all?

5.1 The physics

We will start this section with an exposition of the physics of the Dome as introduced by Norton (2003, 2008). Norton's Dome is a surface as illustrated in figure 5.1 below, which is surrounded by a downward directed gravitational field. It is rotationally symmetric about the origin $r = 0$, which is also the highest point of the Dome. The shape of the Dome is described by the equation

$$h(r) = \frac{2}{3g} r^{\frac{3}{2}} \tag{5.1}$$

where r is the radial distance coordinate in the surface of the Dome (i.e. the distance from the highest point of the Dome along the surface), h is the vertical distance below the apex at $r = 0$, and g is the acceleration due to gravity. A point-like unit mass slides frictionlessly over the surface.

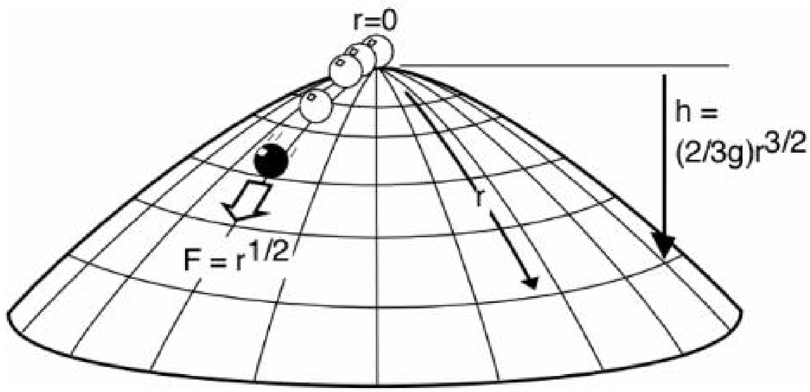

Figure 5.1: Mass sliding on the Dome (from Norton 2008)

Before explaining why, according to Norton, the Dome shows that Newtonian mechanics is indeterministic, we should clarify: why is it usually held that Newtonian mechanics *is* deterministic? Determinism, within the context of Newtonian mechanics, should be taken to mean that the motions of objects – such as the point mass on the Dome – within a Newtonian system are *uniquely determined* by their initial positions and velocities (cfr. e.g. Norton 2003, Norton 2008, Malament 2008, Fletcher 2012). This, in turn, should be taken to mean that Newton's laws of motion, i.e. the differential equations which describe the dynamics of the Newtonian system in terms of its motion as a function of time, given some initial conditions, have a *unique solution* – i.e. there is a unique function describing the position r of the object in terms of the time t such that the initial conditions are satisfied.[1]

Now, Newton's laws of motion give the following picture of the movement of the mass. The net force F acting on the mass is the component of the gravitational force tangent to the surface, i.e. $F = \frac{d(gh)}{dr} = r^{\frac{1}{2}}$ and is directed radially outward. Thus, at any point r on the surface of the Dome, the mass is subject to an outward directed force field of $r^{\frac{1}{2}}$. Since the force acting on the mass is known, Newton's

[1] This uniqueness is guaranteed by a theorem of the theory of ordinary differential equations called the *Picard-Lindelöf theorem*. As we will see in the next subsection, this theorem has a condition – the *Lipschitz-continuity* of F – which in case of the Norton Dome is violated.

second law of motion ("$F = ma$") should be sufficient to describe the motion of the mass. Given that we have to do with a unit mass, it sets the acceleration equal to the force field:

$$a(t) = \frac{d^2r(t)}{dt^2} = r^{\frac{1}{2}} \tag{5.2}$$

If the mass is initially located at rest at the apex $r = 0$, then the initial conditions are the following:

$$r_0 = 0 \tag{5.3}$$

$$\frac{dr}{dt}(0) = v_0 \geq 0 \tag{5.4}$$

5.3 says that the mass is initially located at the apex of the Dome, $r = 0$, and 5.4 that the initial velocity v_0 points in positive r direction (cfr. Malament 2008). Since at the apex there is no tangential component of the gravitational force, i.e. $F = 0$, one obvious solution to 5.2 for all times t is a trivial one:

$$r(t) = 0 \tag{5.5}$$

Equation 5.5 describes the mass simply remaining at rest at the apex for all times t. However, as Norton argues, this is not the only possibility, for also the following large class of unexpected solutions satisfies Newton's second law:

$$r(t) = \begin{cases} \frac{1}{144}(t-T)^4 & \text{for } t \geq T \\ 0 & \text{for } t \leq T \end{cases} \tag{5.6}$$

Note that T here can be any time *whatsoever*. Newtonian mechanics gives no guidance as to what T should be. So what we have is an uncountable infinity of solutions according to which the mass sits at rest at the apex of the Dome for some time T, whereupon it moves off in some arbitrary radial direction. The acceleration in these cases would be:

$$a(t) = \begin{cases} \frac{1}{12}(t-T)^2 & \text{for } t \geq T \\ 0 & \text{for } t \leq T \end{cases} \tag{5.7}$$

We see that also these solutions satisfy Newton's second law, by noting that $a(t)$ as given by 5.7 is the square root of $r(t)$ as given by 5.6, as required by 5.2. Thus, in case of the Norton Dome the motion of the mass is *not* uniquely determined: both 5.5 and 5.6 satisfy conditions 5.3 and 5.4. In other words, there is in Newtonian mechanics no way to determine whether or not the mass will start sliding off the Dome, and if it does, at which time T – and, additionally, in which direction – it will do so. The Dome thus seems to contradict the idea of Newtonian mechanics as a deterministic theory, in that some initial conditions are compatible with many futures. As Norton puts it, Newtonian mechanics allows thus for "spontaneous acceleration" (Norton 2003, Norton 2008).

As Norton points out, one might consider to object that such spontaneous acceleration is prohibited by Newton's first law (in its instantaneous form: "*In the absence of a net external force, a body is unaccelerated*"; cfr. Norton 2003, p. 14). That is, one might argue that given that there is no net force on the mass at $t = T$, by this law the mass should remain at rest. However, as Norton shows, the motions of 5.6 are completely in accordance with Newton's first law: for times $t \leq T$, there is no force applied (since the mass is at position $r = 0$, where the gravitational component of the tangential force $F = 0$) and the mass is unaccelerated; for times $t > T$, there is a net force applied (since the mass is at a position $r > 0$, where $F > 0$), and the mass accelerates in accord with $F = ma$ (ibid.). At the particular time $t = T$, we can see from substitution into 5.6 that the mass is still at the apex and from substitution into 5.7 that it has an acceleration $a(0)$ of zero. Thus, at $t = T$ there is no force and the mass is unaccelerated; at any $t > T$, there is a non-zero force and the mass accelerates accordingly. This, as Norton argues, is exactly what Newton's first law demands.

In sum, Norton holds that his Dome shows that Newtonian mechanics is not a deterministic theory, because there is a Newtonian system, namely the Dome, for which Newton's laws do not determine whether or not the point mass placed on the apex will start to move, and if so, when and in which direction it will do so. Surely, whether or not the Dome shows Newtonian mechanics indeterministic crucially depends on whether or not it is *appropriate* to apply Newton's laws to it – that is, in other words, whether or not the Dome is a proper "Newtonian system". Perhaps surprisingly, what does and what does not count as a proper Newtonian system is not a clear-cut matter, and depends – among other things[2] – on the unsettled issue of which kind of *idealizations* should be admitted within this theory. In the philosophical literature, many have argued that Norton's Dome involves idealizations which are improper for Newtonian systems, and thus that it falls outside the proper "domain of application" of Newtonian mechanics. We will discuss these arguments in the next section.

5.2 The philosophical debate

The common thread in the philosophical literature on Norton's Dome is the thought that one cannot decide for or against the deterministic character of Newtonian mechanics as long as it is not clear what the proper domain of application of this theory is, that is, what counts as a proper "Newtonian system". In particular, it is commonly argued in the literature that it is not clear that the Dome is a case of indeterminism in Newtonian mechanics, since it is not clear that the Dome is a proper Newtonian system. Clearly, the Dome involves idealizations, and whether or not these idealizations are allowed in Newtonian mechanics is up for debate. Surprisingly, different scholars have pointed towards *different* idealizations used in the Dome as being "improper" to Newtonian mechanics. In this section, we will discuss those idealizations identified in the literature as improper which involve, in some way or another, infinity: the geometrical shape of the Dome, the geometrical shape of the mass, and the initial state of the mass on the Dome.

We will start with the geometrical shape of the Dome. As is clear from the previous section, the Dome works as a constraint surface for the particle: it im-

[2] As we will discuss below, Wilson (2009) argues that what is commonly referred to as "classical (or Newtonian) mechanics", actually divides into three different formalisms: point particle mechanics, the physics of rigid bodies and perfect constraints, and continuum mechanics – each of them with descriptive gaps which reach all the way to their cores, and which often give mutually incompatible analyses of the same issue. No combination of these formalisms can fill all the gaps in our physical theorizing, in Wilson's view.

poses conditions on the motion of the point particle and thus on the force F acting on it. There is one point of the Dome which is special: the apex. As we have seen in the previous section, the apex is the only point at which there is no tangential component to the gravitational force, and – consequently – the only point at which the mass remains unaccelerated.

Mathematically, the geometrical shape of the Dome – in particular, at the apex – can be seen as the cause of the indeterminism in the following sense. As we have seen in the previous section, it is usually held that Newtonian mechanics is a deterministic theory, because the motions of objects – such as the point mass on the Dome – in a Newtonian system are uniquely determined by their initial positions and velocities, meaning that Newton's laws of motion, i.e. the differential equations which describe the dynamics of the Newtonian system in terms of its motion as a function of time given some initial conditions have a unique solution. The uniqueness of this solution is guaranteed by the *Picard-Lindelöf theorem* from the theory of ordinary differential equations. This theorem, however, has a *condition*: the differential equations have a unique solution *iff* the function F satisfies a *Lipschitz condition* or is *Lipschitz continuous*.[3] As was already noted by Norton, at the apex of the Dome, F does *not* satisfy the Lipschitz condition (2008, p. 788). Thus, in virtue of its shape, the Dome makes the force F acting on the mass non-Lipschitz, and this, in turn, makes the indeterminism possible.

