PHASE TRANSITIONS IN SCIENCE: SELECTED PHILOSOPHICAL TOPICS

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophie an der Ludwig-Maximilians-Universität
München
vorgelegt von

Patricia Palacios

March 2020

Referent: Prof. Stephan Hartmann
Korreferentin: Prof. Charlotte Werndl
Tag der mündlichen Prüfung: 10.07.2018

La fantasía, lo fantástico, lo imaginable que yo amo y con lo cual he tratado de hacer mi propia obra es todo lo que en el fondo sirve para proyectar con más claridad y con más fuerza la realidad que nos rodea.

Julio Cortázar, Lecciones de Literatura

Abstract

This dissertation examines various philosophical issues associated with the physics of phase transitions. In particular, i) I analyze the extent to which classical phase transitions impose a challenge for reductionism, ii) I evaluate the widespread idea that an infinite idealization is essential to give an account of these phenomena, and iii) I discuss the possibility of using the physics of phase transitions to offer a reductive explanation of cooperative behavior in economics.

Against prominent claims to the contrary, I defend the view that phase transitions do not undermine reductionism and that they are in fact compatible with the reduction of thermodynamics to statistical mechanics. I argue that this conclusion follows even in the case of continuous phase transitions, where there are two infinite limits involved.

My second claim is that the infinite idealizations involved in the physical treatment of phase transitions although useful are not indispensable to give an account of the phenomena. This follows from the fact that the thermodynamic limit provides us with a controllable approximation of the behavior of finite systems. My third claim is that the physics of phase transitions, in particular renormalization group methods, can constitute a promising way of giving a reductive explanation of stock market crashes. This will serve not only to motivate the use of statistical mechanical methods in the study of economic behavior, but also to contradict the claim that renormalization group explanations are always non-reductive explanations.

Zusammenfassung

Diese Doktorarbeit untersucht verschiedene philosophische Probleme, die mit der Physik der Phasenübergängen zu tun haben. Insbesondere i) analysiere ich, ob das klassische Phasenübergänge tatsächlich eine Herausforderung für den Reduktionismus ist, ii) bewerte ich die Idee, dass eine unendliche Idealisierung notwendig ist, um eine Erklärung der Phasenübergänge zu geben und iii) diskutiere ich die Möglichkeit, die Physik der Phasenübergängen zu verwenden, um eine reduktive Erklärung des kooperativen Verhaltens in der Wirtschaft anzubieten.

Gegen prominente Ansprüche auf das Gegenteil verteidige ich die Ansicht, dass Phasenübergänge Reduktionismus nicht untergraben, und dass sie tatsächlich mit der Reduktion der Thermodynamik zur statistischen Mechanik vereinbar sind. Ich behaupte, dass dieser Beschluss sogar im Falle von kontinuierlichen Phasenübergängen folgt, wo es zwei unendliche Limes gibt.

Mein zweiter Anspruch besteht darin, dass die unendlichen Idealisierungen an der physischen Behandlung von Phasenübergängen, obwohl nützlich, nicht notwendig sind, um eine Erklärung der Phänomene zu geben. Das folgt aus der Tatsache, dass die thermodynamische Limes uns mit einer kontrollierbaren Annäherung des Verhaltens von endlichen Systemen versorgt. Mein dritter Anspruch besteht darin, dass die Physik von Phasenübergängen, inbesondere die Renormalisierungsgruppe, eine versprechende Weise einsetzen kann, eine reduktive Erklärung von Börsencrashs zu geben. Das wird nicht nur dienen, um den Gebrauch von statistischen mechanischen Metho-

den in der Studie des Wirtschaftsverhaltens zu verleiten, sondern auch um dem Anspruch zu widersprechen, dass die Erklärung der Renormalisierungsgruppe immer nichtreduktive Erklärungen sind.

Acknowledgements

I am deeply grateful to my supervisor Professor Stephan Hartmann, who not only patiently read and criticized the different chapters that constitute my dissertation but also gave me support and good advice to start building my academic career.

During all these years, I had the opportunity to learn from many outstanding philosophers of science who helped me along the way with superb comments and contribute to almost every single aspect of this work. I am particularly grateful to Giovanni Valente, who spent considerable amount of time helping me organize my thoughts and gave me invaluable support during all these years, to Erik Curiel for all the help, the discussions and very detailed feedback on all the chapters of this work, and to Sam Fletcher, who gave me detailed feedback and motivated my interest in reduction even more.

I am very grateful to Jim Weatherall, who made important suggestions on how to improve my work and invited me to Irvine to work on the project that led to Chapter 4 of this dissertation. I am also indebted to Jennifer Jhun, who worked with us on the project and patiently helped me understand economics.

I thank Charlotte Werndl, for her valuable feedback on Chapters 2 and 3, and for happily agreeing to be part of my dissertation committee. I also thank Jos Uffink for all the conversations in Minneapolis that led to Chapter 3 of this dissertation, and I thank Roman Frigg, who motivated my interest in phase transitions and encouraged me to do a PhD on this topic when I was still an MSc student at the LSE.

I am also indebted to Lapo Casetti, Federico Benitez, Neil Dewar, Ben Feintzeig, Laurenz Hudetz, Katie Robertson, Pauline van Wierst, Paul Weingartner, and Jingyi Wu for helpful feedback and conversations about various points of Chapter 2, to Casey McCoy for his feedback on Chapter 3, and to Alexander Reutlinger for his comments on Chapter 4. I am also grateful to Karim Thébault and Sean Gryb for encouraging me to start thinking about the relationship between phase transitions and black holes.

I would like to thank the Deutscher Akademischer Austauschdienst (DAAD), which awarded me a DAAD scholarship to support my doctoral studies, the National Science Foundation (NSF), which partially supported my research in Irvine, and the Österreichischer Austauschdienst (OeAD), which awarded me with a visiting scholarship to pursue research in Salzburg.

I am sure things would not have turned out so well, if I would not have had the support of my friends and family in Chile and Europe. I am really grateful to those in Chile who helped me when I decided to move abroad and to those who have continued contributing in different ways during all these years: Natalia Santander, Josefa Vega, Carla Arévalo, Rodrigo Henríquez, María Elena Tobar, Roberto Astorga, Katherine Paez, Cristina Santander, Carmen Gloria Valenzuela and Francisco Pereira. I would specially like to thank Patricia Santander, who has been a crucial and constant support during these years. I am also grateful to my friends in Europe, who have inspired me one way or another: Albert Ibanez, Alex Zwez, Aidan Lyon, Pablo Ruiz de Olano, Orlando Zumbo and Lavinia Picollo.

Finally, I would like to dedicate this dissertation to my mother, Marianela Palacios, who strongly and selflessly has always encouraged my academic career.

Contents

1 Introduction **11**

2 Phase Transitions: A Challenge for Reductionism? **16**

2.1 Introduction 16

2.2 From Statistical Mechanics to the Thermodynamics of Phase Transitions 18

2.2.1 The Thermodynamics of Phase Transitions 18

2.2.2 The Importance of the Thermodynamic Limit 21

2.2.3 The Appeal to a Second Limit: Infinite Iteration of RG Transformations 23

2.3 The Concept of Limiting Reduction 30

2.3.1 Nickles' Concept of Limiting Reduction 30

2.3.2 Beyond Nickles' Concept of Reduction 32

2.4 Are Continuous Phase Transitions Incompatible with Reduction? 34

2.4.1 The Problem of "Singular" Limits 35

2.4.2 Implementing RG Methods 38

2.5 Approximation, Topology and the Reduction of Continuous Phase Transitions 42

2.6 Concluding Remarks 46

3 Had We But World Enough, and Time... But We Don't! **48**

3.1 Introduction 48
3.2 The Thermodynamic Limit in the Theory of Phase Transitions 50
3.2.1 The Problem of Phase Transitions 51
3.2.2 Butterfield's Principle and Butterfield's Solution to the Problem of Phase Transitions 54
3.3 The Infinite-time Limit in the Ergodic Explanation of Equilibrium 59
3.3.1 The Problem of the Infinite-Time Limit 59
3.3.2 The Dog-Flea Model and a Straightforward Justification for the Infinite-Time Limit 63
3.3.3 The Importance of the Rate of Convergence 65
3.4 Conclusion: Infinite Limits as Controllable Approximations . . 69

4 Market Crashes as Critical Phase Transitions? 72
4.1 Introduction 72
4.2 Some Financial and Economic Background to the JLS Model . 76
4.3 The JLS model 79
4.4 The Logic of the JLS model 86
4.4.1 Do Market Crashes Constitute a Universality Class? . . 87
4.4.2 On the Reductive and Explanatory Character of the JLS model 91
4.5 Policy Implications 107
4.6 Conclusion 109

5 Conclusion 111
5.1 Prospectus 111
5.1.1 Universal Explanations 111
5.1.2 Quantum Phase Transitons 112
5.1.3 Finite Theories of Phase Transitions 112
5.1.4 Defining Equilibrium for Symmetry-breaking Phase Transitions 113

5.1.5 Analogue Experiments 114
5.2 Summing up 115

Chapter 1

Introduction

A question that has puzzled scholars since Democritus is how exactly is it that the macroworld of our everyday experience arises out of the behavior of the microconstituents of matter. This is in fact the question that motivated the development of statistical mechanics, which is a theory that aims to explain how macroscopic phenomena, especially thermodynamic phenomena, originate in the cooperative behavior of interacting lower level entities. From its origin, in the latter half of the nineteenth century, until now, statistical mechanics has successfully derived many macroscopic thermal phenomena from the laws governing the interactions of microscopic constituents and probabilistic assumptions, however in some important cases such a derivation has been particularly problematic. One of those cases is the irreversible approach to equilibrium and the other case, which is the one that I will focus on in this dissertation, is equilibrium phase transitions.

Phase transitions are those sudden changes in the phenomenological properties of a system that we observe every time that we see liquid water turning into vapor. Other typical examples include the transition from a normal conductor to a superconductor and from a paramagnetic to a ferromagnetic phase in magnetic materials. Surprisingly, the microscopic explanation of these everyday phenomena has constituted one of the major challenges of statistical mechanics. In fact, in order to give such an account it has been necessary the appeal to the thermodynamic limit, whereby the volume and the number of particles go to infinity, and, in the case of continuous phase transitions, the introduction of renormalization group methods, which is an entirely new theoretical framework that basically consists in reducing the number of effective degrees of freedom, losing information about the fine grained details of the system.

The use of the thermodynamic limit and the introduction of renormalization group techniques in the theory of phase transitions motivated an anti-reductionist position among physicists working in condensed matter physics, who considered phase transitions as a genuine example of emergent behavior in physics. In his celebrated paper "More is Different" (Anderson 1972), Philip Anderson says that the use of the thermodynamic limit "is not only convenient but essential to realize that matter will undergo mathematically sharp, singular 'phase transitions'" (p. 395). As a consequence of that, he argues that the properties of a huge number of constituents, all working together, were different from the behavior of a few of these particles. In a similar vein, the statistical mechanic Lebowitz (1999) claims that phase transitions are "paradigms of emergent behavior" (p.2), arguing that the properties of this collective behavior had no counterpart in the behavior of individual atoms.

Recently, considerations of this sort have entered the philosophical debate and have been at the center of philosophical discussions on reduction, idealizations and explanations in science. In this discussion, some philosophers have argued that statistical mechanics cannot provide a full reductive account of phase transitions in finite systems and that this undermines reductionism, i.e. the belief that ultimately all macroscopic laws are reducible to the fundamental microscopic laws of physics (i.e. Batterman 2011, Morrison 2012, Bangu 2009). At the same time, it has been claimed that the infinite idealization is indispensable to give an account of phase transitions

(Batterman 2005, Jones 2006), which has led some of them to the conclusion that discontinuities are physically real (Batterman 2005). Finally, it has been said (Batterman and Rice 2014) that the use of renormalization group techniques leads to a special kind of explanations in science which they regard as non-reductive and non-causal.

In this dissertation, I address these positions and analyze the extent to which the physical treatment of classical phase transitions actually call into question important philosophical theses. Although I will admit that phase transitions challenge various notions of reduction present in the physical and philosophical literature, my main claim is that phase transitions do not undermine reductionism and that they are in fact compatible with the reduction of thermodynamics to statistical mechanics. Against prominent claims to the contrary, I will defend the view that the statistical mechanical treatment of these phenomena actually succeeds in building a connection between the thermodynamic behavior of phase transitions and the cooperative behavior of lower level entities.

My second claim is that the infinite idealizations involved in the physical treatment of phase transitions although useful are not indispensable to give an account of the phenomena. This will follow from the fact that the thermodynamic limit provides us with a controllable approximation of the behavior of finite systems, which is not necessarily the case when other limiting operations, such as the infinite-time limit, are involved.

My third claim is that the use of renormalization group techniques do not necessarily lead to non-reductive explanations. Contra what has been claimed by some philosophers (e.g. Batterman and Rice 2014), I will argue that these methods can constitute a promising way of offering a reductive account for the behavior of collective phenomena not only in physics but also in the social sciences, in particular, in economics.

These claims will be made along the next three chapters, each of which is to a large extent self-contained. In the next chapter (Chapter 2), I will focus on the problem of the reduction of phase transitions and argue that despite the use of the thermodynamic limit and the introduction of renormalization group methods, phase transitions are compatible with inter-theory reduction. The notion of inter-theory reduction that I will endorse is a revised version of Nickles' (1973) notion of limiting reduction. This notion departs from the traditional Nagelian model reduction, but in my view accommodates better the limiting operations involved in the statistical mechanical treatment. In the same chapter, I argue against the idea that the thermodynamic limit is indispensable to give an account of phase transitions.

In Chapter 3, I put the emphasis on the justification for the empirical success of infinite limits. After a systematic comparison between the thermodynamic limit in the theory of phase transitions and the infinite-time in the ergodic theory of equilibrium, I will conclude that what allows for a justification of the empirical success of the thermodynamic limit is that this limit is controllable, which means that one has control over how large the value of the parameter must be assure that the infinite limit is a reasonable substitute for a finite system. This will also serve to undermine claims about the indispensability of the thermodynamic limit in the theory of phase transitions, but it will make salient some problems associated with the justification of the infinite-time limit in statistical mechanics.

Finally, in Chapter 4, I will indirectly address the question of whether renormalization group explanations always constitute non-reductive explanations. I will do so by considering a specific model of econophysics, the Johansen-Ledot-Sornette (JLS) model that treats stock market crashes as critical phase transitions and, therefore, uses renormalization group techniques. After a careful analysis of the epistemic role of this highly idealized model, I will conclude that this model constitutes a promising way of giving a reductive and causal explanation for stock market crashes that can also help visualize possible avenues for intervention. This will serve not only to motivate the use of statistical mechanical methods in the study of economic behavior, but also to contradict the idea that renormalization group explanations are always non-reductive explanations.

Although this dissertation aims to resolve different problems that have raised in the philosophical discussion around phase transitions, there are many other issues associated with phase transitions that I will not be able to address here. In Chapter 5, I will offer an overview of some of these issues with the purpose of motivating the philosophical discussion on these topics.

Chapter 2

Phase Transitions: A Challenge for Reductionism?

> "By convention sweet is sweet, bitter is bitter, hot is hot, cold is cold, color is color; but in truth there are only atoms and the void." [Democritus, trans. Durant 1939]

2.1 Introduction

Phase transitions are sudden changes in the phenomenological properties of a system. Some common examples include the transition from liquid to gas, from a normal conductor to a superconductor, or from a paramagnet to a ferromagnet. Nowadays phase transitions are considered one of the most interesting and controversial cases in the analysis of inter-theory relations. This is because they make particularly salient the constitutive role played by idealizations in the inference of macroscopic behavior from a theory that describes microscopic interactions. In fact, it appears that statistical mechanics – a well-established microscopic theory – cannot account for the behavior of phase transitions as described by thermodynamics – a macroscopic theory – without the help of infinite idealizations in the form of mathematical limits.

In the discussion on phase transitions, physicists and philosophers alike have mainly been concerned with the use of the thermodynamic limit, an idealization that consists of letting the number of particles as well as the volume of the system go to infinity. For many authors (e.g. Bangu 2009, Bangu 2011; Batterman 2005; Batterman 2011, Morrison 2012) this idealization has an important philosophical consequence: it implies that phase transitions are *emergent phenomena*. As a result, they claim that such phenomena present a challenge for reductionism, i.e. the belief that ultimately all macroscopic laws are reducible to the fundamental microscopic laws of physics.

On the other hand, numerous other authors (e.g. Butterfield 2011; Butterfield and Buoatta 2011; Norton 2012; Callender 2001; Menon and Callender 2013) have rejected this conclusion, arguing that the appeal to the infinite limit does not represent a problem for reductionism. Some of them (Butterfield 2011, Norton 2012) have even argued that phase transitions, instead of threatening reductionism, are paradigmatic examples of Nagelian reduction, whereby reduction is understood in terms of logical deduction.

These last remarks, however, have not ended the debate. In particular, the physical treatment of continuous phase transitions that implements renormalization group (RG) techniques is still regarded as especially problematic for the reductionist attitude towards phase transitions (e.g. Batterman 2011, Morrison 2012).

In this chapter, I analyze the extent to which classical phase transitions, especially continuous phase transitions, impose a challenge for reductionism. My main contention is that classical phase transitions are, in fact, compatible with reduction, at least with the notion of reduction that relates the behavior of physical quantities in different theories under certain limiting conditions. I argue that this conclusion follows even if one recognizes the existence of two infinite limits involved in the physics of continuous phase transitions.

To reach my goal, I organize this chapter as follows. In the next section (Section 2.2), I describe the physics of phase transitions, outlining how statistical mechanics recovers thermodynamical behavior. Here I emphasize that in the RG treatment of continuous phase transitions, apart from the thermodynamic limit, there is a second infinite limit involved. Subsequently (Section 2.3), I further develop the concept of limiting reduction suggested by Nickles (1973). Based on that notion of reduction, I contend (Section 4) that, despite some objections, first-order phase transitions satisfy Nickles' criterion of limiting reduction. However, I also show that continuous phase transitions do not satisfy this criterion due to the existence of the second infinite limit. In Section 5, I suggest to liberalize the notion of limiting reduction and I argue that continuous phase transitions fulfill this notion. This paper concludes by describing some attempts to apply RG methods to finite systems, which indeed support the claim that thermodynamical phase transitions are reducible to statistical mechanics.

2.2 From Statistical Mechanics to the Thermodynamics of Phase Transitions

Statistical mechanics aims to account for the macroscopic behavior typically described by thermodynamics in terms of the laws that govern microscopic interactions. In the philosophical literature, the reproduction of the thermodynamic results by statistical mechanics is generally referred to in terms of *reduction*. In this section, I will describe how statistical mechanics recovers the thermodynamic behavior of phase transitions and will explain why phase transitions are an interesting and puzzling case for the project of reducing thermodynamics to statistical mechanics.

2.2.1 The Thermodynamics of Phase Transitions

In thermodynamics, phases correspond to regions of the parameter space (known as a phase diagram) where the values of the parameters uniquely specify equilibrium states. Phase boundaries, in contrast, correspond to values of parameters at which two different equilibrium states can coexist. The coexistence of states expresses itself as discontinuities of thermodynamic quantities, like volume, which are related to the first derivatives of the free energy with respect to a parameter such as pressure or temperature. If the system goes from one phase to another intersecting a phase boundary, the system is said to undergo a *first-order phase transition*. This name is due to the fact that the discontinuous jumps occur in the first derivatives of the free energy. On the other hand, if the system moves from one phase to another without intersecting any coexistence line, the system is said to undergo a *continuous phase transition*, in which case there are no discontinuities involved in the first derivatives of the free energy but there are divergencies in the response functions (e.g. specific heat, susceptibility for a magnet, compressibility for a fluid). An example of a first-order phase transition is the passage from liquid water to vapor at the boiling point, where the quantities that experience discontinuous jumps are entropy and volume, which are first derivatives of the free energy with respect to temperature and pressure respectively. An example of continuous phase transition instead is the transition in magnetic materials from the phase with spontaneous magnetization – the ferromagnetic phase – to the phase where the spontaneous magnetization vanishes – the paramagnetic phase –. (Figure 1)

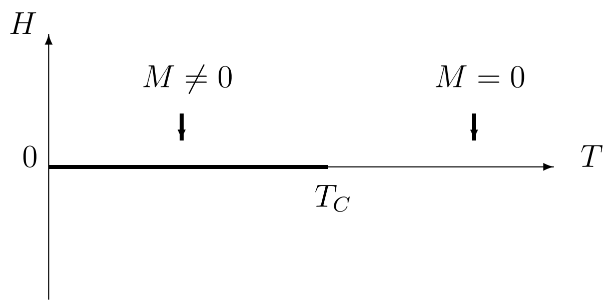

Figure 1: Phase diagram for the paramagnetic–ferromagnetic transition. Here H is the external magnetic field and T the temperature. At the transition or critical point T_C the spontaneous magnetization M vanishes.

Although both first-order and continuous phase transitions are of great interest for the project of reducing thermodynamics to statistical mechanics, the latter kind is considered to be more controversial than the former. The reason is that continuous phase transitions have characteristic properties that are much more difficult to recover from statistical mechanics than first-order phase transitions. One of those properties is that, in the vicinity of a continuous phase transition, measurable quantities depend upon one another in a power-law fashion. For example, in the ferromagnetic-paramagnetic transition, the net magnetization M, the magnetic susceptibility χ, and the specific heat C depend on the reduced temperature $t = \frac{T-T_c}{T_c}$ (the temperature of the system with respect to the critical temperature T_c) as follows:

$$M \sim |t|^{\beta}, C \sim |t|^{-\alpha}, \chi \sim |t|^{-\gamma},$$

where β, α, γ are the *critical exponents*. Another remarkable property of continuous phase transitions is that radically different systems, such as fluids and ferromagnets, have exactly the same values of critical exponents, a property known as *universality*.

Finally, continuous phase transitions are also characterized by the divergence of some physical quantities at the transition or critical point. The critical exponents α and γ are typically (although not always) positive, so that the power laws that have negative exponents (and the corresponding quantities like specific heat and susceptibility) diverge as $T \to T_c$. The divergence of the magnetic susceptibility χ implies the divergence of the *correlation length* ξ, a quantity that measures the distance over which the spins are correlated, which also obeys power-law behavior: $\xi \sim |t|^{-\nu}$. The divergence of the correlation length is perhaps the most important feature of continuous phase transitions because it involves the loss of a characteristic scale at the transition point and thus provides a basis for universal behavior.

The inference of the experimental values of critical exponents – or adequate relations among them – together with the account of universality has been one the major challenges of statistical mechanics. We will see next that, in order to provide such an account, it was necessary to appeal to infinite idealizations and to RG methods, an entirely new theoretical framework, which basically consists in reducing the number of effective degrees of freedom of the system.

2.2.2 The Importance of the Thermodynamic Limit

We saw in the previous section that the macroscopic behavior of first-order phase transitions is defined in terms of singularities or non-analyticities in the first derivatives of the free energy. Gibbsian statistical mechanics offers a precise definition of the free energy F, given by:

$$F(K_n) = -\kappa_B T \ln Z, \tag{2.1}$$

where K_n is the set of coupling constants, κ_B is the Boltzmannian constant, T is the temperature, and Z is the canonical partition function, defined as the sum over all possible configurations:

$$Z = \sum_i e^{\beta H_i}. \tag{2.2}$$

When trying to use statistical mechanics to recover the non-analyticities that describe phase transitions in thermodynamics, the following problem arises. Since the Hamiltonian H is usually a non-singular function of the degrees of freedom, it follows that the partition function, which depends on the Hamiltonian, is a sum of analytic functions. This means that neither the free energy, defined as the logarithm of the partition function, nor its derivatives can have the singularities that characterize first-order phase transitions in thermodynamics. Taking the thermodynamic limit, which consists of letting the number of particles as well as the volume of the system go to infinity $N \to \infty$, $V \to \infty$ in such a way that the density remains finite, allows one to recover those singularities. In this sense, the use of this limit appears essential for the recovery of the thermodynamic values, which motivated Kadanoff's controversial claim: "phase transitions cannot occur in finite systems, phase transitions are solely a property of infinite systems" (Kadanoff, 2009, p. 7).

The appeal to the thermodynamic limit is also found in the description of continuous phase transitions. Consider again the paramagnetic-ferromagnetic transition. This is a continuous phase transition defined in terms of the divergence of the magnetic susceptibility at the critical temperature and characterized by the appearance of spontaneous magnetization in the absence of an external magnetic field. From a statistical mechanical point of view, the appearance of spontaneous magnetization in finite systems is, strictly speaking, impossible. The impossibility is due to the up-down symmetry of the lattice models used in the study of magnetization, including the Ising model. A consequence of up-down symmetry is that for zero external field H the magnetization obeys the symmetry condition $M = -M$, whose unique solution is $M = 0$. That means that the magnetization M with zero external magnetic field H must be zero (Details elsewhere, e.g. Goldenfeld

1992, Sec. 4; Le Bellac, Mortessagne, and Batrouni 2006, Sec. 4). This so-called "impossibility theorem" can be avoided by taking the thermodynamic limit $N \to \infty$ followed by the limits $H \to 0^+$ and $H \to 0^-$:

$$M = \lim_{H\to 0^+} \lim_{N\to\infty} \frac{1}{N} \frac{\partial F(H)}{\partial H} \neq 0$$

$$-M = \lim_{H\to 0^-} \lim_{N\to\infty} \frac{1}{N} \frac{\partial F(H)}{\partial H} \neq 0.$$

Notice that since M and $-M$ have different values and are different from zero, the magnetic susceptibility, defined as the derivative of the magnetization with respect to an external field, diverges to infinity in the neighborhood of the zero external field. One can see, therefore, that taking the thermodynamic limit not only provides the concept of spontaneous magnetization with precise meaning but also allows for the recovery of the divergence of the thermodynamic quantities that characterizes continuous phase transitions.

2.2.3 The Appeal to a Second Limit: Infinite Iteration of RG Transformations

In an ideal scenario, one would expect to perform a direct calculation of the partition function. Unfortunately, analytic calculations of the partition functions have been performed only in particular models with dimension $D = 1$ or $D = 2$; for all other cases, one requires to use approximation techniques.[1] The most useful approximation for the case of first-order phase transitions is the *mean field approximation*, which employs the assumption that each spin acts as if it were independent of the others, feeling only the *average mean field*. Although the mean field approximation proved to be

[1]The first and most famous exact solution of the partition function is the Onsager solution for an Ising model of dimension D=2.

successful in some cases of first order phase transitions, experiments have shown that this account fails to give accurate predictions for the case of continuous phase transitions, in which the correlation length diverges. It is believed that this failure is due to the fact that mean field theories neglect fluctuations whereas fluctuations govern the behavior near the critical point.

A more complete account of continuous phase transitions requires the use of RG methods. These methods are mathematical and conceptual tools that allow one to solve a problem involving long-range correlations by generating a succession of simpler (generally local) models. The goal of these methods is to find a transformation that successively coarse-grains the effective degrees of freedom but keeps the partition function and the free energy (approximately) invariant. The usefulness of RG methods lies in the fact that one can compute the critical exponents and other universal properties without having to calculate the free energy. This methods also allow to account for universality, the remarkable fact that entirely different systems behave qualitatively and quantitatively in the same way near the critical point.

To give a specific illustration of RG methods, let us consider a block spin transformation for a simple Ising model on the two-dimensional square lattice with distance a between spins.[2] Here, the spins have two possible values, namely ± 1. If it is assumed that the spins interact only with an external magnetic field h and with their nearest neighbors through the exchange interaction K (meaning that the coupling constants are only K and h), the Hamiltonian H for the model is given by:

$$H = -K \sum_{ij}^{N} S_i S_j + -h \sum_i S_i. \tag{2.3}$$

[2]For simplicity, I am going to restrict the analysis to real-space renormalization. However, I think that the same conclusions apply to momentum-space renormalization. For details on the difference between real-space and momentum space-renormalization, see Wilson and Kogut (1974) and Fisher (1998). For a philosophical account on the difference between those two frameworks see Franklin (2017)

By applying the majority rule, which imposes the selection of one state of spin based on the states of the majority of spins within a block, one can replace the spins within a block of side la by a single *block spin*. Thus, one obtains a system that provides a coarse-grained description of the original system.

If one assumes further that the possible values for each block spin S_I are the same as in the Ising model, namely ± 1, and also that the block spins interact only with nearest neighbor block spins and an external field, the effective Hamiltonian H' will have the same form as the original Hamiltonian H:

$$H' = -K' \sum_{IJ}^{Nl^{-d}} S_I S_J + -h' \sum_{I} S_I. \tag{2.4}$$

Formally, this is equivalent to applying a transformation R to the original system, so that $H' = R[H]$, in which the partition function and the free energy remain approximately invariant.[3]

Although the systems described by H and H' have the same form, the correlation length in the coarse-grained system $\xi[K']$ is smaller than the correlation length $\xi[K]$ of the original system. This follows from the fact that the correlation length in the effective model is measured in units of the spacing la whereas the correlation length in the original system is measured in units of the spacing a. In other words, the correlation length is rescaled by a factor l. The expression that relates the correlation lengths of the two systems is:

$$\frac{\xi[K]}{l} = \xi[K']. \tag{2.5}$$

[3]The previous example captures the spirit of real-space RG methods. However in practice RG transformations consist of complicated non-linear transformations that do not preserve the form of the original Hamiltonian. This allows for the possibility that new local operators are generated during the RG transformation (Details in Goldenfeld 1992, p. 235).

After n iterations of the RG transformation, the characteristic linear dimension of the system is l^n. Thus the correlation lengths in the sequence of coarse-grained models vary according to:

$$\xi[K] = l\xi[K'] = ... = l^n \xi[K^{(n)}]. \tag{2.6}$$

The idea is that one iterates the RG transformation until fluctuations at all scales up to the physical correlation length ξ are averaged out. In many cases, this involves numerous iterations (Details elsewhere, e.g. Le Bellac et al. 2006, Sec. 4.4.3; Goldenfeld 1992, Sec. 9.3).

It follows from equation (6) that for a large correlation length, the number of iterations should be large. For an infinite correlation length, which is the case of continuous phase transitions, the number of iterations should be infinite.[4] Indeed, if the original correlation length $\xi[K]$ is infinite and we want to eliminate all effective degrees of freedom, i.e. we want the effective correlation length to be small, then we are forced to take the limit $n \to \infty$ in the right hand side of equation (6) such that the following expression holds:

$$\xi[K] = \lim_{n\to\infty} l^n \xi[K^{(n)}] = \infty \tag{2.7}$$

This result is important because it demonstrates the existence of two different infinite limits involved in the theory of phase transitions. The first is the thermodynamic limit that takes us to a system with an infinite correlation length. The second is the limit for the number of RG iterations going to infinity that takes us to a fixed point Hamiltonian, i.e. the Hamiltonian with the coupling constants equal to their fixed point values: $[K^*] = R[K^*]$. These fixed points can be also thought of as stationary or limiting distributions to which the renormalization group trajectories converge after infinite iterations

[4] In order to maintain the system at criticality, one performs a sort of double rescaling process: one changes scale in space and also changes the distance to criticality in coupling space (Details in Sornette 2000, p. 232).

of the RG transformation $n \to \infty$. This point will be crucial for what will be argued in Sections 3.3 and 5.2.

Although the iteration of the RG transformation preserves the symmetries of the original system, it does not preserve the value of the original Hamiltonian, and, therefore, it does not preserve the value of the set of coupling constants $[K]$ associated with the corresponding Hamiltonians. Thus, the iteration of the RG transformation can be thought of as describing a sequence of points moving in a space of coupling constants K^n or a corresponding space of Hamiltonians H. If the sequence describes a system at the critical point, after infinite iterations $n \to \infty$ it will converge to a non-trivial fixed point [K*] given by:

$$[K^*] = R[K^*] \tag{2.8}$$

The other possible fixed points are trivial, namely $K = 0$ and $K = \infty$, which correspond to low and high temperature fixed points respectively.

At fixed points the coupling constants remain invariant under the transformation. Therefore, varying the length scale does not change the value of the Hamiltonian and therefore brings us to a physically identical system. This latter feature associates fixed points with the property of scale invariance, which means that the system looks statistically (and physically) the same at different scales.

It has been shown that by linearizing in the vicinity of the fixed point, one can calculate the values of the critical exponents and the relations between them (Details in Goldenfeld 1992, Sec. 9; Domb 2000, Sec. 7; Sornette 2000, Sec. 11). This is remarkable because it demonstrates that the critical exponents are solely controlled by the RG trajectory near the fixed point and that one does not need to calculate the free energy to determine the behavior of the system in the vicinity of the critical point. This means also that the initial values of the coupling constants do not determine the critical behavior. The latter constitutes the origin of the explanation of *universality* because it tells us that systems that flow towards the same fixed point are governed by the same critical exponents, even if they are originally described by different coupling constants. The systems that flow towards the same fixed point – that are in the basin of attraction of the fixed point – are said to be in the same *universality class*.

In summary, we have seen that the recovery of the thermodynamic properties from statistical mechanics involves: i) first, the introduction of particular assumptions (e.g. lattice structure, a particular kind of degrees of freedom, ranges of values of the degrees of freedom, and dimension) that allow one to build a specific model (Ising model in our case study); ii) second, the assumption of the thermodynamic limit, which brings us to a fine-grained system with infinite number of particles and infinite correlation length;[5] and iii) finally, the assumption of a second infinite limit that consists of an infinite number of iterations of a coarse-graining transformation. This limit takes us to a fixed point Hamiltonian that represents a coarse-grained model. After those steps are made, the most important statistical mechanical approaches can make accurate predictions of the behavior of continuous phase transitions and explain universal behavior. Figure 2 illustrates this process. Notice, however, that in the case of first-order phase transitions, one could in principle derive the thermodynamic behavior just after taking the first limit.[6]

[5] Recently, Norton (2012) has challenged the appeal to an infinite system in the theory of phase transitions. His contention is that the limit system would have properties that are not suitable to describe phase transitions, such as the violation of determinism and energy conservation. This point is relevant for his distinction between idealizations and approximations, which led him to the conclusion that phase transitions are a case of approximation and not idealization. Since we are trying to make a different point here, we are going to adhere to the standard *façon de parler* that refers to the existence of an "infinite system" (e.g. Kadanoff 2009; Fisher 1998; Butterfield 2011). This does not mean that our view is incompatible with Norton's view.

[6] One should bear in mind that although RG methods are not required to infer the behavior of first-order phase transitions, they can be (and have been) used to describe these kinds of transitions as well. See Goldenfeld (1992, Sec. 9).

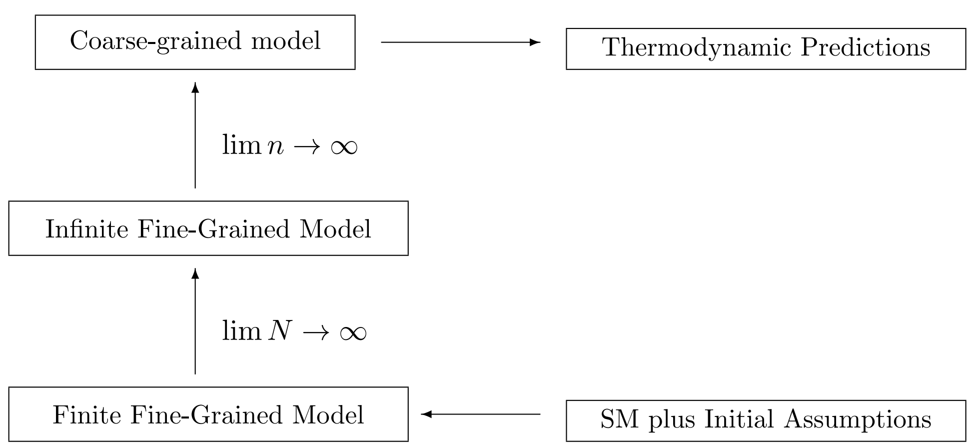

Figure 2: Inter-theory relation for continuous phase transitions.

2.3 The Concept of Limiting Reduction

What has been at stake in the philosophical debate around phase transitions is whether the thermodynamic description of these phenomena *reduces* to statistical mechanics. Even if the previous section showed that statistical mechanics can reproduce the non-analyticities that describe phase transitions in thermodynamics, the appeal to the infinite idealizations throws suspicion to the legitimacy of such a reduction. The main aim of Sections 4 and 5 is to evaluate whether the infinite idealizations mentioned in Section 2 are compatible with the reduction of phase transitions. However, given that the term "reduction" is notoriously ambiguous, before we can assess this issue, some clarifications as to how this term is constructed in this context are necessary. This is the task of the present section.

2.3.1 Nickles' Concept of Limiting Reduction

Since we are interested in relating the thermodynamic treatment of phase transitions with another theory that aims to describe the same phenomena, we are treating phase transitions as a potential case of inter-theory reduction, where reduction is taken as a relation between two theories (or parts of theories). This kind of reduction is to be distinguished from other types of reduction such as whole-parts reduction.[7] More specifically, since the description of the phenomenon in the two theories coincides only by assuming a limit process, the case of interest is a candidate for a specific class of inter-theory reduction sometimes called *limiting* or *asymptotic* reduction (Landsman, 2013).

Nickles (1973, pp. 197-201), who was the first to distinguish *limiting* reduction from other classes of inter-theory reduction, calls this type of reduction *reduction*$_2$ (henceforth LR_2) to distinguish it from *reduction*$_1$, which corresponds to Nagelian reduction. He characterizes LR_2 in the following

[7]See Norton (2012) for a clear distinction between these two kinds of reduction.

way:

> *LR_2*: A theory T_B (secondary theory) reduces to another T_A (fundamental theory), iff the values of the relevant quantities of T_A *become* the values of the corresponding quantities of T_B by performing a limit operation on T_A.[8]

According to Nickles, the motivation for this type of reduction is heuristic and justificatory. The development of the new (or fundamental) theory T_A is motivated heuristically by the requirement that, in the limit, one obtains the same values as the predecessor (or secondary) theory T_B for the relevant quantities. As such, T_A is also justified as it can account adequately for the domain described by T_B. Nickles is also emphatic in pointing out that this kind of reduction is to be distinguished from *$reduction_1$*, which, as I said above, corresponds to Nagelian reduction. He clarifies that whereas Nagelian reduction requires the old (or secondary) theory to be embedded entirely in the new theory, limiting reduction only requires that the two theories make the same predictions for the relevant quantities when a limiting operation is performed. In this way, *$reduction_2$*, in contrast to *$reduction_1$*, does not require the logical derivation of one theory from another and, therefore, does not require logical consistency between the two theories (Nickles, 1973, p. 186). Since Nickles' *$reduction_2$* does not make any reference to explanation, logical deduction or the ontological status of reduction, which are aspects of more standard philosophical conceptions of reduction, this type of reduction

[8] Nickles (1973) inverts the order of "reducing" theory and theory "to be reduced" used by philosophers. According to him, the "reducing theory" is the theory that results from the limit operation and the theory "to be reduced" is the theory in which the limit operation is performed. This terminology is motivated by the way in which physicists use the term "reduce to". Since this notation is not relevant for Nickles' general concept of limiting reduction, I will use the term according to the philosophers' jargon and not following Nickles' terminology.

is often regarded as the "physical sense" of reduction (e.g. Nickles 1973; Rohrlich 1988; Batterman 2016).

2.3.2 Beyond Nickles' Concept of Reduction

In order to evaluate potential cases of limiting reduction, it is useful to have a formal definition at hand. Batterman (2016) advances such a definition by proposing the following schema (which he calls *Schema R*, henceforth SR):

> SR : A theory T_B reduces asymptotically to another T_A iff:

$$\lim_{x \to \infty} T_A = T_B,$$

where x represents a fundamental parameter appearing in T_A. T_A is generally taken as the fundamental theory and T_B is typically taken as a secondary or coarser theory.[9] For Batterman, the relation between two theories can be called "reductive" if the solutions of the relevant laws of the theory T_A smoothly approach the solutions of the corresponding laws in T_B, or in other words, if the "limiting behavior" of the relevant laws, with $x \to \infty$, *equals* the "behavior in the limit", where $x = \infty$.