Should geometrical shapes such as the Dome with its singularity at the apex, which make the forces acting on them non-Lipschitz, be allowed in Newtonian mechanics? Some philosophers argued that non-Lipschitz forces do *not* belong to Newtonian mechanics (e.g. Earman & Friedman 1973). Others, instead, argue that requiring Lipschitz continuity is *ad hoc*, and one cannot rule out these Lipschitz-indeterministic systems without relying on the notion of determinism (Fletcher 2012). Importantly, as Norton pointed out, Lipschitz continuity a *sufficient*, but *not* a *necessary* condition for determinism (Norton 2008, p. 797).

There are good arguments to exclude the Dome as a proper Newtonian system in virtue of its shape at the apex. Malament (2008) has shown that at the apex,

[3] F is *Lipschitz continuous* or satisfies a *Lipschitz condition* on its domain D if and only if there is a constant $L \geq 0$ such that $\forall x, y \in D, |F(x) - F(y)| \leq L|x - y|$. Lipschitz continuity thus expresses that the function has a bounded derivative. The force function F on the Dome is not Lipschitz at the apex: $F' = \frac{1}{2\sqrt{r}}$ at $r = 0$ is infinite, and thus not bounded. For a proof that the Lipschitz condition is violated, see the appendix to Norton (2008).

the Dome surface is once, but not twice differentiable (elsewhere it is infinitely differentiable).[4] This means in particular that the Gaussian curvature of the surface blows up and goes to infinity as one approaches the apex (Malament 2008, p. 2). Importantly, the second derivative of the Dome surface corresponds to the mass' acceleration, and thus this singularity at the apex implies that at that point the acceleration of the particle is not well-defined (Malament 2008, p. 16).[5] Furthermore, Malament showed that this means that a mass placed at the apex of the Dome can *only* stay on the Dome's surface if its initial speed is 0; with positive speed, no matter how small, the mass would fly off the Dome immediately (Malament 2008, section 5). Norton responded to this concern that we could consider the mass to be constrained to the surface of the Dome by a perfectly rigid wire that provides the necessary constraint force to keep the mass on the surface of the Dome (Norton 2008, p. 790). However, according to Stemeroff & Dyer (2016), this would in no way alleviate Malament's concerns, for according to them, the problem that Malament raised stems from the fact that the differential calculus simply cannot be applied on the Norton Dome (2016, p. 22).

According to Stemeroff & Dyer, motion, as construed within any conception of the world based on the differential calculus, is defined to have a certain structure "in the small" (2016, p. 15). This means, in particular, that we can define a differential to the force function that approximates the behavior of the function in the infinitesimal neighbourhood of any point. According to them, the fact that the force function depends on the velocity function, and the velocity function does not have this structure in the small (its derivative w.r.t. time, i.e. the acceleration is not well-defined at the apex), the force function does not have the right structure in the small either (Stemeroff & Dyer 2016, p. 23). Since, as they put it, the differential structure breaks down in the infinitesimal neighbourhood around the apex, Newton's second law cannot be applied there (Stemeroff & Dyer 2016, p. 21). Therefore, as they argue, Newtonian mechanics simply cannot accommodate motion of a mass on the Norton Dome.

We will now proceed with a discussion of another idealization employed in the Norton Dome which has been diagnosed in the literature as improper for Newto-

[4] Note that F has to be twice differentiable in order to satisfy the above Picard-Lindelöf theorem (cfr. Malament 2008, p. 2).

[5] One might argue that it is essential to Newtonian mechanics that we be able to assign accelerations to particles; Malament does not take a stand on this issue (Malament 2008, p. 16).

nian theories: the geometrical shape of the mass. As discussed before, the mass is a point, and therefore has zero extension in space and infinite density. There have been presented in the literature two different reasons why this is problematic: first, because this requires the infinite idealization of the Dome as *infinitely rigid*, and by the slightest relaxation of this idealization the indeterminism disappears, and second, because this gives a *conceptual mismatch* with the extended Dome. We will discuss these views here in turn.

As was pointed out by Korolev (2007), the fact that the mass is an extensionless point means in particular that the force is applied to just one point and thus produces *infinite pressure* on the surface of the apex. In order to compensate for this infinite pressure, the Dome has to be idealized as absolutely non-deformable, i.e. *infinitely rigid*. Without this idealization, Korolev argues, the infinite pressure would cause an infinite well in the Dome, as illustrated in figure 5.2.

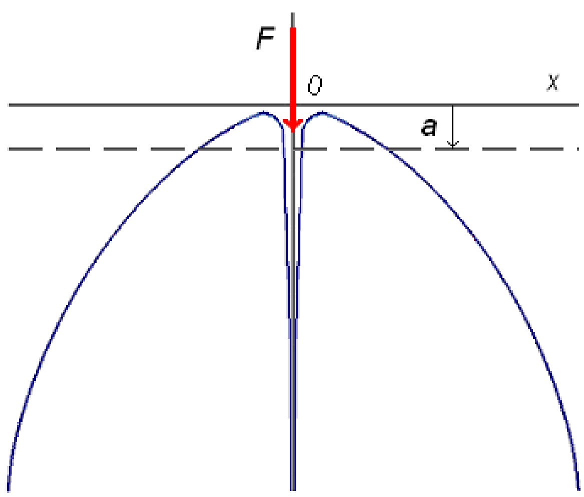

Figure 5.2: The Dome deformed by the point mass (after Korolev 2007)

The idealization of the Dome as infinitely rigid comes down to ignoring *elastic phenomena*, namely the deformation that would occur in the Dome as a response to the force applied to it by the mass. Korolev demonstrates that if some elasticity was taken into account, no matter how small, it would deform sufficiently to prevent the indeterminism. To be more precise, the elasticity destroys the singularity at the apex, which means that the Lipschitz indeterminism disappears. As

Korolev argues, the problem of the point mass on the apex of the Dome "cannot even be set up properly within the classical theory of elasticity" (Korolev 2007, p. 13). Thus, on Korolev's view, the indeterminism of the Dome case is as a result of an infinite idealization (i.e. the idealization of the Dome as infinitely rigid) which makes the Dome *inconsistent* with the theory of elastic phenomena.

According to Wilson, the problem with the point mass on the Norton Dome involves a *conceptual mismatch* between an extended and an unextended object (Wilson 2009, p. 185 n. 11). Wilson argues that what is commonly referred to as "classical (or Newtonian) mechanics", actually divides into three different conceptual frameworks which often give mutually incompatible analyses of the same issue: *mass point particle mechanics*, the *physics of rigid bodies and perfect constraints*, and *continuum mechanics*. Each of these three conceptual frameworks, according to Wilson, has descriptive gaps which reach all the way to their cores; they appeal to each other locally to fill these gaps, but this cannot be done globally, and there will always remain some explanatory gaps in our physical theorizing (Wilson 2009, p. 174). Now, as he explains, unextended point-masses such as the one on the Norton Dome are object of mass point particle mechanics, whereas extended yet perfectly rigid geometrical shapes such as the Dome are object of the physics of rigid bodies (Wilson 2009, p. 176). This is according to him particularly problematic with regard to the force acting on the particle: in Norton's example, the total applied force appears as the constant downward directed gravitational force, which is then apportioned into two sub-categories in different amounts, depending on the position at the Dome; such a decomposition of force is part of the physics of rigid bodies and perfect constraints, but is rejected by mass point particle mechanics (Wilson 2009, pp. 177-178). According to Wilson, this division of forces is how Norton's loss of determinacy secretly enters the scene (Wilson 2009, p. 178).[6] Thus, in Wilson's view the idealization of the mass as an extensionless point-particle is problematic in combination with the extended Dome, for both are covered by different and incompatible parts of classical mechanics.

We will now proceed with discussing the third idealization, the initial conditions: the mass is placed at perfect rest at the center of the apex. As already pointed out by Norton himself, this is an idealization, because it involves ignor-

[6] Roberts (2009) has shown, however, that the decomposition of forces is neither necessary nor sufficient for the indeterminism to obtain.

ing perturbations and is incompatible with quantum mechanics (Norton 2008, p. 793). Wenmackers & Vanpoucke (2016) propose a nonstandard model of the Dome, with (infinitesimal) perturbations.[7] This model is deterministic, and moreover it makes it possible to assign probabilities to the possible values of T in the standard Dome. Given that their contribution is, to the best of my knowledge, the only contribution to the debate on infinite idealizations in which some model based on non-classical (in this case, meaning nonstandard) mathematics is proposed as a solution to philosophical problems, I will discuss their contribution in detail.

As we have seen in section 5.1, in the standard model of the Dome, the time parameter – as every other parameter – is taken to be continuous. Wenmackers & Vanpoucke, to the contrary, take the time parameter to be discrete ($t = n\Delta t$). Given the discrete time parameter, the differential equations of the standard model have to be replaced by their discrete analogue, i.e. *difference equations*. As Wenmackers & Vanpoucke explain, this is standard praxis in numerical analysis, which is often used in physics. Within the scope of standard analysis, the discrete approach using difference equations is only an approximation to the continuous case described by differential equations – an approach that improves as the time step Δt decreases. Wenmackers & Vanpoucke use the discrete approach however within the framework of nonstandard analysis: they define $\Delta t = \frac{t}{N}$ for $N \in {}^*\mathbb{N}$ infinite, i.e. their Δt is infinitesimal, and thus smaller than any strictly positive real number (see section 2.4). Thus, we could say that within the framework of nonstandard analysis, the discrete approach is actualy an *improvement* on the continuous case. In the following, we will see which consequences this has for the (in)determinism of the Dome.

On the discrete approach of Wenmackers & Vanpoucke, the velocity of the mass $\frac{dr}{dt}$ is approximated by the sequence $\frac{R_n - R_{n-1}}{\Delta t}$ and its acceleration $\frac{d^2r}{dt^2}$ by $\frac{R_n - 2R_{n-1} + R_{n-2}}{\Delta t^2}$. This means that $R_n = 2R_{n-1} + \Delta t^2\ R_{n-1}{}^{\frac{1}{2}} - R_{n-2}$.

As we have seen above, the indeterminism arises in the standard Dome when the mass is initially located at rest at the apex. As Wenmackers & Vanpoucke argue, in the nonstandard model this initial condition has two possibile interpre-

[7] My discussion of Wenmackers & Vanpoucke (2016) is based on an extended abstract and presentation slides from 2016, since the related paper is not published at the time of writing of this dissertation. Many thanks to Sylvia Wenmackers for kindly providing me with the abstract and slides.

tations: (1) it could mean that *even with infinitesimal precision* the mass is at rest at the apex, i.e. $r_0 = 0$ even when seen through an "infinitesimal microscope", or (2) it could mean that *merely with standard precision* the mass is at rest at the apex, i.e. $st(r_0) = 0$, but not necessarily $r_0 = 0$ when seen though and "infinitesimal microscope" (see section 2.4). Now, if we go for interpretation (1) of the initial conditions and consider $R_0 = 0$ and $R_1 = 0$, then the solution to the difference equation describing the motion of the mass is unique: it is the constant sequence $R_n = 0$ for all n, which corresponds to the trivial solution $r(t) = 0$ in the continuous case (5.5 above). If we go for interpretation (2) however, and we relax the initial conditions to $R_0 \cong 0$ and $R_1 \cong 0$, i.e. we allow the initial position and velocity to be any infinitesimal, then we obtain for every initial condition a unique solution. For example, take $R_0 = \Delta t$ and $R_1 = R_0$ (so $V_0 = 0$). Then: $R_0 = \Delta t; R_1 = \Delta t;\ R_2 = \Delta t + \Delta t^{\frac{5}{2}}; R_3 = \Delta t + 2\Delta t^{\frac{5}{2}} + \Delta t^2(\Delta t + \Delta t^{\frac{5}{2}})^{\frac{1}{2}};\ \ldots\ .$[8] Thus, since each different choice for R_0 and R_1 results in a unique sequence R_n, the model of Wenmackers & Vanpoucke is deterministic. Further, as Wenmackers & Vanpoucke point out, the standard parts of the results agree with the family of solutions found in the standard model: for small n, R_n is infinitesimal, and for large n, R_n tends to $\frac{1}{144}(n\Delta t - T)^4 + O(\Delta t^2)$, and thus equals the solution of the standard model (see 5.6 in section 5.1 above) plus some infinitesimal $O(\Delta t^2)$. Consequently, for all n, R_n is from a standard perspective indistinguishable from $r(t)$.

In sum, Wenmackers & Vanpoucke showed that if from a nonstandard perspective the mass is initially at rest at the center of the Dome, i.e. $R_0 = 0$, then $r(t) = 0$ is the only solution, i.e. the mass will stay put for all time. If, however, from a standard perspective, but not necessarily from a nonstandard perspective, the mass is initially at rest at the center of the Dome, i.e. $st(R_0) = 0$, but not necessarily $R_0 = 0$, then, for every initial condition, there is a different, unique trajectory. So, Wenmackers & Vanpoucke showed that the indeterminism of the Dome is a model-dependent property.

Besides the fact that there is no indeterminism in the nonstandard model of Wenmackers & Vanpoucke, there is another benefit of the nonstandard approach: it can serve to assign probabilities to the standard solutions. Norton wrote that

[8] Wenmackers & Vanpoucke (2016) wrote $R_2 = \Delta t + \Delta t^{\frac{3}{2}}$ and $R_3 = \Delta t + 2\Delta t^{\frac{3}{2}} + \Delta t^2(\Delta t^2 + \Delta t^{\frac{3}{2}})^{\frac{1}{2}}$, but this seems me to be a mistake.

assigning probabilities to the times at which the particle starts to move might be desirable, but argues that there is no way to do so given the infinite import (Norton 2003, pp. 13-14). This is false: with a nonstandard probability theory it can be done (see Benci et al. 2016). Without going into all the details, we will shortly discuss Wenmackers & Vanpoucke's application of nonstandard probability to the Dome.

Wenmackers & Vanpoucke use a nonstandard probability theory to obtain a probability distribution over the values of the offset T, by assigning a probability distribution to the (infinitesimal) values of the perturbations. As they explain, they assign probability values in a robust manner, i.e. not just assuming a uniform distribution of infinitesimal positions. The relation between initial conditions and offset T values is highly nonlinear, with a higher probability for smaller T-values. The result is that the (hyperreal) probability of $st(T) > 0$ is infinitesimal (with standard part $= 0$), and the (hyperreal) probability of $st(T) = 0$ is one minus an infinitesimal (with standard part $= 1$). In other words, the mass will start rolling of the Dome immediately almost surely; the probability that the mass will stay put for some strictly positive standard time is infinitely small.

In sum, in this section we discussed three idealizations employed in Norton's Dome which are considered problematic in the literature: the geometric shape of the Dome, the geometric shape of the mass, and the initial conditions of the mass being placed at perfect rest exactly at the center of the Dome. It is argued that the geometric shape of the Dome is problematic, because it involves a singularity at the apex, which makes that the force function is non-Lipschitz and that the mass' acceleration is not well-defined. It is argued that the geometric shape of the mass, being point-like, is problematic, first, because as a consequence the mass produces infinite pressure on the surface of the Dome (and the Dome has to be infinitely rigid to compensate for that), and second, because conceptually it does not match with the extensive Dome: the Dome and the mass are covered by different and incompatible parts of classical mechanics. It is argued that the initial conditions are problematic, for they ignore perturbations. In the nonstandard model of the Dome proposed by Wenmackers & Vanpoucke (2016), (infinitesimal) perturbations are taken into account, which makes the model deterministic. Moreover, on the nonstandard approach it is possible to show that the probability that the mass will stay on the Dome for some positive time is infinitely small.

5.3 Analysis

As we have seen in section 5.1, Norton takes his Dome to illustrate that Newtonian mechanics is indeterministic. In section 5.2 we have seen that various scholars rejected Norton's conclusion, because according to them, certain (infinite) idealizations employed in the Dome make that it is not a proper Newtonian system. In the present section, our aim is to analyze to which extent the conclusion that Newtonian mechanics is indeterministic is relative to classical mathematics. We will discuss, first, two kinds of indeterminism that seem to be involved in the Dome case and whether they disappear in alternative (i.e. discrete) models of the Dome, second, what is the justification for using alternative models, and lastly, what the Dome teaches us about how far we may take infinite idealization.

As we have seen in section 5.1, Norton takes his Dome to illustrate that Newtonian mechanics is indeterministic. 'Indeterminism' can mean different things, and it seems that Norton uses the term in two different senses. First, Norton uses it as meaning that a *single past is compatible with multiple futures*. This is the kind of indeterminism that Norton demonstrates with the derivations from Newton's second law, as we have seen in section 5.1: the mass on the Dome stays at rest on the apex until some time $t = 0$, and may or may not move at any time after that. There is nothing in Newtonian mechanics that determines whether or not the mass will move, and if it does, when, and therefore this theory is indeterministic.

But Norton uses the term in a second sense as well. For he takes the Dome case to show that there is *no first cause* which sets the mass in motion (Norton 2003, section 6). As he writes, the mass moves during the time interval $t > T$ only and there is no first instant of motion in this time interval at which to locate the first cause (Norton 2003, p. 25). Indeterminism means here the absence of a first cause of the motion of the mass. Norton locates the indeterminism in this sense specifically at the moment $t = T$: "[t]he failure of causality arises specifically at time $t = T$ when the system spontaneously accelerates. Before and after, the system is quite causal [...]" (Norton 2003, p. 24).

Note, to start, that it is slightly misleading to say (as Norton does in the quote in the previous paragraph) that "at time $t = T$ the system (spontaneously) accelerates". For, as we have seen at the end of section 5.1, at time $t = T$ the mass is at the apex and has an acceleration $a(0)$ of zero, while at any $t > T$ the mass is *not* at the apex and has positive acceleration. This means that the instant $t = T$ is *not* the first instant at which the mass moves; it is the *last* instant at which the mass is at rest (cfr. Norton 2008, p. 788). As Malament (2008) pointed out, the acceleration is not well-defined at the apex. Thus, rather than saying that at $t = T$ the mass accelerates, we should say that at time $t = T$ everything is as it was before, which means that there is no movement and no (well-defined) acceleration.

Further, the problem of indeterminism in the second sense seems to be that there is no first instant of motion during the interval $t > T$ at which to locate the first cause of the motion of the mass (for this interval, being open, has no first instant) *and* there is nothing in the state of the system at $t = T$ which is productive of the acceleration (for this state is identical to the states at all $t < T$ at which the mass is at rest at the apex). Thus, the indeterminism or "spontaneous" movement of the mass seems to mean here that there is no external intervention or change in the physical environment which causes the movement, i.e. the movement of the mass is 'uncaused' (cfr. Norton 2003, p. 12). The complaint thus seems to be that the movement of the mass at time $t > T$ is – in some sense – *disconnected* from the state of the system at all times $t \leq T$.