It could be objected, however, that Batterman's *Schema R* is not precise enough since, strictly speaking, the limit is taken on functions representing quantities (or properties) of a theory rather than on the theory itself. Moreover, even if two functions representing the same physical quantity in T_A and T_B respectively coincide when a limit is taken, that does not guarantee the reduction of an entire theory to another. In fact, it might be possible for the functions representing a given quantity in the fundamental and secondary theory to be related by limiting reduction while for another quantity the

[9] In the original formulation, Batterman (2016) defines schema R, using $\epsilon \to 0$ instead of $x \to \infty$. For consistency with other parts of this paper, I instead express schema R as considering the limit to infinity $x \to \infty$. Whether one formulates $x \to \infty$ or $\epsilon \to 0$ does not make a difference in the content of this schema.

corresponding functions fail to do so. A more precise definition of limiting reduction, formulated only in terms of the quantities to be compared, is as follows:

> LR_3: A quantity Q^B of T_B reduces asymptotically to a quantity Q^A of T_A if:

$$\lim_{x \to \infty} {Q_x}^A = Q^B,$$

where x represents a parameter appearing in T_A, on which the function representing ${Q_x}^A$ depends. According to this definition, one is thus allowed to call a relation between quantities "reductive" if the values of the quantity ${Q_x}^A$ smoothly approach the values of the quantity Q^B when the limit $x \to \infty$ is taken. Naturally, in order to obtain the reduction of one theory to another, one would require that the values of *all* the physically significant quantities of the reduced theory coincide with the values of the quantities of the fundamental theory under certain conditions.[10] Proving this in every case is a huge enterprise, but note that, according to the above framework, the failure of reduction of one of the relevant quantities suffices to infer the failure of reduction of an entire theory to another. As it will be seen in the next section, this is exactly what is at stake in the case of phase transitions.

Before going there though, more specifications regarding the concept of limiting reduction are necessary. For example, it can still be argued that definition LR_3 is far too strict since it requires that the values obtained by performing a limit operation on a quantity ${Q_x}^A$ are exactly the same as the values of Q^B. In most cases this condition is not satisfied. Take, for instance, the concept of entropy as it is defined in thermodynamics and in Bolzmannian statistical mechanics. In thermodynamics, such a quantity

[10]Note, however, that here we assume that the two quantities have some qualitative features in common that make them candidates for reduction. An important topic that deserves to be addressed in future research regards the issue of whether quantitative coincidence suffices to infer correspondence between two quantities of different theories.

reaches its maximum value at equilibrium and does not allow for fluctuations. In contrast, Bolzmannian entropy is a probabilistic quantity that fluctuates every now and then even when the system has reached equilibrium. Cases like this motivated many authors (including Nickles himself) to allow for "approximate reduction". Accordingly, one can reformulate LR_3 as follows :

> LR_4: A quantity Q^B of T_B reduces asymptotically to a quantity Q^A of T_A if:

$$\lim_{x \to \infty} {Q_x}^A \approx Q^B,$$

where "≈" means "approximates", "is similar to", or "is analogous to". This means that a quantity ${Q_x}^A$ reduces another quantity Q^B if the values of ${Q_x}^A$ *approximate* the values of Q^B when the limit $x \to \infty$ is taken.

2.4 Are Continuous Phase Transitions Incompatible with Reduction?

In order to judge whether phase transitions correspond to a case of reduction, one needs to specify which quantities of T_A and T_B are expected to display the same values when a certain limit is taken. Subsequently, one needs to evaluate whether these quantities relate to each other according to the definitions provided in the previous section.

In both first-order and continuous phase transitions one is interested in comparing quantities of statistical mechanics with quantities of classical thermodynamics, where statistical mechanics is taken as the reducing theory T_A and classical thermodynamics as the theory to be reduced T_B. As it was shown in Section 2.2, in the case of first-order phase transitions one takes the thermodynamic limit to obtain the singularities in the derivatives of the free energy that successfully describe the phenomenon in thermodynamics. Following definition LR_4, one will say that the derivatives of the free energy in thermodynamics are reduced to the corresponding quantities in statistical mechanics if

$$\lim_{N\to\infty} F_N^{SM} \approx F^{TD},$$

where F^{SM} represents a derivative of the free energy as defined in statistical mechanics and F^{TD} the corresponding quantity in thermodynamics.

The case of continuous phase transitions is different, because, in general, one is not interested in computing the free energy but rather in calculating the universal quantities, like the critical exponents, and in explaining universality. In other words, one uses the thermodynamic limit and the infinite iteration limit to calculate the critical exponents that control the behavior of the system close to the critical point.

2.4.1 The Problem of "Singular" Limits

The view that phase transitions are not a case of limiting reduction has been most notably developed by Batterman (2001; 2005; 2011). He argues that this is a consequence of the "singular" nature of the thermodynamic limit.[11]

Using Batterman's terminology, a limit is singular "if the behavior in the limit is of a *fundamentally different character* than the nearby solutions one obtains as $\epsilon \to 0$" (Batterman, 2005, p. 2). According to him, the thermodynamic limit is *singular* in this sense because no matter how large we take the number of particles N to be, as long as the system is finite, the derivatives of the free energy will never display a singularity. As a consequence, he says that taking the limit of the free energy of finite statistical mechanics F^{SM} does not allow us to construct a model or theory that approximates the thermodynamic behavior.

> The idea that we can find analytic partition functions that "approximate" singularities is mistaken, because the very notion of

[11] Similar views are also held by Rueger (2000) and Morrison (2012).

approximation required fails to make sense when the limit is singular. The behavior at the limit (the physical discontinuity, the phase transition) is qualitatively different from the limiting behavior as that limit is approached (Batterman, 2005, p. 14).

This means that phase transitions would not even satisfy definition LR_4 stated in Section 4.

Although Batterman's argument is plausible, Butterfield (2011) (and Butterfield and Buoatta (2011)) challenges his reasoning using the following mathematical example. Consider the following sequence of functions:

$$g_N(x) = \begin{cases} -1 & \text{if } x \leq -1/N \\ N_x & \text{if } -1/N \leq x \leq 1/N) \\ 1 & \text{if } x \geq 1/N \end{cases}$$

As N goes to infinity, the sequence converges pointwise to the discontinuous function:

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

If one introduces another function f, such that

$$f = \begin{cases} 1 & \text{if } g \text{ is discontinuous} \\ 0 & \text{if } g \text{ is continuous} \end{cases}$$

then one will conclude, in the same vein as Batterman, that the value of f_∞ at the limit $N = \infty$ is *fundamentally different* from the value when N is arbitrarily large but finite. However, Butterfield warns us that if we look at the behavior of the function g, we will see that the limit value of the function is approached smoothly and therefore that the limit system is not "singular" in the previous sense.

According to Butterfield, this is exactly what happens with classical phase transitions and, for the cases analyzed here, he seems right.[12] Consider again the paramagnetic-ferromagnetic transition discussed in Section 2.1. This transition is characterized by the divergence of a second derivative of the free energy - the magnetic susceptibility χ - at the critical point. If we introduce a quantity that represents the divergence of the magnetic susceptibility and attribute a value 1 if the magnetic susceptibility diverges and 0 if it does not (analogously to the function f in Butterfield's example), then we might conclude that such a quantity will have values for the limit system that are considerably different from the values of the of systems close to the limit, i.e. for large but finite N. As a consequence, we will say that definition LR_4 fails. However, if we focus on the behavior of a different quantity, namely the magnetic susceptibility itself χ, we will arrive at a different conclusion. In fact, as N grows, the change in the magnetization becomes steeper and steeper so that the magnetic susceptibility smoothly approaches a divergence in the limit (analogous to the function g). This result is important because it tells us that definition LR_4 holds:

$$\lim_{N\to\infty} {\chi_N}^{SM} \approx \chi^{TD},$$

where χ^{SM} and χ^{TD} are taken as the magnetic susceptibility in statistical mechanics and thermodynamics respectively. The existence of finite statistical systems whose quantities approximate qualitatively the thermodynamic quantities for the case of first-order and continuous phase transitions has been also corroborated by Monte Carlo simulations (I will come back to this

[12] Even if Butterfield aims to make a more general claim, this does not hold for all cases of "singular" limits. Landsman (2013) shows that for the case of quantum systems displaying spontaneous symmetry breaking and the classical limit $\hbar \to 0$ of quantum mechanics, the situation is different and much more challenging. It seems therefore that the analysis of singular limits and the way of "dissolving the mystery" around them should be done on a case-by-case basis.

point in Section 6).

The important lesson from Butterfield's argument is that the "singular" nature of the thermodynamic limit does not imply that there are no models of statistical mechanics that approximate the thermodynamic behavior of phase transitions, for N sufficiently large but finite. If we assume that inter-theory reduction is consistent with the fact that the quantities of the secondary theory are only approximated by the quantities of the fundamental theory (as suggested by schema LR_4), then we arrive at the important conclusion that the "singular" nature of the thermodynamic limit is not *per se* in tension with the reduction of phase transitions.

One needs to be cautious, however, in not concluding that the previous argument solves all the controversy around the reduction of phase transitions. First of all, it is important to bear in mind that we are referring only to classical phase transitions and that quantum phase transitions have not been considered.[13] Second, one needs to note that we have not considered the use of renormalization group methods yet, in which there are two infinite limits involved. This is precisely the issue that we are going to address next.

2.4.2 Implementing RG Methods

As was shown in Section 2, the inference of the thermodynamic behavior of continuous phase transitions generally requires the appeal to RG methods. Batterman (2011) has suggested that the assumption of RG methods imposes a further challenge for the project of reducing phase transitions to statistical mechanics. He attributes this difficulty to the need for the thermodynamic limit in the inference of fixed point solutions, which are said to be necessary for the computation of critical exponents and for giving an account of universality. He claims (2011, p. 23):

> Notice the absolutely essential role played by the divergence of the

[13]For an analysis of quantum phase transitions see Landsman (2013).

> correlation length in this explanatory story. It is this that opens up the possibility of a fixed point solution to the renormalization group equations. Without that divergence and the corresponding loss of characteristic scale, no calculation of the exponent would be possible.

Why is it that the thermodynamic limit appears to be so important in the inference of non-trivial fixed points? The reason is that in every finite system there will be a characteristic length scale associated to the size of the system. Therefore, the application of a coarse-graining transformation beyond that length will no longer give identical statistical systems and the "RG flow" will inevitably move towards a trivial fixed point, with values of the coupling constants either $K = 0$ or $K = \infty$.

Figure 3 describes a contour map sketching the topology of the renormalization group flow and serves to illustrate the previous situation. Here the RG flows are represented by the trajectories R and D in a space S of Hamiltonians. Each point in this space represents a physical system described by a particular Hamiltonian associated with a set of coupling constants K. In this topology, the elements of S can be classified according to their correlation lengths ξ. Therefore, one can define surfaces containing all Hamiltonians $H \in S$ with a given correlation length. For example, the critical surface describes the set of all Hamiltonians with infinite correlation length $\xi = \infty$. In the figure, p represents a system with a Hamiltonian that inhabits the critical surface $\xi = \infty$, whereas s represents a system with a Hamiltonian that is infinitesimally close to p but is not on the critical surface; p^* and p_0 are fixed points. As one can see, the trajectory starting from s will stay close to trajectory R, describing a system at criticality, but eventually will move away towards a trivial fixed point. This follows because in a finite system the RG transformation will constantly reduce the value of the correlation length, moving the system away from criticality and resulting in a system with trivial values of coupling constants. As a result, two neighbor systems will approach far away fixed-points when a RG transformation is repeated infinitely many times, i.e. when $n \to \infty$, and therefore the two neighbor systems will approach two different limiting distributions describing physically diverse systems. Since the values of the critical exponents can be calculated by linearizing around non-trivial fixed points, this naturally means that iterating the RG transformation infinitely many times in a finite system will lead us to a fixed point from which one will be able neither to compute the critical exponents nor to give an account of universality. Taking into account that the critical exponents describe the behavior of the physical quantities Q close to the critical point, one can formally express this fact as follows. For N being arbitrarily large but finite:

$$\lim_{n\to\infty} {Q_{N,n}}^{SM} \not\approx Q^{TD},$$

where n is the number of iterations, Q^{SM} represents a quantity of statistical mechanics controlled by the critical exponents, whereas Q^{TD} represents the corresponding quantity in thermodynamics whose values match with the experimental results.

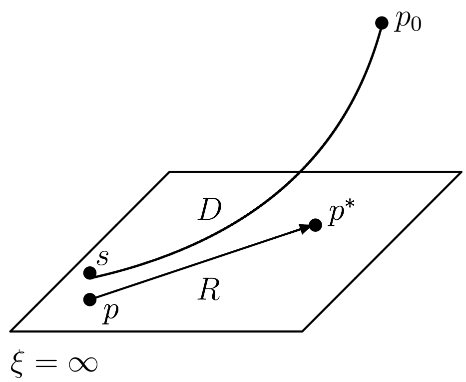

Figure 3: Contour map sketching the topology of the renormalization group flow (R). s and p represent systems infinitesimally close to each other. p^* is a critical fixed point and p_0 is a trivial-fixed point.

This is what led Batterman and others, for example Morrison (2012), to stress the importance of the thermodynamic limit. In fact, one can see from the argument given above that only systems with infinite correlation length (associated with a loss of characteristic scale) will approach non-trivial fixed points after infinite iterations of the RG transformation. The point that these authors do not emphasize is, however, that it is by taking the *infinite iteration limit* $n \to \infty$ that one approaches trivial fixed points from which one can neither explain universality nor calculate the critical exponents. If one realizes this, then the question that arises is whether in a finite system one can recover the experimental values of the critical exponents only after a finite number of iterations of the renormalization group transformations, i.e. without taking the second limit. This will be addressed in the next section.

2.5 Approximation, Topology and the Reduction of Continuous Phase Transitions

Before assessing the reducibility of continuous phase transitions, let us discuss the notion of approximation involved in the concept of limiting reduction. In the definition suggested by Nickles (and also in the revised versions mentioned in Section 4), there is implicit a precise criterion of approximation given by the convergence of the values of quantities in the fundamental theory to the values of the corresponding quantities in the secondary theory (See also Scheibe 1998, Hüttermann and Love 2016, Fletcher 2015).[14] We saw that, in cases where the quantitative and qualitative behavior of phase transitions can be inferred solely by taking the thermodynamic limit, this criterion of approximation well captures the idea of the reducibility of the quantities that describe phase transitions. The cases mentioned in Section 4.1 are examples of this.

Unfortunately, one cannot use the same criterion of approximation in cases in which taking the thermodynamic limit is not sufficient to infer the thermodynamic behavior. The reason is that, as we saw, in the case of continuous phase transitions one generally infers the thermodynamic behavior and explains universality only after performing a second limiting operation, which consists of applying repeatedly an RG transformation in the parameter space until the trajectory converges towards a non-trivial fixed point. Such a convergence does not, however, give us the criterion of approximation that can be used to determine whether phase transitions are a case of reduction. This is because when we ask about reduction, we are interested in analyzing the behavior of finite systems. Instead, the points of the RG trajectory describing a system at criticality are confined to the critical surface, corresponding to points with infinite correlation length $\xi = \infty$, and that does not give us any information about the behavior of finite systems.

[14] The convergence involved in limiting relations is generally pointwise and not uniform.

The challenge that the reductionist needs to face is that every point in a space of coupling constants that describes a system with finite correlation length will approach a trivial fixed point when the infinite iteration limit is taken. In this sense, if one sticks to the criterion of convergence to establish similarity or approximation between different physical quantities, one will conclude that the values of the quantities of statistical mechanics do not approximate the values of thermodynamic quantities. As a consequence, and in agreement with Batterman (2011), one would claim that limiting reduction fails for the case of continuous phase transitions.

But, what forces us to understand approximation only in terms of convergence towards a certain limit? Imagine that we could delimitate a region in the neighborhood of a fixed point p^*, as illustrated in Figure 4. Imagine further that we could show that the RG trajectory D generated by a finite system s intersects the region U around the fixed point p^*, after a large but finite number of iterations. Finally, imagine that linearizing around a point d' of the trajectory D which resides inside the region U allows us to calculate, at least approximately, the experimental values of the critical exponents. Could we say, then, that we have succeeded in deriving, at least approximately, the experimental values of the physical quantities from finite statistical mechanics? I think we could. Let me now show that this is actually the case.

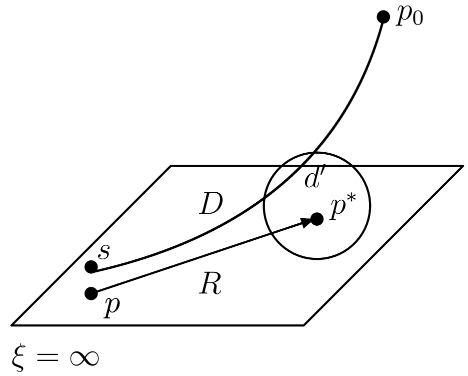

Figure 4: The region around the fixed point p^* represents neighboring points.

Wilson and Kogut (1974, Sec. 12) demonstrated by using ϵ-expansion approximation that in principle, and for an idealized case, if one starts from a point which is close enough to the critical surface, the RG trajectory will move close to the critical trajectory until it reaches the vicinity of a non-trivial fixed point p^*.[15] Once the trajectory reaches the neighborhood U of the fixed point p^* will stay there for a long time (which means, for repeated iterations of the RG transformation), thereby acting as it were a fixed-point. Finally, as $n \to \infty$, the trajectory will eventually move away from that region approaching a trivial fixed point.

What is relevant for us is that within the neighborhood U of the fixed point linearization is indeed possible, which implies that from a finite system one can obtain the values of the critical exponents after a finite number of iterations of the RG transformation. In order to derive accurate values of

[15] The ϵ-expansion is an asymptotic expansion for which ϵ takes values from $\epsilon = 1$ to $\epsilon << 1$. Since the exponents are not analytic at $\epsilon = 0$ one faces convergence problems which are treated by sophisticated summation methods that are nowadays under control.

the critical exponents, the number of iterations of the RG transformations should be large enough so that all details which are not universal, namely all details specific to a model, are washed out. If the number of iterations is not large enough the coupling constants will be sensitive to details of the model and the calculations of critical exponents will not be accurate (For details see also Le Bellac 1998).

If the ultimate goal of limiting reduction is to justify the fundamental theory by showing that the relevant quantities display values that approximate the values of the secondary theory, then, based on the previous argument, we have good reason to say that the quantities that describe continuous phase transitions in thermodynamics reduce to the quantities that describe the same phenomena in statistical mechanics, at least in this idealized case.

The formal expression that describes reduction in this particular case is as follows:

> LR_5**:** A physical quantity Q^{SM} in statistical mechanics reduces asymptotically to the analogous quantity Q^{TD} in thermodynamics, if for N sufficiently large:
>
> $$\exists n_0 \text{ such that } Q^{SM}_{N,n_0} \approx Q^{TD},$$

where n_0 corresponds to a finite range of iterations of the RG transformation. It should be noticed that the values of Q^{SM}_{N,n_0} also approximate $\lim_{n\to\infty} \lim_{N\to\infty} Q^{SM}$, which represent the values of the given quantity after taking both the thermodynamic limit and the infinite iteration limit.

One might object that the results obtained in this section rely too much on an idealized case and that in actual practice things are more complicated. Although it is true that in practice things are less straightforward, numerical simulation gives an important support for what has been said here. Since 1976 there have been attempts to use the numerical Monte Carlo simulation in the framework of renormalization group methods for the study of critical exponents. The first contribution in this direction was made by Ma (1976), who suggested an application of real space RG methods that required the calculation of the renormalized Hamiltonians. However, since calculating the renormalized couplings accurately enough proved to be too difficult, this approach did not succeed in determining the fixed point Hamiltonian with significant precision. Pawley, Swendsen, and Wilson (1984) made further progress in this direction by suggesting an approach based on expectation values of the correlation functions that did not rely on the calculation of renormalized Hamiltonians. Using this approach, they showed that for an Ising square lattice with 64 number sites, the system approaches the behavior of an infinite system after two iterations of a RG transformation. After more iterations, however, the system was shown to depart from the expected results flowing towards a trivial fixed point. A plausible explanation for this cross-over was that after more iterations the correlation length became comparable to the size of the system and finitary effects became relevant.[16]

One should bear in mind, however, that for some models the convergence is not as rapid as for the 2D-Ising lattice. Therefore, in order to avoid finite size effects in the renormalized systems, one should use large lattices. In the past years there has been significant improvement in this direction. See, for example, Hsiao and Monceau (2002) and Itakura (2003).