That these two uses of "spontaneous" movement are not the same and do not necessarily come together can be seen by comparing the movement of the mass on the Dome with that of a mass on a surface where the force acting on the mass turns on smoothly from zero to non-zero magnitude. As pointed out by Zinkernagel (2010, p. 9), in such cases, too, $F = 0$ up to some moment $t = T$, and $F > 0$ for $t > T$, i.e. the system does not continue in its original state for any $t > T$ even though $F = 0$ for $t \leq T$ (see also Fletcher 2012). Also in this case there is no first instant of motion during the interval $t > T$ at which to locate the first cause of the motion of the mass, and there is nothing in the state of the system at $t = T$ which is productive of the acceleration. Thus, in case of smooth forces, there is indeterminism in the second sense, but *not* in the first sense: a single past implies a single future (that is to say, the equations of motion given some initial conditions have a unique solution), but there is "spontaneous" movement in the sense that there is no instant at which to locate the first cause of the motion.

The complaint that the movement of the mass at time $t > T$ is disconnected from the state of the system at all times $t \leq T$, reminds of the critique of Dummett that the state of his 'supertask' pendulum at time $t < 1$ does not tell us anything about its state at instant $t = 1$ (Dummett 2000, see section 3.3 in this thesis). According to Dummett, the fault lies with the classical model of time, in which time is made up of extensionless instants. As Dummett argues, modeling time as a classical continuum makes it possible to model physical systems such as his pendulum and the mass on the Dome in such a way that its state at some $t < 1$ does not tell us anything about its state at $t = 1$. This, in Dummett's view, goes against our *conception* of physical quantities and events: "we do not suppose that events are as loose and separate as this" (Dummett 2000, p. 504). The problem with the indeterminism in the second sense in case of Norton's Dome, as well as in case of smooth forces, seems exactly to be that the independence of the state of the system at $t = 1$ from its state at $t < 1$ goes against our conception of movement and forces.

There is no indeterminism in the first sense in *discrete* models of the Dome. Zinkernagel (2010) considers a (standard) discrete approach (also known as *Euler's method* to approximate the solution to differential equations): he takes time step size $h = \Delta t$ so that $t_n = nh$, and defines the position of the mass $r(t + \Delta t) = r(t) + v(t)\Delta t$ and the velocity $v(t + \Delta t) = v(t) + r(t)^{\frac{1}{2}}\Delta t$, with initial conditions $r(0) = v(0) = 0$. The difference equation for the mass on the Dome is found to be $(r(t_{n+2}) - 2r(t_{n+1}) + r(t_n))/h^2 = (r(t_n))^{\frac{1}{2}}$.[9] Since, as Zinkernagel writes, $r(t_0) = r(t_1) = 0$, this equation has the unique solution $r(t_n) = 0$ for all n (and thus the same solution in the limit $h \to 0$).

Interestingly, not only the Norton Dome under Euler's method – which is considered to be only an *approximation* to the continuous case – but also the nonstandard Dome – which, as we have seen in section 5.2, is in some sense an *improvement* on the standard continuous model – is deterministic in the sense that the solutions to the equations of motion are unique. Further, in the discrete case, the mass will *not* move, *unless* there is a force acting at the beginning of each time segment, and thus there is also no indeterminism in the sense that the movement seems 'uncaused'. This confirms the hypothesis that the paradox of Norton's Dome is relative to classical mathematics.[10]

---

[9] There are many ways to discretize a differential equation, so this difference equation for the Dome is not unique (Zinkernagel 2010, p. 9 n. 12).

[10] It has been pointed out by Waaldijk (2005, p. 24), that in many discussions in modern physics, especially in the one on determinism, it is taken for granted, without explanation, that

Importantly, as Zinkernagel (2010) points out, in standard examples of Newtonian systems involving continuously varying forces the physical situation may just as well be described via a sequence of discrete forces. That is, in such cases modeling a continuously varying force and a sequence of discrete forces is *equivalent*. This equivalence is *absent* in the Dome case – which already can be seen from the fact that difference equations, as opposed to the differential equations, are deterministic. Some have argued that the absence of equivalence between differential equations and their approximations in terms of Euler's method in case of non-Lipschitz continuous functions is a *failure* of Euler's method (Wilson 2006, pp. 213-217; Fletcher 2012, p. 11). However, we could also argue for the contrary: the fact that non-Lipschitz continuous functions behave erratically under Euler's methods can be seen as a *virtue* of Euler's methods. For, as we have seen above, as was pointed out by Stemeroff & Dyer, in case of non-Lipschitz continuous functions the differential structure breaks down. If we accept (the well-definedness of) the differential calculus as a prerequisite for (appropriate) application of Newtonian mechanics, as Stemeroff & Dyer do, then we could say that the inequivalence of differential and difference equations help us to distinguish those systems that are covered by Newtonian mechanics from those that are not.

One point that deserves further attention is the *justification* of using discrete models instead of the standard continuous model. Wenmackers & Vanpoucke (2016) justify their NSA approach to the Dome by pointing out that, since any physical measurement is finitely precise, it is not possible to experimentally distinguish between zero and infinitesimal quantities. However, as we have seen in section 3, since any physical measurement is finitely precise, it is also not possible to experimentally identify a specific real number as the value of a physical quantity, for classical real numbers are *infinitely precise*. As we have seen in section 3.3, Dummett objected against modeling physical quantities by classical real numbers since it amounts in his view to adopting *super-realism*: it posits states of affairs independently even of the theoretical possibility of us knowing them.

Now, as we discussed in section 5.1 above, in the nonstandard Dome of Wenmackers & Vanpoucke, by relaxing the initial conditions to $R_0 \cong 0$ and $R_1 \cong 0$ (allowing the initial position and velocity to be any infinitesimal), one makes room

classical mathematics is the only means available to describe the real world. As he notes, for example the influential book Earman (1986) is guilty of making this tacit assumption.

for perturbations. These perturbations are *infinitesimal* (with standard part zero) and thus can never be empirically detected. Wenmackers & Vanpoucke argue that given that the nonstandard model is empirically indistinguishable from the standard model of the Dome, *we might as well* adopt the nonstandard model to obtain the desired result (determinism). However, if empirical indistinguishability is the criterion for which model (and which initial conditions) are admissible, then other – one might say: less extreme – methods would suffice. For instance, one could adopt, instead of infinitesimals in the sense of NSA, i.e. hyperreal numbers which are smaller than any strictly positive classical real number, numbers which are only *relatively infinitesimal*, i.e. numbers which are strictly smaller than the measurement error.[11] That is, in the standard continuous model, as well as in the standard discrete model, one could adopt initial positions and velocities smaller than the measurement error and as such obtain determinism. Thus, it seems that it is not the discrete (as opposed to continuous) model *per se* which solves the issue of indeterminism in the Norton Dome, but rather undoing the *idealization* of infinitely precise initial conditions (i.e. the mass being placed at perfect rest exactly at $r = 0$).

However, it seems that Norton considers the idealization of the mass being placed at perfect rest exactly at the apex as an essential feature of his Dome, and thus the possibility to make such idealizations as an essential feature of Newtonian mechanics. It is in his view not relevant that these initial conditions cannot be realized or measured in the physical world. As he writes:

> "The dome is not intended to represent a real physical system. The dome is purely an idealization within Newtonian theory. On our best understanding of the world, there can be no such system. For an essential part of the setup is to locate the mass exactly at the apex of the dome and exactly at rest. Quantum mechanics assures us that cannot be done. What the dome illustrates is indeterminism within Newtonian theory in an idealized system that we do not expect to be realized in the world" (Norton 2008, p. 793).[12]

[11] It has been argued that this is the sense in which physicists use the term "infinitesimal" anyway.

[12] It is curious in this context that elsewhere Norton calls the Dome a "plausible physical instantiation" of a violation of a Lipschitz condition (Norton 2008, p. 788, n. 1). Apparently, according to Norton a plausible physical instantiation is not something which can be realized in

In other words, Norton is concerned with the *mathematical structure* of Newtonian mechanics, not with how the theory applies to the physical world.

It should also be mentioned that not only the initial conditions cannot be realized in the physical world: the same applies to the makeup of the Dome and the mass. As Norton writes himself (2008, p. 795), we can consider the Dome with its apex with infinite curvature as the limit of a sequence of domes with apices with finite curvature. Likewise, according to Norton, the mass is the limit of successively smaller, perfectly spherical masses of correspondingly greater density. Both mass and Dome are thus idealizations recovered in the limit of more realistic structures, and these limit systems themselves cannot exist in the real world.

Importantly, as Norton writes, on any dome up to the limit, i.e. on any dome with finite curvature at the apex, the mass will stay indefinitely according to Newtonian theory. Recalling section 1.2.2, one might thus argue that the Dome with its infinite curvature and the extensionless mass does not qualify as a *legitimate* idealization. For, as we saw in that section, Norton argued in his (2012, 2014) papers on idealization and approximation that a system is a legitimate idealization *only if* the limiting process results in a system that is consistent *and has the intended properties.*[13] Since the indeterminism will not obtain for any (more) realistic dome and mass, we have a case in which limit property and the corresponding property of the limit system *disagree*: all systems upto the limit, i.e. all domes with apices with finite curvature, are deterministic, whereas the limit system, i.e. the Dome with its infinite curvature, is indeterministic. For this reason, according to Norton's own account, in the Dome case we do *not* have a legitimate idealization.

Should we discard the Dome on these grounds? Norton himself argues that the infinite curvature does *not* make the Dome an inadmissible idealization: it is "no reason for discounting the system approached in the limit" (Norton 2008, p. 796). We should just acknowledge, according to him, that the Dome with its infinite curvature is a different case with qualitatively different properties that cannot be recovered by taking limits of the properties of more realistic systems; this, in his view, is in particular *no* reason to infer that the Dome is not a Newtonian

the physical world.