2.6 Concluding Remarks

The arguments presented in this paper give us good reason to think that the appeal to the infinite limits in the theory of phase transitions does not represent a challenge for reduction, at least not for limiting reduction. In fact, contra what has been argued by Batterman (2001, 2009) and Morrison (2012), these arguments suggest that the infinities and divergences characteristic of the physics of phase transitions are not essential for giving an account of the phenomena since from finite statistical mechanics one can recover the

[16]This is also pointed out by Butterfield (Butterfield, 2011, p. 69).

thermodynamic behavior of phase transitions even in the case of continuous phase transitions, as it was shown in section 5.

Nevertheless, this does not mean that phase transitions are not inconsistent with other notions of reduction that have also been discussed in the philosophical literature. Norton (2013), for instance, correctly points out that the case of continuous phase transitions does not satisfy what he calls "few-many reduction", according to which there will be a reduction if the behavior of a system with a few components can be used to explain the behavior of a system with a large number of them. The reason for this is that continuous phase transitions are intrinsically fluctuation phenomena that can only arise when N is sufficiently large.

Likewise, continuous phase transitions also seem to be at odds with the kind of reductive explanation that requires the explanans to give us accurate and detailed information about the microscopic causal mechanisms that produce the phenomenon (e.g. Kaplan (2011)). As it has been pointed out by Batterman (2002), Batterman and Rice (2014) and Morrison (2012), the impossibility of giving such an account is related with the robustness of the fixed point solutions under different choices of the initial conditions. This implies that the critical behavior is largely independent of specific microscopic details characterizing the different models and that the statistical mechanical account of phase transitions does not give us complete information about the microscopic mechanisms underlying the transitions. However, as it was shown in the paper, these senses in which reduction "fails" do not threat the project of inter-theory reductionism in any relevant sense.

Chapter 3

Had We But World Enough, and Time... But We Don't!

Justifying the Thermodynamic and Infinite-time Limits in Statistical Mechanics

> "The divergent series are the invention of the devil, and it is a shame to base on them any demonstration whatsoever" [N. H. Abel 1828]

3.1 Introduction

"Had we but world enough, and time" are the words with which Andrew Marvell begins his passionate poem in which he tells his lover that things would be different if they had infinite space and time. While neither the number of particles in real systems nor the time of measurements are infinite, it is common in statistical mechanics to take the number of particles and time to infinity in order to recover the values of thermodynamic observables. These are called the thermodynamic limit and the infinite-time limit, respectively. This raises the following questions: What justifies the empirical adequacy of scientific models that involve infinite limits? And what is the justification that we have for applying such a theory to finite systems? Sure enough, there would be a straightforward justification for the limits if one could show that, at least for the purpose of inferring the values of the thermodynamic observables, the infinite case is rather similar to the finite case (contrary to the situation described by Marvell!). But, is this so?

As it was seen in the previous chapter, there has been a fervent controversy around the use of the thermodynamic limit in the statistical mechanical treatment of phase transitions, in which has been claimed by some authors (e.g. Batterman 2005, Jones 2006, Batterman 2011, Bangu 2009, Bangu 2011) that the use of the thermodynamic limit – and so of an infinite system – is indispensable to give an account of phase transitions. As a consequence, it has been said that the behavior in the limit is physically real (Batterman 2005) or that phase transitions are not reducible to statistical mechanics (e.g. Batterman 2011, Bangu 2011, Morrison 2012). Others (e.g. Butterfield 2011, Butterfield and Buoatta 2011, Norton 2012) have argued against these conclusions saying that the thermodynamic limit can be justified straightforwardly, because the thermodynamic limit gives an approximate description of the behavior of real systems. They generally arrive at that conclusion by saying that the thermodynamic limit satisfies what Landsman (2013) calls Butterfield's principle, according to which a limit is justified and can be regarded as mathematically convenient and empirically adequate if the same behavior that arises in the limit also arises, at least approximately, "on the way to the limit".

In this chapter, I will take the side of the ones that believe that there is a straightforward justification for the thermodynamic limit, but I will argue against the idea that the so-called "Butterfield Principle" is sufficient to give a straightforward justification for the use of infinite limits in general. I arrive at that conclusion by comparing the use of the thermodynamic limit in the theory of phase transitions with the infinite-time limit in the explanation of equilibrium states, which has generally been left aside from the recent philosophical debate around the use of infinite idealizations in statistical mechanics. In the case of phase transitions, I will argue (Section 3.2) that the thermodynamic limit can be justified pragmatically, since the limit behavior also arises before we get to the limit and for a number of particles N that is physically significant. However, I will contend (Section 3.3) that the justification of the infinite-time limit is less straightforward. In fact, I will point out that even in cases where one can recover the limit behavior for finite time t, i.e. before we get to the limit, one fails to recover this behavior for realistic time scales. In my view this leads us to reconsider the role that the rate of convergence plays in the justification of infinite limits in general and calls for a revision of the so-called Butterfield's principle. I will end this paper (Section 4.4) by offering a criterion for the justification of infinite limits based on the notion of controllable approximations.

3.2 The Thermodynamic Limit in the Theory of Phase Transitions

In recent years, phase transitions have captured the attention of philosophers of science mainly because there seems to be an eliminable appeal to the thermodynamic limit in the statistical mechanical treatment of these phenomena. In this section, I will explain the apparent need for the thermodynamic limit and I will argue – in the same vein as Butterfield (2011)– that, despite some claims about the "singular nature" of the thermodynamic limit, this idealization can be justified pragmatically.[1]

[1] Since the goal here is to relate the problem of the thermodynamic limit in the theory of phase transitions with the infinite-time limit in the explanation of equilibrium, I will be deliberately brief in my exposition of the problem of phase transitions. A more detailed treatment of these topics can be found in the previous chapter as well as in Kadanoff (2009), Butterfield (2011), Batterman (2001), and Butterfield and Buoatta (2011).

3.2.1 The Problem of Phase Transitions

In thermodynamics, phases correspond to regions of the parameter space where the values of the parameters uniquely specify equilibrium states. Phase boundaries, in contrast, correspond to values of parameters at which two different equilibrium states can coexist. The coexistence of different equilibrium states at phase boundaries expresses itself as discontinuities of thermodynamic quantities, which are related to the first derivatives of the free energy with respect to a parameter such as pressure or temperature. If the system intersects a phase boundary when going from one phase to another, i.e., encounters a discontinuity in a macroscopic observable, the system is said to undergo a first-order phase transition. If the system moves from one phase to another without intersecting a phase boundary, the system is said to undergo a continuous phase transition, in which case there are no discontinuities involved in the macroscopic observables, but there are divergences in the second derivatives of the free energy.

In the statistical mechanical treatment of phase transitions, which is generally constructed on the basis of Gibbs' canonical ensembles, one can describe phase transitions in terms of discontinuities or divergencies of the free energy by invoking the thermodynamic limit. However, it appears that one cannot do so without the infinite limit. In fact, in the canonical ensemble, the free energy is defined as the logarithm of the partition function Z:

$$F = -k_B T \ln Z, \tag{3.1}$$

where k_B is the Boltzmannian constant. The partition function is the sum over all states accessible to the system:

$$Z = \sum_i e^{\beta H_i}, \tag{3.2}$$

where $\beta = \frac{1}{k_b T}$ and H_i is the Hamiltonian associated to a particular microstate i. Since the Hamiltonian is usually a non-singular function of the degrees of freedom, it follows that the partition function is a sum of analytic functions. As a consequence, neither the free energy nor its derivatives can have the singularities that characterize phase transitions in thermodynamics. Taking the thermodynamic limit, which consists of letting the number of particles and the volume of the system go to infinity, i.e., $N \to \infty$, $V \to \infty$, in such a way that N/V remains constant, allows one to recover those singularities and provide a rigorous definition for the phenomena that turns out to be empirically adequate.

Since we assume that real systems have a finite number of degrees of freedom, the question that arises is how can one justify the empirical adequacy of the statistical mechanical treatment of phase transitions, notwithstanding the fact that we know that it relies on an infinite idealization. One might think that what explains the success of the theory is that it provides us with a mathematical model that approximates the behavior of finite systems. Following this line of reasoning, one might assume that the quantities that successfully describe phase transitions in the thermodynamic limit (in this case, the derivatives of the free energy) approximate the values of the quantities before we get to the limit, i.e. for finite and large N, and, moreover, that they do so for *realistic* values of N. If this were actually the case, one would have good reason to conclude that the justification for both the success of the theory and the infinite idealization are straightforward. Moreover, we would have good reason to justify the use of the limit as mathematically convenient and empirically adequate, which is what Butterfield (2011) calls "a straightforward justification of the limit".

Unfortunately, the previous reasoning faces at least three difficulties that prevent us from arriving at that conclusion as quickly as we would expect.

1. The first difficulty, pointed out most notably by Batterman in a series of papers (2002, 2005, 2011), concerns the so-called "singular nature" of the thermodynamic limit. According to Batterman, a limit is singular "if the behavior in the limit is of a *fundamentally different character*

than the nearby solutions one obtains as $\epsilon \to 0$" (2005, p. 2), where $\epsilon \to 0$ is taken as the "limiting behavior". Batterman argues that the thermodynamic limit is *singular* in the previous sense because even if we take N to be arbitrarily large, as long as it is finite, the derivatives of the free energy will never display a singularity. It is important to note that he arrives at that conclusion by assuming that the singular behavior of a quantity is qualitatively different from its analytic behavior. As will be seen in the next section, this assumption is far from trivial.

2. The second difficulty regards the apparently essential role of the thermodynamic limit in the renormalization group approach. In order to give an account of the quantitative behavior of continuous phase transitions, it was necessary to incorporate renormalization group (RG) techniques. These techniques rely on the existence of non-trivial fixed points, which are points in a space of Hamiltonians at which different renormalization trajectories arrive after repeated iterations of a renormalization group transformation (details elsewhere, e.g, in Goldenfeld 1992, Wilson and Kogut 1974). It has been claimed (Batterman 2011, Morrison 2012) that the thermodynamic limit is "ineliminable" in this approach, because no matter how large we take N to be, as long as it is finite, the RG trajectory will not converge towards a non-trivial fixed point. This is supposed to follow from the fact that finite systems cannot display a divergence in the correlation length and therefore cannot present a loss in the characteristic length scale, which is necessary to define non-trivial fixed points in the space of Hamiltonians.

3. The third difficulty is the problem of generality. Even if we could show that in some cases the values of the quantities that successfully describe phase transitions in the limit "$N = \infty$" approximate the values of the quantities evaluated for large but finite N, there remains the question

> of whether this is so in all cases in which the thermodynamic limit is used to describe the phenomena of phase transitions. Landsman (2013) argues, for instance, that for the case of quantum systems displaying spontaneous symmetry breaking and the classical limit $\hbar \to 0$ of quantum mechanics, the situation is different and much more challenging than in classical phase transitions.

3.2.2 Butterfield's Principle and Butterfield's Solution to the Problem of Phase Transitions

The difficulties mentioned in the previous section have motivated controversial claims. For instance, it has been argued that the need for the thermodynamic limit in the theory of phase transitions and, especially, in the theory of continuous phase transitions imply the failure of the reduction of thermodynamics to statistical mechanics (Batterman 2011, Morrison 2012, Bangu 2009). Moreover, it has been argued that as a consequence of the "singular" nature of the thermodynamic limit, one should conclude that the singularities that describe phase transitions in the limit are *physically real* (Batterman 2005).

Independently of whether these conclusions actually follow from the problems pointed out above, the fact is that, in light of those difficulties, the empirical adequacy of the theory of phase transitions appears as conceptually puzzling and requires an explanation.

So the question is: can we restore a straightforward justification for the thermodynamic limit in the theory of phase transitions despite the objections mentioned above? Butterfield (2011) actually argued that we can. According to him, the thermodynamic limit is justified and can be conceived as mathematically convenient and empirically adequate because

> there is a weaker, yet still vivid, novel and robust behaviour that occurs before we get to the limit, i.e., for finite N. And it is this

weaker behaviour which is physically real. (p. 1065)

Here "novel and robust" represents the behavior that is novel and robust with respect to the behavior of systems with finite N: in the case of phase transitions that is the discontinuities and singularities in the derivatives of the free energy. And the word "weak " is meant to emphasize that the behavior that arises before one gets to the limit only approximates the behavior that is observed in the limit. In other words, Butterfield thinks that the limit is justified because the value of the relevant quantities before we get to the limit is close to the value of the corresponding quantities evaluated at the limit. In order to support his view, he presents a series of examples to show that the "qualitative" difference between the behavior of the relevant quantities in the limit and close to the limit is only apparent, since it is the consequence of focusing on the wrong quantities. Let me summarize his argument. Consider a sequence of functions:

$$g_N(x) = \begin{cases} -1 & \text{if } x \leq -1/N \\ Nx & \text{if } -1/N \leq x \leq 1/N) \\ 1 & \text{if } x \geq 1/N \end{cases}$$

As N goes to infinity, the sequence converges pointwise to the discontinuous function:

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

If one introduces another function f, such that

$$f = \begin{cases} 1 & \text{if } g \text{ is discontinuous} \\ 0 & \text{if } g \text{ is continuous} \end{cases}$$

then one will conclude that the value of f_∞ at the limit $N = \infty$ is *fundamentally different* from the value when N arbitrarily large but finite: $f_\infty \not\approx f_N$. Consequently we will conclude that the thermodynamic limit is "singular" in Batterman's sense. However, if one looks at the behavior of the function g, one will see that the limit value of the function is approached smoothly and therefore that the limit system is not "singular" in the previous sense. Thus, if one looks only at the quantity f, one will not be able to see what is revealed when one looks at the behavior of the quantity g, namely that the limit is actually an approximate description of the behavior before we get to the limit. According to Butterfield, this is exactly what happens with classical phase transitions, and, for typical examples of phase transitions, he seems right. Consider the paramagnetic–ferromagnetic transition in magnetic materials. This transition is characterized by the divergence of a second derivative of the free energy - the magnetic susceptibility χ - at the critical point. If we introduce a quantity that represents the divergence of the magnetic susceptibility and attribute a value 1 if the magnetic susceptibility diverges and 0 if it does not (analogous to the function f in Butterfield's example), then we might conclude that the limit quantities have values that are considerably different from the values of the quantities for arbitrarily large but finite N. However, if we focus on the behavior of a different property, namely the magnetic susceptibility itself χ, we will arrive at a different conclusion. In fact, the magnetic susceptibility χ is defined as the derivative of the magnetization with respect to an external magnetic field $\chi = \partial M/\partial H$. As N grows, the change in the magnetization becomes steeper and steeper, and the quantity smoothly approaches a divergence in the limit (analogous to the function g). This means that in statistical mechanics one can, in principle, find finite systems that have values of the magnetic susceptibility χ that approximate the thermodynamic behavior.

I take it as a moral of Butterfield's argument that the "singular" nature of a limit is not in conflict with a straightforward justification of the limit.

However, one needs to recognize that this only solves the first of the problems pointed out above and does not allow us to conclude that the same argument applies to other cases of phase transitions (problem (iii)), or to explain the role of the thermodynamic limit in renormalization group techniques (problem (ii)). This last problem was studied extensively in the first chapter and has been addressed also, for example, by Batterman 2011, Morrison 2012, Norton 2012 and Butterfield himself (Butterfield 2011, Butterfield and Buoatta 2011). Since I do not have space to discuss these other issues here, I will restrict my analysis to the cases in which numerical values for finite systems are available: the paramagnetic-ferromagnetic transition described above and the liquid-vapor transition at the critical point in which the compressibility behaves analogously to the magnetic susceptibility. The question that I want to raise here instead is whether, in order to justify the use of the thermodynamic limit, it is sufficient to show that the behavior of finite systems before we get to the limit (for large N) approximates the behavior in the limit, as Butterfield's principle prescribes. Moreover, I wish to discuss whether this can be used as a general principle for justifying the use of infinite limits in physics.

Although Butterfield (2011) does not consider this criterion as a general principle (at least not explicitly), Landsman (2012) does:

> Butterfield's Principle is the claim that in this and similar situations, where it has been argued (by other authors) that certain properties emerge strictly in some idealisation (and hence have no counterpart in any part of the lower-level theory), "there is a weaker, yet still vivid, novel and robust behaviour [...] that occurs before we get to the limit, i.e., for finite N. And it is this weaker behaviour which is physically real." (p. 383)

Likewise, Norton (2012) also seems to take this as a criterion when he suggests that most of the controversy around phase transitions is dissolved after one recognizes that this theory does not require idealizations (i.e. systems that provide inexact descriptions of the target system) but only approximations (i.e. inexact description of the target system) of the behavior of systems with *very large* number of particles. It is important to note, however, that if one wants to transform Butterfield's criterion into a *principle*, one needs to show not only that this criterion is necessary for giving a straightforward justification of infinite idealizations (which seems hard to deny), but also sufficient. In this respect, it is surprising that little attention has been given to the rate of convergence in the justification of infinite limits.

More to the point, if we assume that the limit is justified when we can prove that the idealized mathematical model is just an approximation of the behavior of realistic systems, it does not suffice to show that the behavior of phase transitions can be recovered for large but finite N, but it must also be shown that it is recovered for values of N that are physically significant, i.e. for $N \approx 10^{23}$. In the examples discussed here, it turns out that this is actually the case. For instance, the value of the magnetic susceptibility χ_N for $N \approx 10^{23}$ is approximately the same as the limit value $\lim_{N\to\infty} \chi_N$. Therefore, one can be confident that the idealized model for phase transitions is a good approximation of realistic systems. Butterfield (2011, p. 19) points this out, but he does not emphasize the importance of demonstrating that the infinite limit also provides a good approximation for realistic values of N, nor he includes this explicitly as a condition for the justification of the limit. Sure enough, in the examples of phase transitions he refers to, the values of the quantities for realistic N are so close to the values obtained in the neighborhood of the limit that distinguishing between such values does not seem to be crucial. However, this is not necessarily the case in other examples of infinite limits. Indeed, we will see next that in the infinite-time limit the values of the relevant quantities for very large but finite time can vary significantly from the values obtained for realistic t.

3.3 The Infinite-time Limit in the Ergodic Explanation of Equilibrium

The infinite-time limit, which consists in letting time go to infinity $t \to \infty$, has played an important role in statistical mechanics and, like the thermodynamic limit, has also been matter of controversy in the philosophical literature (e.g. Malament and Zabell 1980, Earman and Rédei 1996, Emch and Liu 2013, Sklar 1995).

In this section, I will first discuss the role of the infinite-time limit in the explanation of equilibrium in Gibbsian statistical mechanics and I will then expose the difficulties for giving a straightforward justification of the limit. Contrary to the case of the thermodynamic limit in the theory of phase transitions, I will argue that these difficulties are not related to whether or not one can recover the limit values of the relevant quantities for finite t, i.e. before we get to the limit, but rather to whether or not one can recover those values for realistic t. This will reveal the important role of the rate of convergence in the justification of infinite limits.

3.3.1 The Problem of the Infinite-Time Limit

In order to understand the use of the infinite-time limit in the Gibbs' framework, one needs to become familiar with the Gibbs formalism. The most important concept here is the notion of *ensemble*, defined as an infinite collection of systems governed by the same Hamiltonian but distributed differently over the phase space Γ. An ensemble can also be understood as a uniform probability distribution ρ over Γ, which reflects the probability of finding the state of a system in a certain region of Γ. The uniform probability distribution on an hypersurface Γ_E of this space Γ is referred to as the *microcanonical ensemble*, where the energy and the number of particles are constant. In the microcanonical ensemble, there is a phase function $f_p : \Gamma_E \to \mathbb{R}$ associated with each relevant physical quantity. The expectation values of those functions will correspond to *phase averages*, defined as follows:

$$\langle f \rangle_p = \int_{\Gamma_E} f_p \rho \, d\Gamma_E \tag{3.3}$$

Phase averages play an important role in this approach because they correspond to the values of the macroscopic quantities measured in experiments. In fact, if we measure the macroscopic quantities of a gas in equilibrium which is enclosed in some container, we will observe that these values coincide with the values predicted by Gibbs' phase averages, even if we do not have any information about the microscopic configuration of the gas.

The question that has puzzled physicists and philosophers of science is why phase averages coincide with values measured in real physical systems. The answer is not clear. First of all, this formalism is built upon the notion of *ensemble*, which is a fictional entity that does not make direct reference to the behavior of a single system. Second, phase averages do not tell us anything about the dynamics, i.e. they do not give us information about how the the system – at the microscopic level – behaves in time. Third, this formalism does not explain why the experimental values always correspond to the average values and are not spread around the mean.[2]

The most intuitive explanation for the success of phase averages consists of associating them with *time averages* $\langle f \rangle_t$. Time averages have a clearer physical meaning because they make reference to the fraction of time that the system spends in the regions of the phase space associated to the mean values of the macroscopic observables. In other words, if we assume that measurements take some time, then we might think that we succeed in measuring phase averages because they correspond to the average values that actually occur during the time of measurement. And here is when the infinite-time limit comes into scene. In order to associate phase averages with time aver-

[2]See Frigg 2008, Uffink 2007 and van Lith 2001 for a more detailed description of the problems associated with the Gibbs formalism.

ages, one generally needs to introduce the infinite-time limit. For example, the Birkhoff theorem tells us that if we define the invariant mean of time $\langle f \rangle_t$ of time dependent functions $f(t)$ as

$$\langle f \rangle_t = \lim_{T \to \infty} \frac{1}{T} \int_0^T f(t)\, dt, \tag{3.4}$$

it follows that for almost all sets (except on a set with measure zero):

1. $\langle f \rangle_t$ exists for every integrable function $f(t)$ in Γ_E.

2. If the system is ergodic, then $\langle f \rangle_p = \langle f \rangle_t$.

Note that in this approach, in order to derive the equivalence between phase averages and time averages $\langle f \rangle_p = \langle f \rangle_t$, one needs to assume that the system is ergodic, which means that as time evolves the dynamic trajectories pass through every point in Γ_E.[3] The assumption of ergodicity has been itself a matter of controversy in the foundations of statistical mechanics, but for the sake of brevity I will leave this discussion aside and focus instead on the appeal to the infinite-time limit for the justification of equilibrium.

The introduction of the infinite-time limit in the definition of time averages is far from trivial, especially if one thinks that the original motivation for relating phase averages with time averages is the belief that the latter have a clearer physical meaning. In fact, we know that measurements do not take an infinite amount of time: so, what is that justifies the use of the infinite-time limit in this context? One might try to give a straightforward justification for the limit along the lines of Butterfield's principle by saying that even if the measurement times are short with respect to human macroscopic scales, they are *very long* with respect to the microscopic time scales, i.e. time of

[3]Strictly speaking, this theorem was formulated in terms of metric transitivity instead of ergodicity. Metric transitivity is a property of dynamical systems that captures the same idea as ergodicity but in measure theoretic sense. For more details see Uffink 2007[sec.6], van Lith 2001[ch. 7]

collision between particles, and therefore they are well approximated by infinite time averages (One can find arguments in this direction, for example, in Gallavotti 1999, Emch and Liu 2013). If so, one might think that one has good reason to consider the infinite-time average as a mathematical model that approximate the values obtained in finite time measurements and will have good reason to give a pragmatic justification for it. For example, that it allows us to wash out fluctuations we deem irrelevant, that it is mathematically convenient and that it allows us not having to decide in advance how long the time of measurement should be. Unfortunately, there are difficulties that prevent us from arriving at this conclusion as quickly as we would like.

1. The first is that even if the limit defined in (4) exists, it does not mean necessarily that it describes a system in equilibrium. Uffink (2007, p. 92) expresses this difficulty pointing out that generally:

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(t)\,dt \neq \lim_{t\to\infty} f(t), \tag{3.5}$$

   where the right-hand side describes a constant value of a physical quantity $f(t)$ and the left-hand side represents an average value of the same quantity. Note that for periodical motions the left-hand side exists whereas the right-hand side does not.

2. Second, there is a problem related to the apparent indispensability of the infinite limit in the derivation of the equivalence between phase and time averages (this problem is similar but not equivalent to the problem of "singular" limits discussed above in the context of the thermodynamic limit). We saw that Birkhoff's theorem states that one can derive the equivalence between phase averages and time averages after taking the infinite time limit, but this theorem does not tell us anything about how these two averages are related for large but finite times. Frigg expresses this point as follows: "... the infinity is crucial.

If we replace infinite time averages by finite ones (no matter how long the relevant period is taken to be), then the ergodic theorem does not hold any more and the explanation is false." (2008, p. 147)

3. Finally, there is the difficulty that even if one can show that the limit in (4) converges, this does not imply that it converges rapidly enough to be empirically meaningful. Measurement times generally take a very short time with respect to human macroscopic time scales. Thus, in order to show that the infinite time average is a good approximation for finite time averages, one needs to prove that the infinite time average is approached within realistic measurement time scales.

In the reminder of this paper, I will focus mainly on problem (iii), because it is this problem that reveals the most important difference between the thermodynamic limit discussed in Section 2 and the infinite-time limit.

3.3.2 The Dog-Flea Model and a Straightforward Justification for the Infinite-Time Limit

In order to understand under which conditions one could give a straightforward justification for the infinite-time limit it is useful to consider a toy model. The toy model that can best help us to grasp these conditions is the Dog-Flea model, invented by Tatjana Ehrenfest-Afanassjewa and Paul Ehrenfest (Ehrenfest and Ehrenfest-Afanassjewa 1907). A version of the model is as follows. Consider two dogs, Poomba and Woori, that share a population of N fleas. Assume further that N is even and that the fleas are labeled by an index from 1 to N. The macroscopic observables of the model are n and m, representing the number of fleas in Poomba and Woori, respectively. A microscopic description of the system corresponds to the specification of the positions of all fleas in each dog. The time evolution of the system is described like this: At every second, a number from 1 to N is taken randomly from a bag and announced. When hearing its name, the corresponding flea jumps immediately from the dog it pestered to the other. The model predicts that in the long run ($t \to \infty$), and independently of the initial distribution, the process leads to a time-invariant distribution that is symmetric around the value $p = N/2$ and it is very peaked at that value, all the more so when N is large. It is important to emphasize that the model admits only one time-invariant probability distribution, which is the same as the distribution in classical probability theory that in a sequence of N trials of a fair coin, exactly p heads come up. In this way, the model illustrates quite nicely that under certain statistical assumptions, it is possible to obtain the properties that characterize equilibrium. And, analogously to the case described above, the equilibrium distribution is defined in the limit $t \to \infty$.

Following the strategy used in the previous section, we might think that the asymptotic distribution will approximate the behavior of a finite time measurement if the measurement time (macroscopic time scale) is *very long* with respect to the time that it takes for a flea to jump from one dog to the other (microscopic time scale), which is here one second. If this is the case, we might also say that the infinite-time limit is justified pragmatically, since it is mathematically convenient and it enables us to wash out fluctuations.

An advantage of the Dog-Flea model is that it allows us to perform computer simulations to test our hypothesis. Emch and Liu (2013, sec. 3.4) present the results of these simulations for two different time scales:

1. The first run consists of 10^2 iterations.

2. The second run consists of 10^4 iterations.

In both cases, the number of fleas is $N = 100$. Remarkably, even if the two macroscopic time scales are long with respect to the microscopic scale (a single iteration), the results for (a) are significantly different from the results obtained for (b). Whereas (a) exhibits values of n, m that are constantly changing, (b) exhibits equilibrium behavior (with chaotic fluctuations) that is in good agreement with the equilibrium distribution obtained in the infinite time limit.

Based on these results, we should conclude that the time invariant distribution (for $t \to \infty$) gives us a good approximation for the values of the macroscopic observables in (b) but not in (a). Accordingly, we can say that we are *justified* in using the limit distribution for describing the situation for (b), but not for (a). Note, that this justification is not related with whether or not the system approaches the equilibrium values in a finite time, but rather with whether or not the system approaches those values in a time that is short with respect to the time of measurement. This obliges us to consider the *convergence rate*, which represents the rapidity at which the limit is reached. In the first case (a), the convergence is not rapid enough. Indeed, the system will eventually approach equilibrium, in a long but finite time, but since this time is much longer than the measurement time, the average values of the observables will not coincide with the values predicted by the time invariant distribution. Therefore, the asymptotic average value will not provide a good approximation of the values measured during that time.

3.3.3 The Importance of the Rate of Convergence

For the present discussion, the important lesson of the Dog-Flea model is that talking about "long time" is useless unless we specify the relevant time scales of the problem under investigation. In this sense, if we want to justify the infinite-time limit in the explanation of equilibrium, it does not suffice to argue that the time of measurement is "very long" with respect to microscopic time scales, but rather we need to specify the rate of convergence and guarantee that the asymptotic value will be reached within the time scales that we are interested in. As one might suspect, specifying the convergence rate is not a trivial task. To give a more precise idea, let $\langle f(T) \rangle$ represent the average value calculated at time T, that is:

$$\langle f(T)\rangle = \frac{1}{T}\int_0^T f(T_t)dt. \tag{3.6}$$

Then in order to determine the convergence rate, one needs to find a finite $\epsilon(T)$ such that:

$$||\langle f(T)\rangle - \langle f\rangle_t|| \leq \epsilon(T), \tag{3.7}$$

where $\langle f\rangle_t$ is the time invariant mean defined in (4). Even in simple models, to obtain definite values of $\epsilon(T)$ is often difficult in both theory and practice, and to demonstrate that this value is very small, i.e, $\epsilon(T) \approx 0$, for realistic measurement times is even harder. More importantly, it is perfectly conceivable to have a situation in which the values of the functions are constantly changing so that the time needed to attain the time average is of the order of the recurrence time, i.e. the time necessary to visit the entire surface Γ_E. One can estimate that the recurrence time for a small sample of diluted hydrogen gas is unimaginably longer than the age of the universe, and this time is even longer if we consider more complicated systems. In situations like this, there might well exist a finite $\epsilon(T)$ that satisfies eq.(7). However, the time for which ϵ is sufficiently small will be much longer than realistic measurement times, which means that for realistic time scales T', say $2/10\,sec.$, $\langle f(T')\rangle \not\approx \langle f\rangle_t$.[4]

The previous argument just tells us that, even if we could demonstrate that the asymptotic average will be reached within finite but very large times (or in other words "on the way to the limit" as in Butterfield's principle), this does not imply that the asymptotic average will be reached for realistic t and, therefore, it does not imply that we can interpret the limit as giving us a good approximation of the systems that we are interested in. This has an important philosophical consequence because it tells us that the so-called "Butterfield's principle" is not sufficient to justify the limit in this case.

[4]For a quantitative estimation, see Gallavotti 1998

Boltzmann himself was aware of the problem of the rate of convergence in the justification of the infinite-time limit, and in order to reconcile this limit with the rapid approach to equilibrium, assumed that the "the macroscopic observables, had an essentially constant value on the surface of given energy with the exception of an extremely small fraction ϵ of cells" (1874 [quoted in Gallavotti 1999, p. 16]). Unfortunately, this assumption is not uncontroversial, and to some extent it does not really solve the problem. In fact, even if we accept the premise postulating, for example, that the functions satisfy symmetry conditions, we still need an argument to associate phase averages with time averages. In other words, we still need an argument that allows us to conclude that the system does not spend so much time in the small fraction of cells that differ from the mean phase values. Ironically, this seems to beg the question, in that it brings us back to the original problem for which the infinite time limit entered the picture, namely the problem of deriving the equivalence between phase and time averages.

Different alternatives have been offered in the literature to deal with this and the other problems associated with infinite time averages. Maybe the most radical was the proposal by Malament and Zabell (1980), where they argue that one can explain the empirical adequacy of phase averages without appealing to time averages at all. Their argument is based on two assumptions: i) the system exhibits small dispersion with respect to the phase average (analogously to Boltzmann's assumption), and ii) the microcanonical measure represents the probability of finding a system in a particular region of the phase space. According to them, these two assumptions taken together lead to the conclusion that the probability that phase functions are always close to their phase averages is very large, without making any reference to infinite time averages. Even if this view looks appealing, two main criticisms have been raised in the literature. The first is that in order to justify assumption (ii), they invoke a version of ergodicity, which is an hypothesis that has been questioned in the foundations of statistical mechanics (e.g. Earman and

Redei 1998, Frigg 2008, van Lith 2001). The second, which is more important for us, is that they justify assumption (i) based on Khinchin-Lanford dispersion theorems, which tell us that for functions that satisfy strong symmetry conditions, the dispersions from the mean will go to 0 in the thermodynamic limit. The appeal to the thermodynamic limit would not be problematic, if we could demonstrate that – like the case of phase transitions – there is a straightforward justification for it. Unfortunately, the use of the thermodynamic limit in this context appears to be less straightforward than in the case of phase transitions, because Butterfield's principle is not enough to justify the limit. In fact, for realistic $N \approx 10^{23}$, one can estimate, based on Khinchin's theorem, that the probability that there is a relative deviation from the mean of more than a tiny ϵ is very small, but not sufficiently small to discard that these states will occur in nature. This means that one cannot regard (at least not without risks) the asymptotic results obtained in this and other similar theorems as providing us with a good approximation of the behavior of realistic systems. This problem is also referred in the literature as the measure-epsilon problem (See Uffink 2007, van Lith 2001 and Frigg 2008).

An alternative approach can be found in Earman and Redei (1996). They do not invoke ergodicity for the explanation of the success of phase averages, but they are quite sympathetic towards the explanatory role of "ergodic-like behavior". According to them, ergodic-like behavior only requires weak mixing behavior with respect to a set of finite observables. It is important to note that the definition of mixing offered by them still requires the appeal to the infinite-time limit. Interestingly for what we are discussing here, they explicitly include *rapid convergence* as an additional condition for the explanation of equilibrium. To justify this assumption they suggest (although not necessarily endorse) two possible routes: a) The first is to make reference to matter-of-fact initial states. b) The second is to assume that systems are subjected to perturbations from outside that act as a kind of 'stirring' mechanism which rapidly drive the observed values of the macroscopic quantities.[5] Even if one should not discard that some progress can be done in each of these lines of research, one should recognize that they are methodologically complicated since they oblige us to the consider specific features of the systems of interest.

The explanation of the empirical success of phase averages is still an open problem in the foundations of statistical mechanics. Although there is some skepticism in the philosophical literature towards the idea of explaining this success via infinite time averages, the infinite-time limit continues playing an important role in physics. It is far beyond the scope of this paper to offer a final assessment for the appeal to the infinite-time limit in the explanation of equilibrium. However, it suffices for our purposes to have shown that much of the problems for providing a justification for such a limit come from the conceptual and methodological difficulties to specify the rapidity at which the limit is approached. I argued that this has an important consequence for the current philosophical literature on infinite limits, because it teaches us an important lesson about the role of the convergence rate in the justification of infinite limits.

3.4 Conclusion: Infinite Limits as Controllable Approximations

Although there is no consensus regarding the status of infinite limits in physics, it seems reasonable to interpret these idealizations as mathematical models that approximate the behavior of finite systems. The question that one needs to ask, however, is under which conditions are we allowed to arrive at that conclusion. In the debate on phase transitions, it is often assumed that we are allowed to interpret the infinite limit as providing an

[5] A review of this attempts can be found in Lanford 1973

approximation of finite systems as long as the behavior that arises in the limit also arises, at least approximately, "on the way to the limit", which is what we called here the " Butterfield principle". However, in this paper I argued that in the case of the infinite-time limit this condition is not sufficient to justify the limit. This is because in this case the values of the relevant quantities "before we get to the limit", that is for finite but very large t, can take values significantly different from the values obtained for realistic time scales t.

The above result leads us to a revision of Butterfield's principle that would apply more generally than the original formulation. A proposal is as follows:

> We can justify infinite limits, when $x \to \infty$, as being mathematical models that approximate the behavior of real finite systems, iff (i) the behavior that arises in the limit also occurs, at least approximately, before we get to the limit, i.e., for finite x., and (ii) it also arises for *realistic values* of x.

A concept that captures the main idea of the previous statement is the notion of *controllable* approximations. Emch and Liu (2002, p. 526) define controllable approximations as the ones in which the deviations of the model with respect to realistic systems can be quantitatively estimated. When no such estimation can be given, the approximation is said to be uncontrollable. Uffink (2007, p. 109) makes this notion more precise, suggesting that in the case of controllable approximations involving infinite limits one has control over how large the value of the parameter must be to assure that the infinite limit is a reasonable substitute for a finite system. Since we are interested in the behavior of realistic systems, I claim that this "control" should also involve a specification of the rate of convergence. This will allow us to warrant that the limit is reached for realistic values of the parameters and therefore that it is a good approximation of the target systems. In the cases of phase transitions analyzed in Section 3.2, the thermodynamic limit appears to be controllable in this sense. However, for what has been argued in Section 3.3, we do not seem to be in the position of deriving the same conclusion for the case of the infinite-time limit.

Chapter 4

Market Crashes as Critical Phase Transitions?

Reductive Explanations and Idealizations in Econophysics[1]

> "Essentially, all models are wrong, but some are useful" [George Box 1987]

4.1 Introduction

The success of formal methods to explain natural phenomena in physics prompts the question of whether similar methods can be applied to explain phenomena in social sciences. Or more specifically, whether the same mathematics employed in physics can be used to explain and predict phenomena in economics and politics. There are reasons to think that this is possible. In the last thirty years, a great number of models originally designed in the context of statistical mechanics have been reinterpreted to recover certain

[1] Chapter based on the paper "Market Crashes as Critical Phenomena: Explanation, Idealization, and Universality in Econophysics", co-authored with Jennifer Jhun and James O. Weatherall (Jhun, Palacios, and Weatherall in press)

regularities in economics. Work in this tradition has come to be known as econophysics, a term coined by H. Eugene Stanley in 1996. [2] In this tradition, important models have tried to account for cooperative behavior in economics using the physics of phase transitions. The idea of using the physics of phase transitions to build models in social sciences is principally motivated by the fact that phase transitions are the prototypical example of cooperative phenomena, in which the correlations between particles extend to very large distances, even though the microscopic interactions remain local. There is, therefore, the thought that the physics that successfully explains the first case will serve to explain the other analogous cases.

Despite the apparent empirical successes of some models in econophysics, the field has not been widely embraced by economists. The few who have engaged have been strongly critical. For example, Lo and Mueller. (2010) have argued that econophysics is doomed because "human behavior is not nearly as stable and predictable as physical phenomena" (1), and thus the strategies available in physics are not at all suitable for dealing with economic phenomena. [3] Our strategy will not be to address the general criticisms and we do not mean to argue that all models from econophysics, or even most or many models, are successful. Instead, we will focus on just one model that, we will argue, has two features of interest: it (1) draws on a significant analogy with phase transitions, in a way that goes beyond standard modeling methods in economics; and (2) has real explanatory power. Our principal goal is to elaborate and defend how we take the model to work, including where and how the analogy with phase transitions enters, and to articulate what sorts of novel insights into market behavior we believe it offers. In this sense, we take the model we consider as "proof of concept", while simultaneously providing

[2] For more on the relationship between physics, finance, and econophysics, see Weatherall (2013); for further technical details and overviews of recent work, see Mantegna and Eugene (1999), McCauley (2004), and Cottrell, P., G., Wright, and V. (2009).

[3] Despite the prevalence of this sort of criticism, it is far from clear that physics is more guilty of oversimplification than economics when it is applied to economic facts.

a case-study for the sorts of explanatory goals that arise in econophysics.

The model that we evaluate is the Johansen-Ledoit-Sornette (JLS) model of "critical" market crashes (Johansen, Ledoit, and Sornette 2000), which uses methods from the theory of critical phase transitions in physics to provide a predictive framework for financial market crashes.[4]. This model is of particular interest because it aims both to predict and describe market-level phenomena – crashes – and to provide microscopic foundations that explain how that behavior can result from interactions between individual agents. More specifically, in addition to its predictive role, the JLS model aims to explain two "stylized facts" associated market crashes. [5]. The first is the fact that stock market returns seem to exhibit power law behavior in the vicinity of a crash, and the second is so-called *volatility clustering*, which is the fact that market returns seem to exhibit dramatic, oscillating behavior before crashes, with large changes followed by other large changes.[6]

The plan of the paper is as follows. In section 4.2, we will present some (limited) background on mainstream modeling in financial economics that will help place the JLS model in context.[7]. In section 4.3, we will introduce the model itself, focusing on the role the analogy with critical phase transitions plays in the model. Then, in section 4.4 we will argue against one tempting way of understanding how the model works, and instead defend a somewhat different understanding. On the view we will defend, the principal achievements of the model are to explain why crashes occur endogenously in markets and to provide a possibly predictive signature for impending crashes.