[13] Admittedly, these papers on idealization and approximation were published several years after his papers on the Dome (2003, 2008), so maybe Norton changed his views in the mean time.

system (ibid.). Indeed, with respect to the question whether or not Newtonian mechanics is indeterministic, whether or not the Dome is a legitimate idealization according to the criteria presented in section 1.2.2 is not relevant; the relevant question is whether or not it is a proper Newtonian system. For if it is a proper Newtonian system, then Newton's laws should govern the motion of the point mass, no matter how the system is obtained.

However, as I see it, this does raise some questions about how far we may want to take idealizations. In virtue of the idealizations employed in the Dome, e.g. the infinite curvature at the apex, and the extensionless mass, the Dome is a *purely mathematical object.* But from a *mathematical* point of view the non-uniqueness of solutions to a non-Lipschitz differential equation is *completely trivial.* Physicists are not interested in such cases: in physics textbooks one simply disregards non-Lipschitz forces (Wenmackers & Vanpoucke 2016).

As a mathematical structure, Newtonian mechanics shows what is *mathematically* possible within that theory. One is of course free to decide, as Norton does, that that is what one is interested in. However, both the indeterminism and the singular solution of the mass staying on the Dome forever, are only (*mathematically*) possible *in virtue of idealization*: de-idealizing by taking into account perturbations, elastic phenomena, extension of the mass and apex, etcetera, would rule out the singular solution and the indeterminism. The question is: should we talk about "domes" and "masses", if we idealize them to be infinitely small, infinitely dense, infinitely rigid, etcetera? It seems a mistake to identify, for example, the apex of the Dome with the apex of some concrete dome, for it has completely different properties (e.g. infinite curvature, and no spatial extension). All we know about the Dome and other idealized – e.g. Lipschitz indeterministic, infinitely rigid, infinitely small, etcetera – systems we know mathematically, so, as I see it, we should not be surprised by the Dome's indeterminism.

5.4 Conclusion

So: *is* Newtonian mechanics indeterministic? From a certain perspective, Newtonian mechanics is indeed indeterministic. That is, it is the case that there are systems for which Newton's second law does *not* uniquely determine the future evolution of the system given certain initial conditions; if one is of the opinion that these systems fall within the proper domain of the theory, then one can say that Newtonian mechanics is indeterministic. However, as we have seen, several idealizations are *necessary* in order to obtain the indeterminism, and there are good arguments to the effect that these particular idealizations are not admissible in Newtonian mechanics.

And: *are* the philosophical problems with Norton's Dome relative to classical mathematics and its notion of actual infinity? To a certain extent. Several of the idealizations which are necessary to obtain the indeterminism are characteristic of classical mathematics and its notion of infinity – or, to be more precise, its notion of the continuum. For example, the infinitely small, extensionless apex and mass. Other idealizations necessary to obtain the indeterminism, such as the absence of elastic phenomena and perturbations do not seem to have anything to do with the classical mathematical framework.

To some extent, we can say that the philosophical problems with Norton's Dome are relative to *continuous time* models. For as we have seen, both the Norton Dome under Euler's method – which is considered to be only an *approximation* to the continuous case – and the nonstandard Dome – which is in some sense an *improvement* on the standard continuous model – are deterministic. Both discrete models can be seen as undoing the idealization of the infinitely precise initial conditions in the continuous case, i.e. of the mass being completely at rest exactly at the center of the apex.

But even though the indeterminism of the Dome can be countered by de-idealizing the infinitely small apex and mass and the infinitely precise initial conditions, these idealizations taken together are not sufficient to obtain the indeterminism. As we have seen, a sufficient condition for the indeterminism is the force acting on the mass on the apex being *non-Lipschitz* (which in turn is a consequence of the shape of the Dome). It has been argued in the literature that requiring Lipschitz continuity in Newtonian mechanics is *ad hoc*, but in my view there are good arguments to the effect that this is not so. Namely, with the failure of Lipschitz continuity comes the ill-definedness of the mass' acceleration on the apex, or as Stemeroff & Dyer (2016) put it, the break-down of the differential structure. As I see it, a well-defined differential structure is a very natural requirement to make in Newtonian mechanics (at least to the extent that basic concepts such as acceleration are well-defined), and thus I agree with Stemeroff

& Dyer (2016) that a well-defined differential structure is a prerequisite for the appropriate application of Newtonian mechanics. Another way to obtain – as far as I can see – the same result, is to require the equivalence of the differential equations and the difference equations. On this basis, we can dismiss the Dome as a Newtonian system, and the paradox is dissolved.

Further, we can take away the following from this chapter. We have seen that there are two kinds of indeterminism involved in the Dome: first, a single past is compatible with multiple futures (for given the point mass at rest at the apex of the Dome, there are multiple solutions to the equations of motion), and second, if the mass moves there is no instant at which to locate the first cause of the movement. The first kind of indeterminism is particular for the Dome (and some other systems, e.g. those which constrain the force working on them to be non-Lipschitz), the second kind of indeterminism arises also in classical continuous models in which the force turns on smoothly from zero to some non-zero value. In a discrete model both kinds of indeterminism are absent.

There is something ironic about this. For we could say that on both discrete approaches we are in a situation in which we (acknowledge to) have *incomplete data*: the standard discrete approach captures that (our measurements of) time and its derivatives are only finitely precise, and the nonstandard approach captures that there might be perturbations which from an empirical or standard perspective are undetectable. The continuous approach, to the contrary, idealizes the initial conditions and our knowledge of them to be infinitely precise. Thus, the continuous model gives us the *illusion of having the complete data* necessary to compute the evolution of the system. But, as we have seen, we obtain complete information about the time-evolution of the system – namely, a deterministic model – *only* on the discrete approach and *not* on the continuous approach. Thus, from incomplete data we obtain complete information about the time-evolution of the system, whereas from complete data we obtain incomplete information about the system's time evolution.

Further, some of the idealizations in the Dome which are necessary for the indeterminism to occur involve *external inconsistencies*. Namely, the idealization of the Dome as infinitely rigid makes the Dome inconsistent with the theory of elastic phenomena – or with atomism in general. Also the geometrial shape of the Dome as completely continuous, and the apex consisting of an extensionless point, make the Dome inconsistent with atomism. The idealization of the mass initially being at perfect rest exactly at the apex is inconsistent with quantum mechanics, for it ignores perturbations.

Importantly, we have seen that in virtue of the idealizations employed, the Dome is *essentially* a *mathematical object*, for it has properties that no concrete object can have. Given that the Dome is a mathematical object, we should not be very surprised by, nor very interested in, the Dome's indeterminism in my opinion. The most relevant question that the Dome raises is, as I see it: How far do we want to take idealization? Should we talk about "domes" and "masses", if we idealize them to be infinitely small, infinitely dense, infinitely rigid, etcetera?

Conclusion to Part III

I think that what is truly infinite may just be the abyss of our knowledge.
- Rovelli (2011)

The infinite idealizations in the two case studies undertaken in this thesis play a very different role and introduce very different philosophical problems. In statistical-mechanical theories of phase transitions, infinite idealizations – in the form of the *thermodynamic limit* – are introduced in order to obtain the discontinuities in the partition function, which are taken to represent phase transitions. The problem with the thermodynamic limit is that it attributes (1) discontinuous changes at the microlevel and (2) an infinite particle number and volume to finite systems, which is taken to be impossible in virtue of the atomic theory of matter. In Norton's Dome, infinite idealizations – in the form of modeling the Dome as continuous, infinitely rigid, and with infinite curavature at the apex, and modeling the mass as infinitely small, and located precisely at the center at the apex in perfect rest – are introduced in order to show that Newtonian mechanics is not a deterministic theory. The problem in case of the Norton Dome is that Newtonian mechanics is usually taken to be the paradigm of a deterministic theory, but the Dome seems to constutite a very simple counterexample to the determinism of this theory.

In both cases, the infinite idealizations lead to *qualitative* changes in the mathematics, and consequently in the physical model that this mathematics embodies. This is most clearly the case with statistical-mechanical theories of phase transitions, where these infinite idealizations are obtained by taking the thermodynamic limit. The thermodynamic limit introduces, to start, *mathematical discontinuities* in the partition function. The partition function represents the system's microlevel properties, and consequently the discontinuities in the partition function represent *discontinuities in the microlevel properties* of the system. In turn, these discontinuities in the microlevel properties introduce a *clear-cut distinction* both between different phases of the system, and between phase transitions and changes within a phase. Furthermore, the thermodynamic limit introduces an *infinite particle number and volume*, which in turn implies the *absence of borders* and of *fluctations*.

In case of the Norton Dome, the situation is a bit more complicated. Modeling the Dome as *continuous* as well as the particular *shape* that is attributed to it, make that the *curvature* at the apex diverges, which in turn enables the force acting on the mass on the apex to fail the *Lipschitz-continuity* condition for determinism. Further, the *initial conditions* for the equation of motion of the mass being place exactly at rest at the center of the apex, are made possible by the idealization of the Dome as being *infinitely rigid*, by *ignoring perturbations*, as well as by the *continuous time* model. Taken together, these infinite idealizations lead to a model that is *indeterministic*, and thus in this respect qualitatively different from more realistic structures.

As we have seen in chapter 1, the introduction of infinite idealizations into a theory often raise issues of *consistency*. We distinguished between two different kinds of inconsistency, namely, what we called *internal* and *external* inconsistency. The first regards inconsistency between different *pieces of theory*, e.g. *within* the theory of water waves there are two contradicting propositions: one asserting a finitely deep, and one an infinitely deep ocean. The second regards inconsistencies between *different theories*, e.g. continuum mechanics being inconsistent with atomism. What role did inconsistency play in the philosophical paradoxes in our case studies?

As we have discussed in the conclusion of chapter 4, the thermodynamic limit introduces an *internal inconsistency* in order to eliminate an *external inconsistency*. That is, the thermodynamic limit in statistical mechanics (apparently) asserts that the systems under consideration are infinitely large in particle number and volume, whereas statistical mechanics is set out to be a theory of systems which are finite in particle number and size (namely, concrete systems). In this way, the thermodynamic limit introduces an inconsistency *within* statistical mechanics. But the thermodynamic limit is taken in order to *undo* the inconsistency between statistical mechanics and thermodynamics: since statistical mechanics is supposed to *explain* thermodynamic phenomena, or to be a theoretical foundation for thermodynamics, it is supposed that whenever thermodynamic systems display phase transitions, statistical mechanical systems should do so too. On the assumption that phase transitions should have the same mathematical representation in both theories, this means that it is supposed that whenever thermodynamic systems display a discontinuity in the derivatives of the free energy, statistical mechanical systems should display a discontinuity in the partition function. Note that these discontinuities are *only* problematic in statistical mechanics and *not* in thermodynamics, because the former is a *microlevel* theory, whereas the latter is a *macrolevel* theory, meaning that thermodynamics is concerned with the *phenomena* and *explicitly ignores* the microscopic make-up of the systems under consideration.

In case of Norton's Dome, the infinite idealizations are *not* introduced, as in the case of phase transitions, to undo an external inconsistency. Rather, the infinite idealizations are introduced in order to obtain a certain result: the indeterminacy. Many of the idealizations involved in the Dome *introduce external inconsistencies*: the idealization of the Dome as infinitely rigid makes the Dome inconsistent with the theory of elastic phenomena, or with atomism in general; the geometrial shape of the Dome as completely continuous, with an apex consisting of an extensionless point and with infinite curvature make the Dome inconsistent with atomism; and the idealization of the mass initially being at perfect rest exactly at the apex is inconsistent with quantum mechanics, for it ignores perturbations.

It seems – at least, *prima facie* – that in case of the Norton Dome, the philosophically problematic features of the Dome – including its indeterminism – will disappear if external consistency is required and the Dome is de-idealized accordingly. In case of phase transitions, however, de-idealizing by means of requiring internal consistency does *not* seem to solve all the problems, for without the introduction of the thermodynamic limit, systems in statistical mechanics seem *incapable of displaying phase transitions*. Of course, the latter is true only *on the assumption* that phase transitions in statistical mechanics should be represented by mathematical discontinuities.

To a certain extent, both paradoxes arise from issues concerning mathematical *representation*. This is most obvious in the case of phase transitions, for the assumption that statistical mechanics should follow thermodynamics in representing phase transitions by mathematical discontinuities is problematic. Especially given the fact that statistical mechanics is a *microlevel* theory which analyzes system's behavior in terms of statistical and mechanical microlevel properties, whereas thermodynamics is a *macrolevel* theory which deals with phenomena and explicitly ignores what happens at the microlevel, it is not at all clear why these two different theories should conform in their mathematical representation of phase transitions. One might argue that similar mathematical representation is necessary for the reason that statistical mechanics should serve as a theoretical foundation for thermodynamics, but it is unclear to me that if thermodynamics and statistical mechanics would represent phase transitions in a different way, this would threathen statistical mechanics' foundational role.

The issue of representation in case of Norton's Dome is that the Dome is idealized to such an extent that it is very far from realizable in the physical world. This concerns not only the idealization of the Dome as a continuum and of the mass as a point, but also the disregard of elastic phenomena and perturbations, as well as the infinite curvature at the apex. Thus, whereas above we wrote that *prima facie* it seems that in case of the Norton Dome, the philosophically problematic features disappear if the Dome is de-idealized, one might say that the philosophical problems already disappear by recognizing that the Dome is an *essentially mathematical object*. That is to say, *mathematically* there is nothing suprising nor worrisome about the non-uniqueness of solutions to the differential equation describing the motion of the "mass" on the "Dome". The indeterminism of the model starts to be disturbing only when we give the model a *physical interpretation*. In which sense can we call the extensionless highest point of the Dome an "apex"? In which sense should we identify the motions ascribed to the massive point on the Dome with the movement of a physical body?

However, Norton is right in arguing that it is not so easy to discard the Dome as an "unphysical" system. For although the notion of physicality seems to play a strong role for physicists in their intuitions regarding which systems are interesting to study, a precise definition or explication of this concept still has not been given. Norton (2008) presents various notions of unphysicality according to which the Dome does *not* qualify as an "unphysical" system. However, given that no physicist seems bothered by the indeterminism of the Dome, it seems that for them the Dome actually is a unphysical system – if not uninteresting on other grounds.

The problem of representation or applicability arose also the other way around. That is, whereas the issue we just discussed concerns attributing a physical interpretation to a mathematical model, issues arise as well concerning attributing a mathematical interpretation to certain physical models and their properties. For as we have seen, in the philosophical literature it was argued that physical systems displaying discontinuous changes in their microscopic properties would presuppose matter to be a continuum. However, it was not clear at all, first, what a "continuum of matter" was taken to be, and second, how matter being a continuum would be presupposed by these discontinuous changes. It seemed that from the fact that only real-valued functions could display discontinuous changes, it was inferred that only microlevel properties of a continuum of matter could display discontinuous changes. It was not clarified at all in which sense a continuum of matter was supposed to resemble a mathematical continuum, nor in which way mathematical discontinuities were supposed to resemble the hypothetical discontinuities in the microscopic properties of physical systems. Furthermore, it seems to have gone unnoticed that the thermodynamic limit – at least, as it is normally described as "letting the particle number and volume of the system go to infinity" – does *not* result in an *uncountable* infinity of particles, nor seems matter in the thermodynamic limit to be – in some appropriate sense – *complete*. Thus, the extent to which such a "continuum of matter" would resemble a mathematical continuum, and thus to which extent the possibility of discontinuities obtaining in the latter say something about the possibilities of discontinuities obtaining in the first, remains a mystery.

Another example of the same issue is Batterman's argument to the effect that the phenomenon of a phase transition is a "physical discontinuity" and for this reason "demands curves with kinks" as their rightful representation. As we have seen, Batterman fails to clarify *exactly in which sense* these "physical discontinuities" are *discontinuous*, that is, what property do they have that justifies the analogy with mathematical discontinuities?

So, how does this all bear on our **central question**: *Are* the philosophical issues surrounding infinite idealizations in physics relative to classical mathematics and its notion of infinity?

In both case studies, the philosphical problems have proven to be model-depend in some sense. In case study 1, the philosophically problematic discontinuities and infinite systems would be absent in a constructive statistical-mechanical theory of phase transitions; in case study 2, the indeterminism disappears in a discrete – either standard or nonstandard – model of the Dome.

However, it is certainly *not* true that classical mathematics is the *complete* cause of the paradoxes. In case study 1, the philosophically problematic idealization – the *thermodynamic limit* – is introduced, because for reasons connected with intertheory relations, one requires that phase transitions have the same *mathematical signature* in statistical mechanics as in thermodynamics: mathematical discontinuities. Given that statistical mechanics is a microlevel theory, such discontinuities cause inconsistencies within statistical mechanics. Thus, more than the classical mathematical framework, the requirement to represent phase transitions by mathematical discontinuities causes the paradox. In case study 2, the cause of the paradox might be attributed to one or more of the idealizations employed in the Dome which are necessary to obtain the indeterminism, or, alternatively, to the fact that we give a physical interpretation to a mathematical object which is far from realizable in the physical world.

In both case studies, the philosophical problems could easily be solved *without* adopting an alternative mathematical framework. In case of statistical-mechanical theories of phase transitions, the most natural way to avoid the philosophical problems is to assign a different mathematical signature to phase transitions – such as for example sufficiently steep gradients of the relevant functions – so that there is no need to take the thermodynamic limit. In case of Norton's Dome, the most natural way seems to be to require a well-defined differential structure, at least to the extent that acceleration is well-defined on the whole surface of the Dome. In case of phase transitions, avoiding the paradox by means of avoiding to take the thermodynamic limit has the consequence that there will be some *vagueness* re-introduced into the theory, for the distinction between different phases as well as the distinction between phase transitions and changes within a phase will no longer be clear-cut; this can be considered a drawback if one holds that a theory of sharp, unambiguous phase transitions is for some reason better than a theory involving somewhat vague phase transitions, or it can be considered a benefit, because such a theory will model concrete phase transitions more realistically. In case of Norton's

Dome, the consequence of requiring a well-defined differential structure implies a loss of generality; this can be considered a drawback if one is of the opinion that the domain of Newtonian mechanics should be as general as possible, or it can be considered a benefit because it rules out e.g. Lipschitz-indeterminate systems, for which there are good reasons to qualify them as "unphysical".

However, in both case studies, switching to a different mathematical framework has benefits which are not obtained with the above ways to avoid the paradoxes within classical mathematics. In case study 1, developing a constructive statistical-mechanical theory of phase transitions has the benefit that the *meaning* of the mathematics suits better to scientific practice than classical mathematics. Constructive real numbers, as we have seen, are only finitely precise, and therefore in constructive mathematics – in contradistinction to classical mathematics – we do *not* idealize physical quantities and our measurements of them to have infinite precision. In Dummett's terms, by adopting constructive mathematics we avoid the *super-realist* metaphysics which comes with the classical model. In case study 2, adopting a – standard or nonstandard – discrete model of the Dome, has several benefits as well. First, on the discrete approach we de-idealize the infinitely precise initial conditions and allow for perturbations. Second, in the discrete case, the mass will not move unless there is a force acting at the beginning of each time segment, and thus there is also no indeterminism in the sense that the movement seems 'uncaused'. Third, the nonstandard discrete approach makes it possible to assign probabilities to the different time-evolutions of the mass on the Dome, from which it can be shown that the mass staying at rest at the apex for some positive amount of time is *infinitely unlikely*.