[4]For more on this model and related ideas, see especially Sornette, Woodard, Yan, and Wei-Xing 2013 and references therein.

[5]These stylized facts are often treated as qualitative laws or as descriptions of lawlike behavior, capturing "set[s] of properties, common across many instruments, markets, and time periods" (Cont 2001, 223)

[6]This has also been noted by Mandelbrot (1963).

[7]For more on how the JLS model fits into mainstream financial modeling, see Sornette (2003); for background on mathematical methods in finance more generally, see for instance, Joshi (2008)

Central to our argument in section 4.4 will be the observation that although the analogy with critical phase transitions is crucial in motivating and developing the model, in the end the analogy is only partial. In particular, although the model fruitfully draws on the renormalization group theory of critical exponents, financial crashes do not seem to constitute a universality class in the strict sense that one encounters in that area of physics. Nonetheless, we argue, there is a weaker sense in which crashes exhibit universal features. This weaker notion of universality allows one to draw novel inferences about the microscopic mechanisms that might underlie crashes. Since the model helps make salient the possible microscopic mechanisms that could explain the occurrence of a crash, we claim that the model provides an explanation of crashes that is both causal (in the sense of Woodward 2003) and reductive.

In section 5 of the paper, we will explore how the argument just sketched relates to recent debates in philosophy of science concerning explanatory uses of idealized models. We will argue that the JLS model is naturally understood as a "minimal model" in the sense of Batterman and Rice 2014 (see also Batterman 2002; 2005; 2009). Nonetheless, we claim, (apparently) contra Batterman and Rice, that it provides both a causal and reductive explanation of market crashes. As we will argue, this shows that the same mathematical methods may be used for multiple explanatory purposes, and that to understand explanatory strategies in the context even of minimal models, one needs to pay careful attention to the salient why questions.

We conclude with some remarks about possible policy consequences. In particular, we argue that our interpretation of the JLS model as one that yields causal explanations suggests methods by which policymakers could intervene on the economy in order to prevent crashes or to halt the spread of one. The JLS model, we argue, may be used as a diagnostic tool, allowing economists and regulators to formulate new measures or to assess the performance of ones are already in place.

4.2 Some Financial and Economic Background to the JLS Model

Although the JLS model draws extensively on methods and ideas from the theory of critical phenomena in physics, it also builds on a long, mainstream tradition of market modeling in financial economics. Moreover, Sornette and collaborators emphasize this continuity with early work in financial modeling. In the course of analyzing work in econophysics, it seems particularly important to be clear about just where this work diverges from more traditional modeling. And so in this section we will provide some minimal background on methods and ideas from financial modeling that the JLS model builds on.

The JLS model may be broadly located in a tradition of modeling markets as stochastic processes. This tradition originated with groundbreaking work in 1900 by French mathematician Louis Bachelier, who first proposed treating price changes as a random walk and built an options model on this basis (Bachelier and Samuelson 2011). Bachelier's work went largely unnoticed, however, until re-discovered by J. L. Savage and Paul Samuelson in the early 1950s. Independently, in 1959 a physicist named M.F.M. Osborne proposed modeling market returns as undergoing Brownian motion (Osborne 1959). Osborne provided his own empirical support for this model, though it was largely consistent with earlier empirical work on market time series by the Cowles Commission (1933) and by Kendall (1953).

Later, Samuelson (1965) and Fama (1965) explicitly connected the random-walk hypothesis to the *efficient markets hypothesis (EMH).* [8] The EMH is

[8] The EMH has been a topic of considerable controversy. For instance, Shiller (1984) has argued that the argument behind the EMH is invalid. The main worry is that current models neglect (i) agent psychology and (ii) interactions amongst agents as key causal and explanatory features of asset price variations. Once these factors are considered, it seems markets may well be random irrespective of how efficiently markets process information or how accurately prices reflect fundamental values. Meanwhile, as Ball (2009) and others have argued, over-reliance on the assumption of efficiency may affect how market partic-

the claim that markets are informationally efficient and asset prices reflect (all) available information. The EMH is consistent with, and indeed implies, market randomness. This is because if markets are assumed to assimilate information efficiently, then any information available to market participants at a given time will already be factored into the price at that time.[9] Thus only (unaccounted for) news, which is random, changes prices, meaning that changes in stock option prices themselves must be random. Persistent exceptions to this rule, it is argued, are impossible, since if traders were to observe a pattern in asset price time series that could be exploited, they will exploit it, which would tend to wash out the pattern.

More formally, in efficient markets prices follow a martingale process, which is a general stochastic process where the conditional expectation of the next value, given past history and current value, is precisely the current value. That is,

$$E(p_{t+1} - p_t | \Omega_t) = 0,$$

where $\Omega_t = (p_1, p_2, ...p_t)$, the history up till time t.

Here $E(p_{t+1} - p_t | \Omega_t) = 0$, is the expectation value of the change in price in a given time-step. Thus, for an asset that pays no dividends, one should expect the future price to hover around the current value, all other things being equal.

$$E_t[p(t')] = p(t),$$

ipants synthesize information regarding possible asset bubbles. But we will not weigh in on such controversies; our purpose here is not to endorse the EMH, but rather to describe the context of the JLS model and to emphasize its continuity with mainstream economic modeling methods.

[9]Note that this argument appears to suppose that news that will positively affect price is equally likely as news that will negatively affect price. But if there were any information available that would indicate that positive (resp. negative) news was more likely, then that fact alone would count as tradeable information that would affect price.

for all $t' > t$. In other words, we could say that the prices of stocks do not depart from their fundamental or intrinsic value in a way that an investor could systematically predict or exploit to make a profit in the long run. In this sense, the EMH implies that the market will behave unpredictably.

The market models just described have some well-known limitations. For instance, if returns are modeled as a random walk, as Osborne and others proposed, one would generally expect returns to be normally distributed. In fact, however, market returns tend to be "fat-tailed".[10] This means that we see extreme events more often than one would expect if returns were normally distributed. In addition, treating markets as a martingale process leaves out a number features that appear to be good indicators of crises, such as volatility clustering (where large changes in price are followed by further large changes in price). That said, neither the martingale condition nor the EMH is in and of itself inconsistent with fat-tailed distributions or with large asset price changes. Indeed, there is a tradition in economics of modeling *rational bubbles*, which are deviations from fundamental values that are compatible with the martingale condition and the EMH (Blanchard 1979; Santos and Woodford 1997; Sornette and Malevergne 2001) The idea is that under some circumstances markets enter a "speculative regime" in which it is rational to hold onto an asset in anticipation of growing future returns, even though one believes that the current price is not the fundamental price. Here, markets may still be understood to be processing information efficiently – and thus the EMH may be taken to hold – since the endogenous facts about the speculative regime are themselves information bearing on future prices. In this regime, an asset's value grows indefinitely, which itself is not realistic but may be a suitable modeling assumption if persistant increase in value is anticipated over the timescale of interest. Still, rational bubbles models of the sort just described provide no insight into the circumstances under which the speculative regime ends and markets crash. The JLS model is

[10] See, for instance, Mandlebrot (1963) and Cont (2001).

intended to extend rational bubbles models in order to explain and predict market crashes in the speculative regime. The basic proposal is that financial bubbles and subsequent crashes are much like the development of sudden, spontaneous, and drastic behavior in physical systems such as magnets. Like earlier rational bubbles models, the JLS model treats bubbles and crashes without rejecting the EMH. Instead, as we will see in the next section, it attempts to reconcile the EMH with a story about the behavior of interacting traders.

4.3 The JLS model

Important stock market crashes of the twentieth century, including the US crashes of 1929 and 1987 and the Hong-Kong crash of 1997, have been the result of the action of a large group of traders placing sell orders simultaneously. Curiously, this synchronized "herding" behavior seems to arise endogenously, rather than from outside instruction or the influence of communication media. Traders, who are geographically apart and generally disagree with each other, seem to organize themselves to place the same order at the same time. The JLS model concerns the character and dynamics of this self organization between traders. [11]

In physics, critical phase transitions constitute an important class of phenomena that likewise exhibit "self organization". A paradigm example of these kinds of transitions is the paramagnetic-ferromagnetic transition in magnetic materials. In this transition, a large group of spins that are generally pointing in different directions align themselves in the same direction simultaneously, so that the system undergoes spontaneous magnetization. This suggests a potentially useful analogy between critical phase transitions

[11] Note that we mean "self-organization" in the informal sense of coordinated action between agents without any apparent external mechanism. We do not intend to invoke any specific theories of self-organization or self-organized criticality.

and stock market crashes.

Motivated by this analogy, Johansen et al (2000) propose a model (henceforth the JLS model) that elaborates on the rational bubbles models noted in the previous section and other work in econophysics (eg. Sornette, Johansen, and Bouchaud 1996). The main hypothesis underlying this model is that market crashes may be understood as a "critical phenomenon" strongly analogous to critical phase transitions. This hypothesis is made precise by postulating a correspondence between the quantities that are used to describe financial crashes and the physical quantities that describe critical phase transitions. This correspondence then allows one to draw inferences concerning various quantities of interest, including the probability of a crash occurring under various circumstances.

In more detail, on the JLS model a stock market crash occurs when the system transitions between two phases: a phase prior to the crash and a phase after the crash. This transition point is analogous to the critical point for physical systems, and in the present context corresponds to the time at which a stock market crash is most likely to occur. In this model, there are two quantities that are relevant for capturing this behavior of interest. The first is known as the *hazard rate*, $h(t)$. The hazard rate measures the instantaneous rate of change of the probability of the event occurring at time t, given that it has not yet occurred by t. The larger the hazard rate, the more rapidly the probability of an impending crash is increasing, given that the crash has not yet occurred. [12] It may be thought of as the instantaneous rate at which crashes should be expected to occur, if only crashes were repeatable. The second quantity is the price of some asset as a function of time, p(t). These two quantities determine the dynamic equation that will be used to predict

[12]More precisely, if F(t) is the cumulative distribution function of a crash occurring at or before time t, then $h(t) = F'(t)/(1 - F(t))$, where $F'(t) = dF$ is the probability density function. Conversely, one can define a cumulative probability function from a hazard rate by integrating both sides of this equation with respect to t. See, for instance, Cleves (2004, Ch. 2) for further details on interpreting hazard rates.

future crashes and provide a framework for the underlying microfoundational story.

The model begins with a general form for the price dynamics for a time prior to a crash. These dynamics are given by:

$$\log \frac{p(t)}{p(t_0)} = k \int_{t_0}^{t} h(t)dt \tag{4.1}$$

where is the price at some initial time t_0, is the price at a subsequent time t, is a constant, and is the hazard rate. Note that the hazard rate determines the price. [13] This means: the higher the hazard rate, the faster the price of an asset will rise. In other words, the more risky the asset is, the more the trader expects to receive in the future as compensation for taking on that risk.

Note that these dynamics are consistent with the standard financial modeling assumptions described above. In particular, in the special case where the hazard rate vanishes, the expected change in price over any given time interval vanishes, just as one would expect from the martingale condition discussed in Section 2 for a stock that does not pay dividends. Following JLS, we call this the "fundamental regime". When the hazard rate is positive, meanwhile – the so-called "bubble regime" – one expects price to increase exponentially over time. In this regime, the increase in price is driven up by the accumulated risk involved in holding the asset during a period in which a crash is deemed possible. Investors are willing to pay ever higher prices on the grounds that they expect price to continue to increase without bound, as long as a crash does not occur.

[13] It is tempting to interpret the right hand side of Eq. 4.1 as representing the probability of a crash occurring during the period from t_0 to t, but this would be incorrect: the integral of $h(t)dt$ does not yield a probability. (For instance, it may exceed 1.) Instead, this quantity should be understood as a measure of accumulated risk, in the sense that it represents the total number of times you should have experienced a crash during this period, supposing the crash were repeatable. Once again, see Cardy 2004[Ch. 2].

In this general form, these dynamics do not give an account of stylized facts such as the power law behavior we observe in financial time series, nor do they tell us anything about the microscopic mechanism underlying the occurrence of a crash. It is to get these further results that one introduces the qualitative analogy to critical phase transitions. (Up to this point, no such analogy has been invoked.)

To begin, we suppose that markets consist of populations of two types of traders, which JLS call "rational" and "noise" traders. (It is not essential that these populations be distinct; particular traders may sometimes be noise traders and sometimes rational traders.) The rational traders are assumed to trade on the basis of market fundamentals; noise traders, meanwhile, are assumed to base their decisions on trends, imitate others around them, etc. rather than investigating market fundamentals (Kyle (1985)).

The model then assumes that traders are situated in a lattice network, analogous to the lattice of the Ising model, the most important model in the study of phase transitions, including the paramagnetic- ferromagnetic transition mentioned above. (Note, however, that the specific lattice structure will turn out to be distinct from the Ising model.) Agents in this network may be in one of two possible states: a "buy" state or a "sell" state, just as spins in an Ising model may be either "up" or "down" Also like in an Ising model, agents are assumed to imitate their nearest neighbors, so that if a given agent is in a different state from the average of her neighbors, there will be a non-zero probability that the agent will change states. A crash on this model is understood as a moment in which a large group of traders are suddenly in the "sell" state.[14] Therefore, in this model a crash is caused (at the microscopic level) by self-reinforcing imitative behavior between traders.

[14]Sornette (2003) also considers the possibilities of "anti-crashes", wherein a large number of traders suddenly transition to "buy" states; these are taken to be the ends of "anti-bubble" regimes. However, it is important to note that neither Sornette (2003) nor Johansen (2000) explain the fact that crashes are generally caused by "sell" states instead of "buy" states.

This behavior is analogous to a phase transition, during which a large number of nodes in the Ising model adopt the same state.

In statistical mechanics, the quantity that best describes the tendency of particles to imitate one another is the *susceptibility* of the system. In the ferromagnetic-paramagnetic transition mentioned above, this quantity corresponds to the magnetic susceptibility , which is governed by the following power law near the transition point:

$$\chi \approx A|T - T_C|^{-\gamma} \tag{4.2}$$

where A is a positive constant, T_C corresponds to the critical temperature, and is known as the critical exponent. Informally, the susceptibility of the system characterizes the tendency of the system?s average magnetization (which is related with the number of spins in the same state) to change due to the influence of a small external field. One consequence of the power law is that at the critical point, $T = T_C$, χ diverges. The divergence of the magnetic susceptibility implies the divergence of the correlation length, a quantity that measures the average distance over which particles in the system interact. It is due to the divergence of the correlation length at the critical point that distant particles are likely to be mostly in the same state at the same time.[15]

The JLS model posits that the hazard rate has the same general form as the magnetic susceptibility

$$h(t) \approx B|t - t_c|^{-\alpha} \tag{4.3}$$

where t_c is the most probable time for the crash, B is a positive constant, and is a critical exponent that is assumed to have values between zero and one. Note that attributing this form to the hazard rate is really an ansatz:

[15]For more details on the logic of critical phenomena in physics, see Wilson and Kogut (1974), Goldenfeld (1992), Cardy (1996), Fisher (1998), Kadanoff (2000), Sornette (2006), and Zinn-Justin (2007); for a more philosophical take, see Batterman (2002) and Butterfield and Bouatta (2015).

no claim has been made to have derived this power law behavior from any microscopic model (or family of models). Instead, we have made two independent assumptions: the first is that traders may be modeled as agents on a lattice with two states, without specifying any details of the lattice or interactions between agents; and the second is that the hazard rate has a particular form analogous to the magnetic susceptibility. The idea that the hazard rate should be analogous to susceptibility is motivated by the idea that a crash should correspond to large correlation lengths, but this does not fix the form of the equation 4.3.

The final ingredient of the model is phenomenological. Observing the stylized fact that prices exhibit accelerating oscillations in the lead up to a crash, one infers that the critical exponent α in 4.3 is complex.[16] A complex critical exponent modifies the power law to include periodic oscillations in time known as log-periodic oscillations.[17] JLS argue that, to leading order in a Fourier expansion near t_c, the general solution for $h(t)$ is given by:

[16]The argument here is subtle. JLS first present their model generically, without making any assumptions about the details of the network. They then observe that if the network has certain features – in particular, if it is hierarchical in a sense to be explained in section 4.2 – then it will exhibit complex critical exponents, and hence log-periodic oscillations near criticality. They give some plausibility argument for considering hierarchical lattices, but leave the actual lattice structure open until they consider historical data –at which point they conclude that, given the presence of oscillations, the network must be approximately hierarchical and the critical exponents must be complex. It is in this sense that introducing complex critical exponents is "phenomenological". One can also run the argument in the other direction, however, and argue that on the basis of a plausible assumption concerning the hierarchical nature of trader networks, the critical exponents should be expected to be complex; at times, Sornette and collaborators appear to prefer this version of the argument.

[17]An early discussion of log-periodicity and self-similarity is given by Barenblatt and Zeldovich (1971). Extensive work on the existence of complex critical exponents with log-periodic oscillations has been carried out by Sornette and his collaborators (eg. Sornette 1998; 2006; Zhou et al. 2005).

$$h(t) \approx B_0|t_c - t|^{\alpha'} + B_1|t_c - t|^{-\alpha'} cos[\alpha'' \log|t_c - t| + \phi \qquad (4.4)$$

where B_0, B_1, and ϕ are real constants, α' is the real part of α, and α'' is the imaginary part of α.

Having identified this form for the hazard rate, one then plugs h(t) from Eq. 4 back into the general dynamic equation 4.1 to obtain an expression that describes the behavior of price as a function of time given this hazard rate, to obtain:

$$\log[p(t)] = \log[p_c] - \frac{k}{\beta}(B_0(t_c - t)^{\beta} cos[\omega \log(t_c - t) + \phi]) \qquad (4.5)$$

where $\beta = 1 - \alpha' \in (0,1), p_c = p(t_c)$ is the price at the critical time, and ϕ is another constant.

Eq. 4.5 succeeds in capturing the stylized facts observed in the occurrence of extreme events, including volatility clustering and accelerating oscillations (Yalamova and McKelvey 2011). Moreover, as we will elaborate below, it provides an explanation of these observed phenomena – and indeed, of crashes themselves – that appeals to the existence of self-reinforcing imitative behavior between traders. Finally, the model aims to be predictive by providing the tools to anticipate the occurrence of crashes that arise due to endogenous herding behavior, such as panics, by describing a specific form of accelerating oscillations – namely log periodic oscillations – that provide a signature of approaching criticality.

Note that although volatility clustering and accelerating oscillations are taken as stylized facts that are "inputs" for the model that are used to establish that the complex exponent in Eq. 4.3 is complex, the specific form of Eq. 4.5 should be taken as an output of the model. As such, it can be backtested to provide empirical support for the model as a whole, and specifically for the claim that crashes may be understood as critical phenomena. The results of these tests have been reported in several places (Sornette, Johansen, and Bouchaud 1996; Sornette and Johansen 1997; Johansen, Ledoit, and Sornette 2000; Sornette 2003; v. Bothmer and Meister. 2003; Calvet and Fisher. 2008). Perhaps most remarkable is the crash of 1987, where the log-periodic oscillations are visible even to the naked eye (Johansen, Ledoit, and Sornette 2000).

4.4 The Logic of the JLS model

The JLS model, and the analogy between crashes and critical phenomena on which it is based, are highly suggestive. However, one needs to be careful about the limits of the analogy.[18] As we will presently argue, even if one accepts the arguments given in the previous section, the logic of the model is importantly different from that of models from statistical physics on which it is based. First, we will argue that unlike critical phase transitions, "critical" market crashes do not form a universality class in the sense of renormalization group (RG) physics. It follows that explanatory strategies familiar from applications of the RG in physics do not carry over directly to this model. We will then present a different analysis of the logic of the JLS model, emphasizing what sort of explanations we think the model can provide. We will conclude by observing that although the mathematical methods used in the JLS model are similar to those from physics the role that these methods play in application are different.

[18]There are various criticisms of the JLS model that also stress the disanalogies between the JLS model of financial crises and critical phase transitions. For example, Ilinski (1999) casts doubt on a main component of the JLS model: crashes are principally caused by imitative dynamics between individual traders. He objects that different market participants may act over different time horizons (e.g. minutes for speculators, years for managers), so that the instantaneous long-range interactions between traders postulated by the JLS model are implausible. We will not engage with this criticism or others; instead, we want to see how far the analogy goes if we assume that the model is well-motivated and well-supported empirically.

4.4.1 Do Market Crashes Constitute a Universality Class?

To evaluate the analogy between market crashes in the JLS model and critical phenomena in physics, we will begin by describing the situation in physics in some further detail. As noted above, when a system undergoes a critical phase transition, some important physical quantities diverge. For instance, in the ferromagnetic-paramagnetic transition described in section 3, the divergent quantities are the magnetic susceptibility, the specific heat, and the correlation length. The divergence of the correlation length implies that all spins are correlated at the transition point regardless of the distance between them. That is, the measuring distance unit is no longer important. When this happens, the system is said to be scale invariant.

Scale invariance is consistent with the observation of power law behavior of physical quantities near a critical point. The exponents appearing in these power laws – called *critical exponents* – were originally determined experimentally. Surprisingly, radically different systems, such as fluids and ferromagnets, were found to have exactly the same values for their critical exponents. This was particularly striking because the exponents were deemed *anomalous*, which is to say that they were not whole numbers or simple fractions. Systems having the same values of their critical exponents are said to belong to the same *universality class*.[19] One of the great achievements in the theory of phase transitions was the development of RG methods to explain how this universal behavior comes about – i.e., to explain why apparently different systems have the same scaling behavior near criticality.

[19]As will become clear in what follows, by "universality class" we mean the basin of attraction of a given non-trivial fixed point under some RG flow. In cases of critical phase transitions, these correspond to systems with the same critical exponent near the transition point, though RG methods may be applied more generally. Batterman and Rice (2014) suggest a still-broader definition of "universality class" that applies to systems outside of physics where the RG does not apply; as we will see below, market crashes will turn out to form a universality class in this more general sense, but one needs to be careful about the role that the RG plays in the argument for this.

RG methods consist, roughly, in a set of transformations by which one replaces a set of variables by another set of – generally coarse-grained – variables without changing the essential physical properties of a system. The (infinite) iteration of these transformations in a space of Hamiltonians enables one to find so-called fixed points of the transformation, which are Hamiltonians that represent the (coarse-grained) dynamics of a system near a transition point.[20] This procedure is taken to explain universality, as it has been shown that systems in the same universality class flow to the same fixed points, and thus the systems in a given universality class should be expected to have the same dynamical properties near the transition point. The existence of non-trivial fixed points is generally taken to show that a system's microscopic details are irrelevant to its behavior near criticality. In addition, RG methods provide an argument for the use of highly idealized models in the explanation of radically different systems. For instance, by showing that both ferromagnets and fluids are in the same universality class as the Ising model, RG methods justify the use of the Ising model for the study of both systems.

Thus, in physics, the logic of universality arguments goes as follows. One begins with the empirical observation that certain systems exhibit the same behavior ? i.e., have the same critical exponents – near criticality. One then shows that those that systems flow to the same fixed point by iterated application of an RG transformation, thus explaining their observed similarity by establishing that, at a certain level of coarse-graining, these systems have the same dynamical properties. In other words, the thing one is ultimately trying to explain is why a range of apparently different systems are all saliently the same, and the explanation proceeds by showing that the microscopic details of the systems do not matter to the phenomenon in question.[21]

[20]Note that our description of RG methods here follows the "field space" approach, in the sense of Franklin (2017).

[21]Note that it is not essential, here, to begin with an empirical observation –though that is what happened in the physics of phase transitions. In principle, one can demonstrate

Is the same reasoning applied in the JLS model? Note that if this sort of argument could be so applied, it would be very attractive. For one, it would mean that markets could be expected to always have the same log-periodic behavior, with the same critical exponent, in the lead up to a crash, which would yield a strong predictor of crashes. Moreover, there are good reasons to be skeptical about the microscopic details of any market model. Markets have heterogeneous participants whose behavior will depend on a large number of endogenous and exogenous factors that no model can hope to accurately represent. And so, an argument to the effect that, irrespective of the details of how one models market participants' disposition, the same large-scale behavior can be expected would provide both helpful support for the resulting large-scale model, and also alleviate worries about the particular microscopic model that has been adopted. It would also justify adopting highly idealized models of market actors, analogously to how the renormalization group justifies using the Ising model for critical phase transitions.

Unfortunately, however, it would appear that this reasoning cannot be carried over directly. In particular, the first step does not work. While data-fitting supports the idea that the relationship between price returns and hazard can be captured via a power-law (eg. Johansen, Sornette, and Ledoit 1999), analysis of past crashes does not support the hypothesis that crashes constitute a universality class in the sense of all corresponding to the same non-trivial fixed point of some RG flow. This is because crashes do not all exhibit the same critical exponent. Via curve-fitting, Graf v. Bothmer and Meister (2003) show that in 88 years of Dow-Jones-Data there actually are no characteristic peaks in the critical exponent β of equation 4.5. Although JLS showed that the exponent of the crash in 1987 and the crash in 1997 differ by less 5%, Sornette et al. (1996) show that the value of that exponent differs substantially from other important crashes such as the crash in 1929.

that two systems are in the same universality class and thereby predict their behavior near critical points.

The fact that that there is no characteristic peak in the exponent ? has the following consequence. Stock market crashes are neither in the same universality class as the Ising model (or any previously solved model) nor do they constitute a universality class themselves.

One might think, as Sornette and collaborators themselves seem argue to in at least some places, that the fact that crashes do not constitute a universality class entirely undermines the analogy between crashes and critical phenomena.

> If we believe that large crashes can be described as critical points and hence have the same background, then β, ω and δt should have values which are comparable. (Johansen, Ledoit, and Sornette 2000, p. 17)

As we will argue below, however, we do not think that the failure of crashes to constitute a universality class is a major problem for the model. But it does mean that the logic of the model, and the sorts of explanations we can expect from it, are importantly different from in physics. If we cannot expect crashes to constitute a universality class, then the RG story cannot be applied either for the calculation of critical exponents or for the explanation of the universal behavior observed in crashes (or not observed, as it happens). In other words, if there is universal behavior in stock market crashes, this is not the kind of universal behavior that can be explained via RG methods alone. [22]

[22] Note however that this does not mean that RG methods cannot be applied at all in the context of the JLS model. Zhou and Sornette (2003), for instance, use renormalization group methods to obtain an extension of equation (5) that gives an account of larger time scales. Moreover, as we will see, RG methods will reappear in our analysis below, although they will play a different role than in statistical physics.

4.4.2 On the Reductive and Explanatory Character of the JLS model

We saw above that the JLS model apparently does not work by establishing that market crashes form a universality class. This means that one cannot apply the same reasoning as in physics to argue that large-scale market behavior near transition points (i.e., crashes) is independent of the microscopic details of market dynamics. It thus seems that insofar as the JLS model is successful, it must function differently. In this section we will develop a positive account of the logic of the JLS model, describing what we take the model to explain and how. We will argue that the JLS model relies on a subtle interplay between microscopic and macroscopic considerations, by which known mathematical facts familiar from statistical physics are used, in conjunction with empirical considerations, to draw inferences in both directions.

Recall that, whereas the arguments from statistical physics sketched above began with a brute empirical claim –many systems appear to have the same critical exponents – the JLS model began with two separate ingredients. The first, 4.1, was taken from mainstream economics – or at least, from the theory of rational bubbles. The second, Eq. 4.3, was a bare *ansatz*, inspired by statistical physics but in no sense justified by it. In other words, one begins by considering what market dynamics would look like if the hazard rate were governed by a power law near crashes, similarly to how the magnetic susceptibility behaves. These two ingredients, along with the further specification that the exponent in Eq. 4.3 is complex, then lead to Eq. 4.5, concerning the logarithm of market prices near a crash. It is this equation that is the principal predictive output of the model, and also the means by which the model is both calibrated and tested against historical data.

But this is not the whole model. To motivate the *ansatz* that the hazard rate satisfies Eq. 4.3 near crashes, JLS include a third ingredient, which is that microscopic market dynamics may be modeled as a network of agents, interacting with their nearest neighbors via imitation, and that the hazard rate may be interpreted, much like the magnetic susceptibility, as a measure of the characteristic distance scale of correlations between agents. The proposal that market participants form some sort of network of influence is taken as prima facie plausible, and no particular evidence is offered for it; at this stage, no claims are made about the details of the network structure. Drawing on known results from statistical physics, JLS then observe that networks of this sort are very often associated with power laws near criticality for the parameter that is now being interpreted as hazard rate, thereby linking Eq. 4.3 with a class of microscopic models.

One then argues that insofar as Eq. 4.5 is successful, this relationship between network models and power laws lends further plausibility to treating market microdynamics with a network model of this sort, and also that spontaneous herding, which now is understood to correspond to long-distance correlations in a network, explains endogenous market crashes. In particular, the divergence of the hazard rate at the critical point implies the divergence of the correlation length, i.e. the range of interaction between traders.

As we noted above, if the correlation length in a network model of this sort diverges, the system becomes scale invariant. It is under these circumstances that the system is successfully described by power laws. Scale invariance means that, near the critical point, market dynamics are self-similar across scales. In other words, as traders imitate their neighbors, they aggregate into clusters (e.g. companies) that act as individual traders imitating their neighbor companies, and so on, to higher and higher scales. This imitation procedure across different scales accounts for how information propagates so quickly before a crash: "...critical self-similarity is why local imitation cascades through the scales into global coordination" (p. 32).

But now, recall that the critical exponents in the JLS model were determined to be complex, and that the associated power laws exhibited log-periodic oscillations. Not all network models lead to log-periodic power laws

(LPPLs); they typically arise (only) when the underlying network model exhibits discrete scale invariance. Discrete scale invariance means that the system is scale invariant only under special discrete magnification factors; this, in turn, implies that the system and the underlying physical mechanisms have characteristic length scales. As Sornette (1998) points out, this provides important constraints on the underlying dynamics. In particular, it suggests that traders are arranged on a *hierarchical* lattice, which is a lattice in which, by virtue of the network structure, some nodes (traders) have greater influence than others (still via nearest-neighbor interaction). Examples of hierarchical networks such as the Bethe lattice, a fractal tree, or hierarchical diamond lattice. These hierarchical networks tell us not only how information propagates through scales but also how information propagates within the same scale. In figure 1, for instance, one can see that in the Bethe lattice information that starts by one agent propagates within the same scale faster than exponentially.

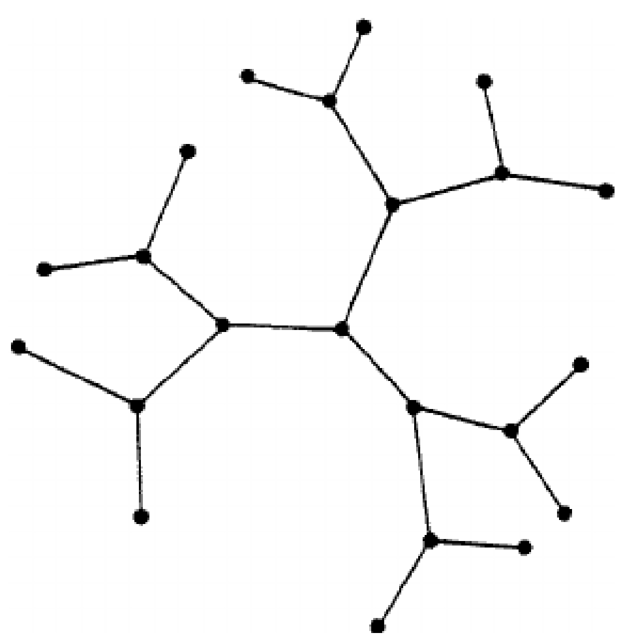

Figure 1: Illustration of a Bethe Lattice, one of the possible network structures underlying the occurrence of a crash according to the JLS model. The point in the center of the figure represents a trader who is source of opinion. The first ring represents the neighbors, who tend to imitate the opinion of the trader at the center. The second ring represents their neighbors, who are indirectly influenced by the opinion of the first trader, and so on. This aims to illustrate how imitation could possibly propagate resulting into global coordination.

Sornette has argued that it is plausible to model the propagation of information in social structures using hierarchical lattices, and also that there is independent empirical support for doing so (Sornette 2003, Ch. 4). But it is not claimed – nor is it necessary to claim – that under general circumstances, the network of traders is fractal, or that it corresponds to some exact hierarchical lattice. Instead, what is claimed is that under general circumstances, the network of traders lies in the basin of attraction of a hierarchical model under RG transformations, so that its critical behavior is the same as that of a hierarchical network, i.e., so that near a crash markets exhibit LPPLs. In other words, interactions between traders must be "approximately" hierarchical, in the sense of lying in the same universality class as some hierarchical network (with imitative dynamics). It is here that RG methods enter explicitly into the JLS model. One might think of the role played by RG methods here as establishing that crashes form a universality class in a more general sense than that discussed above, namely by showing how a wide range of systems flow to fixed points characterized by hierarchical networks of one sort or another. (We will return to this idea below.)

We claim that it is the inference from observed LPPLs to discrete scale invariance of an underlying network structure (or, more generally, from power laws of any kind to scale invariance) that forms the explanatory core of the JLS model. In more detail, what we find here is an explanation of (endogenous) market crashes as arising from the structure of the network of traders at the time the crash occurs. Markets crash in the absence of any external, coordinating event because the network of traders can spontaneously evolve into states that are (discretely) scale invariant, i.e., which have long correlation lengths, so that small, essentially arbitrary perturbations, can propagate rapidly across scales. Perhaps surprisingly given the literature on universality and explanation, this explanation, as we understand it, is causal, in the sense of Woodward's interventionist account of causation (Woodward 2003).[23] On Woodward?s account, causes are variables that one could intervene on in order to reliably influence a system. More precisely, one says that A causes B if (given some background conditions) there is a conditional of the form "if A, then (likely) B", where A can be understood as a single variable that one could, in principle, manipulate. On this account of causation, a relationship such as the one between LPPLs, discrete scale invariance, and transitions, which holds across a range of different condition, can serve as a guide to identifying causal relations. As Woodward puts it, "When a relationship is invariant under at least some interventions...it is potentially usable in the sense that...if an intervention on X were to occur, this would be a way of manipulating or controlling the value of Y" (16). We take it that the moral of the JLS model in its most general form is as follows: if the network of agents participating in a market approaches a (discretely) scale-invariant

[23] Sornette also speaks of this explanation as "causal": for instance, when he writes "?the market anticipates the crash in a subtle self-organized and cooperative fashion, hence releasing precursory "fingerprints" observable in the stock market prices?. we propose that the underlying cause of the crash must be searched years before it in the progressive accelerating ascent of the market price, reflecting an increasing build-up of the market cooperativity" (Sornette 2003 p. 279). As we noted in footnote 8, we do not take the claim that this explanation is causal to be in conflict with the views defended by Batterman (2000; 2002), Reutlinger (2014), or others. The claim is not that there is an explanation of universality in this model that is causal. Rather, the claim is that the explanation of a given crash, or even crashes in general, is causal, because the JLS model identifies how to intervene to produce a crash, or to prevent one – namely, by changing network structure.

state, as signaled by the appearance of LPPLs in price, *then* (it is likely that) a crash will occur. In other words, the model says that crashes occur in many different systems precisely when their (coarse-grained) dynamics become approximately discretely scale invariant. And so, it is the emergence of discrete scale invariance (or, perhaps, scale invariance more generally) that should be identified as the proper cause of the crash. On this view, it is the state of the network as a whole that should be understood as the cause of the crash. But one might worry that this is not an ?event? or ?variable? of the sort that one can intervene on. We believe it is. First, observe that on Woodward?s account, it need not be possible to *actually* manipulate the variable; it need only be the case that one could imagine, within the model, changing just this feature. And indeed, in the present case, one certainly can change the state of the network so that it is no longer scale invariant (discretely or otherwise), and in doing so, one *ipso facto* moves away from the transition point. This is precisely what is needed. More can be said on this point, however. As we will explore in the final section of the paper, we believe there are mechanisms by which an agent – say, a regulatory body – can in fact intervene on the network structure of market participants in order to disrupt scale invariance. If this is possible, then the conditional above not only bears a clear causal interpretation, but in fact has policy implications regarding how to deal with an impending market crash. Before turning to this point, however, we will consider how the analysis of both the logic of the JLS model and its explanatory properties that we have just provided bears on recent debates concerning explanation and universality in the philosophy of science literature. 5. Infinite idealizations, universality, and explanation in the JLS model In the last section, we argued that the JLS model, though bearing important relationships to models of phase transitions in physics, relied on an argument that was importantly different, both in the sense of "universality" at play and in how inferences are drawn about the micro- and macrodynamics of markets. We also presented a positive account of both the logic of the model and the character of the explanation it offers of market crashes. As we argued, this explanation is best construed as *causal*, in the interventionist sense of Woodward (2003).

We made these arguments largely independently of the recent literature on the character of explanations in statistical physics that make use of the methods the JLS model borrows. There was good reason for this: our main contention above was that the logic of the JLS model is different from that of the models of phase transitions on which it is based. That said, there are some features of the JLS model that make it salient from the perspective of recent debates on explanation in philosophy of science. In particular, the JLS model is arguably a *minimal model* in the sense of Batterman and Rice (2014).[24] A minimal model, according to Batterman and Rice, is one that "...is used to explain patterns of macroscopic behavior across systems that are heterogeneous at small scales" (p. 349). More importantly, minimal models are "thoroughgoing caricatures of real systems" whose explanatory power does not depend on their "representational accuracy" (p. 350). Instead, the key feature of a minimal model is that it allows us to say why many different systems turn out to be saliently similar, despite their significant differences at a microscopic level.

The model of critical phase transitions discussed above is a paradigm example of a minimal model in the Batterman and Rice sense. There, the

[24]See Lange (2015) for a different critique of Batterman and Rice (2014) than we give here. Lange argues that Batterman and Rice cannot sustain the distinction they draw between their account and "common feature" accounts such as Weisberg's (discussed below). We take it that one can sustain a distinction between different explanatory goals, one of which might well be to explain why many different systems should be expected to be saliently similar to some highly idealized model, and we think that Batterman and Rice do an adequate job of explaining both how that explanatory goal can be met, and why the strategies for meeting it do not look like they are appealing to common features of a model and a target system. That said, as we will argue, in some cases a single model, including the JLS model, can be used to achieve more than one explanatory goal.

goal is to explain why many different systems have the same behavior near transition points, and moreover, to show why highly idealized models, such as the Ising model, capture the essential behavior of all of these different systems. The RG played an essential role in this story. But Batterman and Rice are clear that it is not only models that use the RG in this way that are to count as minimal models: they also describe an example from biology – the Fisher sex ratio model - and argue that it is a minimal model as well. The essential feature in both cases is that one has a universality class, in the general sense of a collection of models that are all similar in some salient way, and an explanation of why all of the systems in question fall into that universality class.

We argued above that even though the RG plays a different role in the JLS model than in models of critical phase transitions, there is still a sense in which market crashes form a universality class, according to the JLS model. This universality class does not correspond to the basin of attraction of a single non-trivial fixed point under iterated applications of an RG transformation. Instead, it is a collection of systems that are all saliently similar, in the sense that they exhibit LPPLs.

Still, one can explain why a wide range of systems exhibit this same universal behavior: they all exhibit discrete scale invariance near their transition points. Moreover, RG methods play an important role in this argument. Although RG transformations do not take all of the relevant similar systems to the same non-trivial fixed point, they do take such systems to non-trivial fixed points with complex critical exponents, and thus LPPLs. So in this sense, the RG establishes the universality class in the salient (generalized) sense. Finally, although one cannot show that there is some idealized model that has the same critical exponent as every market crash –since not all market crashes have the same critical exponent! – one can show that there are highly idealized models, each exhibiting discrete scale invariance near transition points, that give rise to LPPLs near their transition points. It is on these grounds that we take the JLS model to be a minimal model in the Batterman-Rice sense.