What should we make of this result? What should we take it to mean that the philosophical problems in our two case studies are to a certain extent relative to classical mathematics?

As I see it, it is important to realize that scientifically applicable mathematics does not *need* to be built on or justified by a set-theoretical foundation. Much more modest means suffice, such as constructive, predicative, or even finitist mathematics. Moreover, we do not need to attribute the set-theoretical *meaning* to the concepts involved in the physics. Talk of infinitely big systems does not need to be interpreted as concerning *actually* infinite systems, real numbers do not need to be interpreted as infinitely precise, etcetera. In different foundational schemes, mathematical concepts have a slightly different meaning, which can be used to philosophical advantage in certain contexts of application. For example, the idea that all science is just *approximation*, can already be built-in at the level of the mathematical framework in which the theory is formulated.

To finish, how about the indispensability argument? Should we go along with Quine and Putnam and regard the empirical confirmation of our best physical theories which employ mathematics based on set theory as an empirical confirmation for (the relevant parts of) set theory itself? The fact that, as we have seen, scientifically applicable mathematics is *not necessarily* founded on set theory, seems already to threathen this argument. Still, we might argue that given the fact that our best theories *actually* apply set theory rather than one of its alternatives, counts as evidence that set theory somehow applies better to the physical world than its alternatives, and thus we can conclude that set theory is empirically confirmed.

However, there is an alternative explanation available for the fact that our best physical theories make use of classical mathematics, or mathematics based on set theory. For as we discussed in the conclusion of Part II of this thesis, the debate on the foundations of mathematics proves itself to be a *subjective* and also *very social affair*. Most mathematicians and physicists are *only* familiar with *classical* mathematics, because this is what is taught in high school and regular university classes. Objections against classical mathematics are often not known and rarely well-understood, and undoubtedly the same holds for the philosophical benefits that alternative foundational schemes might have in certain contexts.

Thusly conceived, it is no wonder that classical mathematics – and not an alternative foundational scheme – is heavily used in our current best physical theories. From this fact we should *not* infer, in my opinion, that it is also classical mathematics which applies *best* to the physical world. I think that the philosophical paradoxes which arise from the application of classical mathematics in physics should urge us not to take classical mathematics for granted, but rather to keep an open mind as to the question which mathematical framework to adopt in which context.

Bibliography

Bagaria, J. (2001), Set theory, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2017 edn.

Bangu, S. (2009), 'Understanding thermodynamic singularities: Phase transitions, data, and phenomena', *Philosophy of Science* **76**(4), 488–505.

Bascelli, T., Bottazzi, E., Herzberg, F., Kanovei, V., Katz, K. U., Katz, M. G., Nowik, T., Sherry, D. & Shnider, S. (2014), 'Fermat, Leibniz, Euler, and the gang: The true history of the concepts of limit and shadow', *Notices of the AMS* **61**(8).

Batterman, R. (2002), *The Devil in the Details: Asymptotic reasoning in Explanation, Reduction and Emergence*, Oxford University Press.

Batterman, R. (2005), 'Critical phenomena and breaking drops: Infinite idealizations in physics', *Studies in the History and Philosophy of Modern Physics* **36**, 225–244.

Batterman, R. (2010), 'On the explanatory role of mathematics in empirical science', *British Journal for the Philosophy of Science* **61**, 1–25.

Batterman, R. (2014), 'The inconsistency of Physics (with a capital "P")', *Synthese* **191**, 2973–2992.

Bauer, A. (2012), Intuitionistic mathematics and realizability in the physical world, *in* H. Zenil, ed., 'A Computable Universe', World Scientific.

Bauer, A. (n.d.), 'Proof of negation and proof by contradiction', http://math.andrej.com/2010/03/29/proof-of-negation-and-proof-by-contradiction/.

Beeson, M. (1985), *Foundations of constructive mathematics: Metamathematical studies*, Springer-Verlag, Berlin and Heidelberg.

Bell, J. (2005), Continuity and infinitesimals, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', summer 2017 edn.

Bell, J. L. & Korté, H. (2009), Hermann weyl, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2016 edn.

Benci, V. (1995), I numeri e gli insiemi etichettati, *in* 'Conferenza del seminario di matematica dell'Università di Bari', Vol. 261, Laterza, pp. 1–29.

Benci, V., Bottazzi, E. & Di Nasso, M. (2014), 'Elementary numerosity and measures', *Journal of Logic and Analysis* **6**.

Benci, V. & Di Nasso, M. (2003), 'Numerosities of labelled sets: A new way of counting', *Advances in Mathematics* pp. 50–67.

Benci, V., Di Nasso, M. & Forti, M. (2006), 'An Aristotelian notion of size', *Annals of Pure and Applied Logic* **143**, 43–53.

Benci, V., Horsten, L. & Wenmackers, S. (2016), 'Infinitesimal probabilities', *The British Journal for the Philosophy of Science* .

Bolzano, B. (1851), *Paradoxien des Unendlichen*, Bernard Bolzano Gesamtausgabe, C.H. Reklam sen., Leipzig.

Bolzano, B. (1978), Aphorismen zur physik, *in* J. Berg & J. Louzil, eds, 'Bernard Bolzano Gesamtausgabe series II: Nachlaß', Vol. 12, 3: Vermischte philosophische und physikalische Schriften 1832-1848, Frommann-Holzboog, Stuttgart-Bad Cannstatt.

Bridges, D. & Dedui, L. (1997), Paradise lost, or paradise regained?, *in* 'CDMTCS Research Report Series'.

Bridges, D. & Palmgren, E. (1997), Constructive mathematics, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2016 edn.

Brouwer, L. (1927/ 1967), On the domains of definitions of functions, *in* J. van Heijenoort, ed., 'From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931'.

Butterfield, J. (2011), 'Less is different:Emergence and reduction reconciled', *Foundations of Physics* **41**, 1065–1135.

Callender, C. (2001), 'Taking thermodynamics too seriously', *Studies in History and Philosophy of Modern Physics* **32**(4), 539–533.

Cantor, G. (1882), 'Über unendliche, lineare punktmannichfaltigkeiten', *Mathematische Annalen* **20**.

Cantù, P. (2013), An argumentative approach to ideal elements in mathematics, *in* A. Aberdein & I. Dove, eds, 'The Argument of Mathematics', Springer, pp. 79–99.

Cartwright, N. (1989), *Nature's capacities and their measurement*, Oxford University Press, Oxford.

Cartwright, N. (1999), *The dappled world: A study of the boundaries of science*, Cambridge University Press, Cambridge.

Chaitin, G. (2006), 'How real are real numbers?', *International Journal of Bifurcation and Chaos* (16), 1841–1848.

Colyvan, M. (2008), 'The ontological commitments of inconsistent theories', *Philosophical Studies* **141**(1), 115–128.

Colyvan, M. (2009), Applying inconsistent mathematics, *in* 'New Waves in Philosophy of Mathematics'.

Compagner, A. (1989), 'Thermodynamics as the continuum limit of statistical mechanics', *American Journal of Physics* **57**(106).

Crosilla, L. (2009), Set theory: Constructive and Intuitionistic ZF, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', summer 2015 edn.

Dauben, J. (1988), Abraham Robinson and non-standard analysis: History, philosophy, and foundations of mathematics, *in* W. Aspray & P. Kitcher, eds, 'Minesota Studies in the Philosophy of Science', University of Minnesota Press, Minneapolis.

Di Nasso, M. (1999), 'On the foundations of nonstandard mathematics', *Mathematica Japonica* **50**(1), 131–160.

Dummett, M. (1977), *Elements of Intuitionism*, Clarendon Press, Oxford.

Dummett, M. (2000), 'Is time a continuum of instants?', *Philosophy* **75**, 497–515.

Dummett, M. (2005), 'Hume's atomism about events: A response to Ulrich Meyer', *Philosophy* **80**(1), 141–144.

Earman, J. (1986), *A Primer on Determinism*, D. Reidel Publ., Dordrecht.

Earman, J. & Friedman, M. (1973), 'The meaning and status of Newton's law of inertia and the nature of gravitational forces', *Philosophy of Science* **40**(3), 329–359.

Feferman, S. (1998), *In the light of logic*, Oxford University Press, New York.

Feferman, S. (2005), Predicativity, *in* S. Shapiro, ed., 'The Oxford Handbook of Philosophy of Mathematics and Logic', Oxford: Oxford University Press, pp. 590–624.

Feferman, S. (2009), 'Conceptions of the continuum', *Intellectica* **51**(1), 169–189.

Ferreirós, J. (2004), 'The motives behind Cantor's Set theory: Physical, biological and philosophical questions', *Science in Context* **17**, 1–35.

Ferreirós, J. (2007), The early development of set theory, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', fall 2016 edn.

Fletcher, P. (2002), 'A constructivist perspective on physics', *Philosophia Mathematica* **10**, 26–42.

Fletcher, P. (2007), Infinity, *in* D. Jacquette, ed., 'Philosophy of Logic', Vol. 5 of *Handbook of the Philosophy of Science*, Elsevier, Amsterdam, pp. 523–585.

Fletcher, P., Hrbacek, K., Kanovei, V., Katz, M. G., Lobry, C. & Sanders, S. (2017), 'Approaches to analysis with infinitesimals following Robinson, Nelson, and others', *Real Analysis Exchange* **42**(2), 193–252.

Fletcher, S. (2012), 'What counts as a Newtonian system? The view from Norton's dome', *European Journal for Philosophy of Science* **2**(3), 275–297.

Frigg, R. & Hartmann, S. (2006), Models in science, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', fall 2012 edn.

George, A. & Velleman, D. (2002), *Philosophies of Mathematics*, Blackwell Publishers, Massachusetts and Oxford.

Goldenfeld, N. (1992), *Lectures on phase transitions and the renormalization group*, Westview Press, Boulder and Oxford.

Griffiths, R. (1972), Rigorous results and theorems, *in* C. Domb & M. Green, eds, 'Phase Transitions and Critical Phenomena: Volume 1 - Exact Results', Academic Press, London and New York.

Harris, T. (2003), 'Data models and the acquisition and manipulation of data', *Philosophy of Science* **70**, 1508–1517.

Heinonen, J. (2005), *Lectures on Lipschitz Analysis*, Lectures for the 14th Jyväskylä Summer School, University of Jyväskylä.

Hilbert, D. (1926), 'Über das Unendliche', *Mathematische Annalen* **95**, 161–190.

Horsten, L. (2007), Philosophy of mathematics, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', spring 2018 edn.

Iemhoff, R. (2008), Intuitionism in the philosophy of mathematics, *in* 'The Stanford Encyclopedia of Philosophy', winter 2016 edn.

Impens, C. & Sanders, S. (2007), ERNA at work, *in* 'The strength of nonstandard analysis', Springer, pp. 64–75.

Isham, C. & Butterfield, J. (2000), Some possible roles for topos theory in quantum theory and quantum gravity, *in* 'Conference on Towards a New understanding of Space, Time and Matter', Kluwer Academic/Plenum Publ., pp. 1707–1735.

Keisler, H. (2007), *Foundations of Infinitesimal Calculus.* On-line Edition: `https://www.math.wisc.edu/~keisler/foundations.html`.

Korolev, A. (2007), 'Indeterminism, Asymptotic Reasoning, and Time Irreversibility in Classical Physics', *Philosophy of Science, PSA 2006 contributed papers* **74**(5), 943–956.

Ladyman, J. (2008), Idealization, *in* S. Psillos & M. Curd, eds, 'The Routledge Companion to Philosophy of Science', Routledge Taylor & Francis Group, London and New York, pp. 358–366.

Lisker, R. (1996), 'Barrier theory: Finitism and Intuitionism in physics', `http://philsci-archive.pitt.edu/1306/`.

Liu, C. (1999), 'Explaining the emergence of cooperative phenomena', *Philosophy of Science* **66**, 92–106.

Longo, G. (1999), The mathematical continuum, from intuition to logic, *in* J. Petitot, F. J. Varela, B. Pacoud & J. Roy, eds, 'Naturalizing Phenomenology', Stanford University Press.

Lui, C. (2001), 'Infinite systems in SM explanations: Thermodynamic limit, renormalization (semi-)groups and irreversibility', *Philosophy of Science* **68**, 325–344.

Maddy, P. (1992), 'Indispensability and practice', *Journal of Philosophy* (6), 275–289.

Mainwood, P. (2005), 'Phase transitions in finite systems', `http://philsci-archive.pitt.edu/8340/`.

Malament, D. (2008), 'Norton's Slippery Slope', *Philosophy of Science* **75**(5), 799–816.

Maloney, M. (2008), 'Constructivism: A realistic approach to math?', `http://www.maa.org/press/periodicals/convergence/homsigmaa-student-paper-contest-winners`.

Mancosu, P. (2008), Explanation in mathematics, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', summer 2015 edn.

Mancosu, P. (2009), 'Measuring the size of infinite collections of natural numbers: Was Cantor's theory of infinite number inevitable?', *The Review of Symbolic Logic* **2**(4).

Mayo, D. (1996), *Error and the Growth of Experimental Knowledge*, The University of Chicago Press, Chicago.

McAllister, J. (1997), 'Phenomena and patterns in data sets', *Erkenntnis* **47**, 217–228.

McComb, W. (2004), *Renormalization Methods: A Guide for Beginners*, Clarendon Press, Oxford.

McMullin, E. (1985), 'Galilean idealization', *Studies in History and Philosophy of Science* **16**, 247–73.

Meyer, U. (2005), 'Dummett on the time-continuum', *Philosophy* **80**(311), 135–140.

Moore, M. (2007), 'The completeness of the real line', *CRITICA, Revista Hispanoamericana de Filosofia* **117**, 61–86.

Morrison, M. (1999), Models as autonomous agents, *in* 'Models as Mediators'.

Moschovakis, J. (1999), Intuitionistic logic, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', spring 2015 edn.

Norton, J. (2003), 'Causation as Folk Science', *Philosophers' Imprint* **3**, 1–22.

Norton, J. (2008), 'The Dome: An Unexpectedly Simple Failure of Determinism', *Philosophy of Science* **75**(5), 786–798.

Norton, J. (2012), 'Approximation and idealization: Why the difference matters', *Philosophy of Science* **79**, 207–232.

Norton, J. (2014), Infinite idealizations, *in* F. Stadtler, ed., 'European Philosophy of Science – Philosophy of Science in Europe and the Viennese heritage: Vienna Circle Institute Yearbook', Springer, Dordrecht-Heidelberg-London-New York.

Norton, J. (2016), 'The impossible process: Thermodynamic reversibility', *Studies in history and philosophy of modern physics* **55**, 43–61.

Palmgren, E. (1998), 'Developments in constructive nonstandard analysis', *The Bulletin of Symbolic Logic* **4**(3), 233–272.

Parker, M. (2013), 'Set size and the part-whole principle', *The Review of Symbolic Logic* **6**(4), 589–612.

Pincock, C. (2014), 'How to avoid inconsistent idealizations', *Synthese* **191**, 2957–2972.

Pincock, C. (n.d.), The applicability of mathematics, *in* 'Internet Encyclopedia of Mathematics'.

Potter, M. (2004), *Set Theory and its Philosophy: A Critical Introduction*, Oxford University Press, Oxford.

Putnam, H. (1967), Mathematics without foundations, *in* 'Mathematics, Matter, and Method: Philosophical Papers', Vol. I, Cambridge, New York.

Roberts, B. (2009), 'Wilson's case against the dome: Not necessary, not sufficient'.

Sanders, S. (2017), 'Reverse formalism 16', *Synthese* pp. 1–48.

Sanders, S. (2018), 'To be or not to be constructive, that is not the question', *Indagationes Mathematicae* **29**(1), 313–381.

Schechter, E. (2001), 'Constructivism is difficult', *The American Mathematical Monthly* **108**(1).

Scoville, N. (2012), 'Georg Cantor at the dawn of Point-set topology', *Convergence* .

Shech, E. (2013), 'What is the paradox of phase transitions?', *Philosophy of Science* **80**, 1170–1181.

Shech, E. (2015), 'Scientific misrepresentation and guides to ontology: the need for representational code and content', *Synthese* **192**, 3463–3485.

Shech, E. & Gelfert, A. (2016), 'The exploratory role of idealizations and limiting cases in models', `http://philsci-archive.pitt.edu/13338/`.

Simons, P. (2015), 'Bolzano's monadology', *British Journal for the History of Philosophy* **23**(6), 1074–1084.

Sommer, R. & Suppes, P. (1997), 'Dispensing with the continuum', *Journal of Mathematical Psychology* **41**, 3–10.

Stanley, H. (1971), *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, Oxford.

Stemeroff, N. & Dyer, C. (2016), 'On the differential calculus and mathematical constraints', `http://philsci-archive.pitt.edu/12569/`.

Suppes, P. (1969), *Studies in the methodology and foundations of science: Selected papers from 1951 to 1969*, Synthese Library, Springer Science Business Media b.v., Dordrecht.

Trench, W. (2013), 'Introduction to real analysis', Trinity University, Faculty Authored and Edited Books & CDs, 7, `https://digitalcommons.trinity.edu/mono/7`.

van Bendegem, J. (2002), Finitism in geometry, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', summer 2015 edn.

van Dalen, D. (1981), *Brouwer's Cambridge Lectures on Intuitionism*, Cambridge University Press, Cambridge.

van Fraassen, B. (2008), *Scientific Representation: Paradoxes of Perspective*, Clarendon Press, Oxford.

van Wierst, P. (2016), 'Profili: Bernard Bolzano', *APhEx* (14).

van Wierst, P. (2017), 'The paradox of phase transitions in the light of constructive mathematics', *Synthese* .

Waaldijk, F. (2005), 'On the foundations of constructive mathematics - especially in relation to the theory of continuous functions', *Foundations of Science* **10**(3), 249–324.

Weisberg, M. (2007), 'Three kinds of idealization', *Journal of Philosophy* **104**(12), 639–659.

Weisstein, E. (n.d.), 'Cardinal number', `http://mathworld.wolfram.com/CardinalNumber.html`.

Wenmackers, S. & Vanpoucke, D. (2016), 'Neo-leibnizian analysis of indeterminism in Newtonian physics'.

Weyl, H. (1949), *Philosophy of Mathematics and Natural Science*, Princeton University Press, Princeton.

Wilson, M. (2006), *Wandering Significance: An Essay on Conceptual Behavior*, Oxford University Press, New York.

Wilson, M. (2009), 'Determinism and the mystery of the missing physics', *British Journal for the Philosophy of Science* **60**, 173–193.

Wilson, M. (2013), What is "Classical Mechanics" anyway?, *in* R. Batterman, ed., 'The Oxford Handbook of Philosophy of Physics', Oxford University Press, New York.

Ye, F. (2011), *Strict Finitism and the Logic of Mathematical Applications*, Springer, Dordrecht.

Zalta, E. (1998), Frege's theorem and foundations for arithmetic, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', winter 2016 edn.

Zimba, J. (2008), 'Inertia and determinism', *British Journal for the Philosophy of Science* pp. 1–12.

Zinkernagel, H. (2010), Causal Fundamentalism in Physics, *in* 'EPSA Philosophical Issues in the Sciences', Springer, Dordrecht, pp. 311–322.