The JLS model also has another feature that, though not part of the official definition of minimal models, seems characteristic of them (Batterman 2005; 2009): the JLS model relies on an *infinite idealization*. (This provides one sense in which the model "caricatures" real markets.) That is, the JLS model assumes that the network of market participants includes infinitely many agents. Moreover, this feature is necessary for the model as we have described it, and it is assumed in all versions of the model we know of in the literature. The reason it is necessary is that scale invariance, including discrete scale invariance, means that some property of the model must hold – i.e., be "invariant" – at all scales, no matter how large. Thus only an infinite model may be truly scale invariant. Likewise, only an infinite model can exhibit the sort of infinite correlation lengths that we identify with a transition point.[25]

These features of the JLS model, and especially the role that the infinite idealization plays in establishing scale invariance near the critical point, are common across applications of the RG methods. And Batterman puts considerable weight on the infinities that arise in models that use these methods: rather than anomalies to be avoided or removed, they are sources of important information.

> I'm suggesting that an important lesson from the renormalization group successes is that we rethink the use of models in physics.

[25]This is not to say that the model could not be reconfigured as one that is invariant across some scales, but not under arbitrary scale transformations. In other words, we do not mean to deny what is sometimes known as "Earman's principle", that idealized models can only be explanatory if one can imagine removing the idealization and still being able to explain the same phenomenon (Earman 2004; J. Butterfield 2011). But doing so would require substantial changes in the analysis, and would effectively produce a different model from the one under consideration. Our interest is in the explanatory role of the infinite idealization in the present version of the model.

> If we include mathematical features as essential parts of physical modeling then we will see that blowups or singularities are often sources of information. (Batterman 2009, p. 11)

It seems that something similar is going on in the JLS model: there, too, one encounters not only infinite systems, but also divergent quantities – including both the hazard rate and the correlation length between traders. And it is these blowups that signal that a crash is impending. This singular behavior is at the very core of the model.

So it seems that the JLS model has the hallmarks of a minimal model. But if so, there is a tension between what we say above and Batterman and Rice's account of how minimal models explain. In particular, Batterman and Rice emphasize that the sorts of explanations they consider are *non-causal* and *non-reductive*.[26] Moreover, they argue minimal models are not *representational*, in the sense that their success does not depend on "some kind of accurate mirroring, or mapping, or representation relation between model and target" (351). On our view, however, the JLS model does provide a causal explanation; moreover, this explanation is arguably both reductive and representational. We have already seen the sense in which the JLS model provides a causal explanation: it may be understood to yield a conditional statement, the antecedent of which is a variable on which one can, in principle, intervene. Thus, on an interventionist account of causal explanation, the model appears to allow us to say that it is (discrete) scale invariance that causes market crashes – or, to put it in more evocative terms, it is herding at all scales that causes market crashes. Some readers will balk at this claim: after all, as just noted, only infinite systems can be truly scale invariant, and realistic markets are not infinite. So, in what sense could a feature that no actual market could have cause a behavior that realistic markets exhibit? Or to put it another way, how could actual market crashes be caused by scale invariance? The answer, as we see it, is that the JLS model explains crashes

[26]See also Morrison (2006) for a related point.

by showing that in some networks, correlation lengths can become long, relative to the overall size of the network, and that when this happens, crashes become likely. It is the infinite idealization that allows one to precisely characterize the relationship between long correlation lengths, scale invariance, and crashes, and it is not clear that one could establish this relationship as neatly in a finite system as one can in the infinite system. But what the infinite system is ultimately telling us is something about the causal relationship between correlations between traders and market-wide crashes. [27] We should emphasize that, although we take this explanation to be causal, it is only on a particular account of causation (i.e., the Woodward (2003) account). Of course, there are many other analyses of causation on which this may well not be a causal explanation (Salmon 1984; Strevens 2008). More importantly, we do not claim that crashes are being explained, here, by appeal to particular details concerning interactions between individual agents. In this sense, it is not a "causal-mechanical" or "mechanistic" explanation (Craver 2006; Kaplan 2011; Kaplan and Craver 2011). Indeed, the model is not committed to any particular network model at the microscale, just a class of models that exhibit discrete scale invariance. Sornette puts the point as follows.

> It turns out that there is not a unique cause but several mechanisms may lead to DSI. Since DSI is a partial breaking of a continuous symmetry, this is hardly surprising as there are many ways to break down a symmetry. We describe the mechanisms that have been studied and are still under investigation. The list of mechanisms is by no mean exhaustive and other mechanisms may exist. (Sornette 1998, p. 247)

[27] Here there is a relationship both to "Earman's principle", as noted in footnote 35, and also to Butterfield (2011), who argues that in cases where one takes an unrealistic infinite limit, one should expect to see the qualitative behavior that arises in the limit appearing already on the way to the limit.

Thus, the model does not even include a specific account of how agents interact with one another. It is rather a generic feature of a range of possible networks that plays the causal role.

This last point is also closely related to the senses in which we take the JLS model to be reductive and representational. The antecedent of the conditional described above refers to the micro-constituents of the market. It is in this sense that we take the explanation to be reductive. But it does not follow that the model supposes an atomistic conception of the economy, i.e. it does not determine the law governing the behavior of any arbitrary agent. But, given some behavioral assumptions, it does constrain the kinds of structures they might reside in. In this case: hierarchical structures that (sometimes) exhibit discrete scale invariance. This does not require any particular arrangement of individuals because those particular details are in some sense irrelevant; what does matter are these structural details.

Likewise, the model is representational in the sense that its success depends on the fact that it represents certain stylized facts about market participants: they influence one another, at least sometimes, by imitation, and their interactions are hierarchical, in the sense that some traders are able to influence larger groups than other traders. Of course, this is far from a complete or accurate representation of market participants. But if actual market participants do not bear relations to one another that are adequately represented by a network with these features – or if markets are not discretely invariant across at least some scales – then the JLS model would fail to support the causal explanation we have described here. And so, it seems that the success of the explanation does depend on the representational accuracy of the model, at least with regard to these particular features. This weak sense of being "representational" indicates that the JLS model may (also) be understood as an example of what Weisberg (2007) calls "minimalist models": "[A] minimalist model contains only those factors that make a difference to the occurrence and essential character of the phenomenon in question" (Weisberg 2007, p. 642). It also invokes Strevens' (2008) account of idealized models: "the content of an idealized model, then, can be divided into two parts. The first part contains the difference- makers for the explanatory target... The second part is all idealization; its overt claims are false but its role is to point to parts of the actual world that do not make a difference to the explanatory target" (318). Strevens, too, argues that this sort of idealization is compatible with causal explanation. Of course, Batterman and Rice's *minimal models* and Weisberg's *minimalist* models are supposed to be fundamentally different; worse, those philosophers who have mistaken minimal models for minimalist models have "almost universally misunderstood" the explanatory structure of these models (Batterman and Rice 2014, 349). And yet, it would seem that the JLS model is an example of both. How could this be? The tension can be resolved if one distinguishes between, on the one hand, features of a model – what sorts of idealizations it involves; in what senses, if any, it is representational; what sorts of mathematical relationships and methods it relies on – from the sorts of explanations one can give by appealing to the model, i.e., the why questions one is able to answer (citealt-Vanfraassen1980). [28] Batterman and Rice define minimal models as models used to give certain sorts of explanations involving universality classes. Since the JLS model can be used to explain why market crashes form a universality class (in the broad sense), the JLS model counts as a minimal model. These explanations, they argue, are neither causal nor reductive, and their success does not depend on the accuracy with which the models represent target systems; using the JLS model to explain the universal behaviors associated with crashes (namely, LPPLs, discrete scale invariance, etc.) is presumably also non-causal, at least insofar as Batterman and Rice's arguments are con-

[28] This point mirrors one made by O'Connor and Weatherall (2016): there are many different purposes for which models may be constructed, and to which they may be put. This includes different explanatory purposes, and so one should be cautious about attempts to classify or taxonomize models on the basis of how they may be used to explain.

vincing. [29] But the fact that the JLS model can be used for this sort of explanation does not bear on whether one can also use it to provide other explanations; nor does it bear on which explanations seem most salient in the context in which the JLS model was developed. In other words, we claim that the JLS model may be used to answer the question, 'Why do markets generically exhibit volatility clustering, log-periodic oscillations, etc. near market crashes, even though market conditions otherwise vary dramatically?" To do so, one uses RG methods to show that a large variety of different networks exhibit discrete scale invariance and satisfy LPPLs near transitions points. In answering this question, we give the sort of explanation that Batterman and Rice are pointing to, and it is for this reason that the JLS model is a minimal model. But we claim that we can also use the JLS model to answer the question, "Why do stock markets crash?", where this question is understood to be about the causes of crashes. And in this case, the answer is: because hierarchical networks can spontaneously evolve into states featuring discrete scale invariance, and scale invariance of any sort allows vanishingly small perturbations to cascade across scales.[30] It is in answering this question that the Woodwardian conditional described above is crucially invoked. And it is in answering this question that the minimalist representational features of the JLS model matter. There are several points to emphasize here. The

[29] We tend to think that they are convincing, or at least, we agree that explanations of universality of the sort Batterman and Rice discuss are non-causal. (See also Reutlinger 2014 for a different argument concerning why these explanations are non-causal.)

[30] Note that there is another interpretation of "Why do stock markets crash?" that does not demand a causal explanation, but rather another minimal model explanation: namely, "Why do markets fall into a universality class of systems that exhibit crashes, as opposed to tamer sorts of transitions?" Of course, this is a legitimate explanatory demand, and the answer, invoking the JLS model, would look more like the answer to the first question than the second. The difference between these two understandings of the question "Why do stock market crash?" invokes van Fraassen's (1980) analysis of the logic of why questions. Explanatory demands, van Fraassen convincingly argues, involve, in addition to the explinandum, both a contrast class and a relevance relation.

first is just to clarify our argument, lest our claims above be misconstrued: As should now be clear, when we argued above that the JLS model provides a causal explanation, we did *not* mean to imply that the explanation one can give for why market crashes form a universality class is a causal explanation (contra Batterman and Rice), nor (*ipso facto*) that *all* explanations are causal.[31] The point is rather that the JLS model, despite having the characteristic features of a minimal model, may nonetheless be used to give causal explanations (in addition to minimal model explanations). And pulling apart these different explanatory tasks requires careful attention to precisely what question one is trying to answer. A second point to emphasize is that, even though the why questions described above are distinct, there is a subtle interplay between them. It is precisely *because* the JLS model can be used to explain why market crashes form a universality class in the relevant sense that it can (also) be used to provide a certain kind of causal explanation of market crashes, since it is the relationship picked out by this universality class, between discrete scale invariance and LPPLs near transition points, that makes true the conditional that forms the basis of the causal explanation. More, for precisely the same reason, the infinite idealization in the JLS model is essential precisely because it helps one identify the common mechanism underlying the phenomenon of interest ? and thus, it is the infinite idealization that permits the causal explanation. Conversely, it is precisely because the relationship encoded by the Woodwardian condition holds that market crashes fall in a universality class (in the broad sense) in the first place.

This situation raises a question. If the JLS model can be both a minimal model and also a minimalist model, can we understand the other models that Batterman and Rice discuss, including models of critical phase transitions, as also providing interventionist causal explanations (in addition to minimal

[31] For other examples of explanations that seem to be even more clearly non-causal, see Weatherall (2011; 2017).

model explanations)? In a sense, the answer must be "yes", at least if what we argue above is correct. For instance, in the phase transition case, one can use the Ising model to answer the question, "Why do critical phase transitions occur?", construed causally, by showing that the Ising model, and a wide range of other models in its universality class, can evolve into states that are (approximately) scale invariant, and thus vanishingly small perturbations can cascade across scales. This explanation is causal in just the same sense that the corresponding explanation invoking the JLS model is. Once again, there is a subtle interplay between this explanation and the minimal model explanation using the same model, since the fact that real systems are in the same universality class as the Ising model is precisely what isolates scale invariance as the difference-maker (or, perhaps better, the manipulable variable).

All that said, there is still a difference between the JLS model and critical phase transitions in this regard. It concerns which explanatory demands seem most salient. As we noted above, one of the most striking features of critical phase transitions is the fact that many different systems have the same critical exponents. The salient issue is not to explain why transitions occur at all, but rather to explain why transitions in different systems are so similar. Of course, this does not prohibit one from asking the other question; it is just a matter of emphasis. (Besides, background theory, such as mean field theory, seems to explain this well, without explaining universality.) In the case of financial markets, the situation seems to be reversed: there, one wants to explain why (endogenous) market crashes occur at all, particularly given that crashes are often taken to be in tension with the EMH and other standard market modeling assumptions. And for this reason, it is the causal explanation using the JLS model that seems to be the salient one.

4.5 Policy Implications

We argued above, particularly in section 4.2, that the sense in which we take the JLS model to provide a causal explanation is interventionist: it depends on identifying a potential conditional relationship, the antecedent of which can be understood as a variable that can be manipulated, at least in principle. Moreover, the JLS model provides an observable signal of when that antecedent obtains. But having identified such a variable means that we have also identified a potential target for policy intervention. If we accept the JLS model, how might a regulatory agency intervene to prevent crashes? The answer is to disrupt the network structure on which traders reside.

How might one do this? One possibility would be through structural changes. Hierarchical networks have interesting dynamical properties because their inhabitants tend to cluster together and thus disseminate risk in particular ways.

> ...hierarchical networks are resilient to peripheral crises, but very fragile in the face of crises in the center. In these systems, the risk of contagion falls as the system integrates around the center. (Oatley, Winecoff, Pennock, and Danzman 2013, p. 135)

Thus, one possible intervention would be to try to identify regions of the network that are peripheral, and try to introduce further connections – i.e., increase integration –between them, as this can make hierarchical networks more resilient to contagion.

It is not clear that this sort of proposal could serve as a response to an impending crash, however. Another proposal that might be more effective in this regard is given by Petter, Kim, Yoon, and Han. (2002). They borrow from computer science to suggest that sometimes the performance of a system can be improved by selectively deleting vertices and edges in a network (i.e. the relationships between nodes/agents):

> If one wants to protect the network by guarding or by a temporary isolation of some vertices (edges), the most important vertices (edges), breaking of which makes the whole network malfunctioning, should be identified. (1)

Here the suggestion would be to identify, in advance, particular relationships – say, relationships between major banks, or within banks – and intervene on them when LPPLs appear in market data, perhaps by blocking information from being exchanged between particular actors.

The JLS model can also be used as a diagnostic tool for evaluating current regulatory tools. For instance, one type of intervention that is actually used as a financial regulatory tool is the "trading curb". A trading curb works by temporarily halting activity if a very large, sudden drop occurs in the stock market. For instance, the New York Stock Exchange (NYSE) currently has in place several "circuit breakers", which kick in depending on how much the Dow Jones Industrial Average (DJIA) has moved within a short period of time, with longer time-out periods for larger sudden drops.

> [T]he circuit-breaker halt for a Level 1 (7%) or Level 2 (13%) decline occurring after 9:30 a.m. Eastern and up to and including 3:25 p.m. Eastern, or in the case of an early scheduled close, 12:25 p.m. Eastern, would result in a trading halt in all stocks for 15 minutes. If the market declined by 20%, triggering a Level 3 circuit-breaker, at any time, trading would be halted for the remainder of the day. ("NYSE: NYSE Trading Information" 2016)

Circuit breakers may also be assigned to a particular stock, rather than to the market as a whole. For instance, "limit up, limit down" measures employed in some markets prevent a stock from being traded outside a certain price band for a few minutes (Pisani 2013). For instance, a 5 % movement within five minutes (e.g. say a stock drops to $5 at that time) would mean that for 15 minutes, it would not be allowed to trade for less than $5.

One motivation behind trading curbs is that in the period during the halt, investors will "calm down," i.e. behave more rationally rather than contributing further to a bubble of irrational exuberance (or pessimism). Unfortunately, some studies indicate that curbs can actually *encourage* such behavior, especially if agents know what the trading curbs are and whether the relevant limits are being approached (Goldstein and Kavajecz. 2004). The JLS model provides some insight into why this might be. In particular, if stock markets crash because of long-range correlations between traders, then a trading curb merely slows down trading, without disrupting the underlying network state that causes the crash. Worse, the trading curb itself can serve as a coordinating signal to the entire network that the market is in a precarious state, in a way that actually *increases* correlations.

4.6 Conclusion

In the foregoing, we have argued that the JLS model provides a compelling causal explanation of market crashes, with potential predictive power. The model is consistent with mainstream models in financial economics, but clearly goes beyond them – and does so by exploiting an analogy with physics. As noted in the introduction, we take this as a proof of concept: econophysics at least has the capacity to contribute to our understanding of economic phenomena, even while remaining within the general realm of mainstream economic thought.

We have also used the JLS model to explore how idealized models that use the physics of phase transitions may be used to provide a reductive explanation. We argue that the JLS model may be understood as both a minimal model and a minimalist model, and that the apparent tension between these accounts dissolves once one recognizes the different explanatory demands that a single model may be used to answer. The JLS model offers a causal explanation of why markets crash: namely, they crash because markets can evolve into states that are approximately discretely scale invariant, with long correlation lengths, such that small perturbations can have outsized effects. But this is not the only explanation one can give using the JLS model; one can also explain why crashes generically exhibit certain features, such as volatility clustering, by showing that crashes lie in a universality class, in the generalized sense described in the paper. That the same model may be used to offer two different explanations – one causal, and one, presumably, non-causal –points to the importance of separating questions concerning the explanatory purposes to which a model can be put from attempts to classify or characterize models themselves.

Chapter 5

Conclusion

5.1 Prospectus

Before summing up the main points of this dissertation, I would like to mention some problems associated with phase transitions that remain open in the philosophical literature and that I could not address here.

5.1.1 Universal Explanations

Critical phase transitions are a well-established case of universal behavior, in which one can demonstrate that systems as diverse as fluids and magnets have exactly the same critical exponents, which means that they instantiate the same macrobehavior. The existence of universal behavior is the result of the insensitivity of critical exponents to short scale effects, which is demonstrated using renormalisation group techniques. An important discussion in philosophy in the last years regards the kind of explanation that renormalization group approaches provide. Many (Batterman 2000, Reutlinger (2014) and Lange 2015) agree that renormalization group explanations are scientific examples of non-causal explanations. However, they disagree with respect to why RG explanations are non-causal. For some (Batterman 2000, Batterman and Rice 2014), renormalization group explanations are non-causal because they ignore causal details. For others (Reutlinger 2013, 2015 and Lange 2015), they are non-causal explanations because their explanatory power is due to the application of mathematical operations, which do not serve the purpose of representing causal relations. Questions that deserve to be addressed in future research are: what are the core aspects involved in renormalization group approaches that make them constitute non-causal explanations? Are all explanations involving renormalization group methods non-causal or does this depend on the question that we are asking?

5.1.2 Quantum Phase Transitons

All what has been said in this dissertation concerns classical phase transitions, but important questions arise when we consider quantum phase transitions. Landsman (2012) argues, for instance, that quantum phase transitions impose more philosophical challenges than classical phase transitions. It would be interesting to investigate to what extent the solutions to the problems that we have offered here also apply to quantum phase transitions.

Another interesting question would be whether a quantum foundation of thermodynamics can give us different insights on the problems that arise when we consider classical phase transitions. In particular, D. . Wallace (2014), 2015 argues that quantum mechanics is in a better position than classical statistical mechanics to explain many of the foundational problems that arise in the context of thermodynamics. A question that remains to be addressed is whether quantum mechanics can actually solve problems surrounding phase transitions better than classical statistical mechanics.

5.1.3 Finite Theories of Phase Transitions

Very recently, physicists have suggested alternative microscopic explanations of phase transitions that do not invoke the thermodynamic limit. At least two approaches are especially relevant for the questions that have been addressed in this dissertation: (i) the proposal that relates phase transitions with microcanonical singularities (e.g. Franzosi, Pettini, and Spinelli 2007) and (ii) the proposal that relates these processes with the topology of configuration space (Casetti and Kastner 2006). The importance of these new programs lies in the fact that they can be applied to finite systems. Given that the main problem of reducing phase transitions regards the assumption of the thermodynamic limit, the question that arises quite naturally is to what extent these new proposals provide a decisive argument in favor of the reduction of phase transitions. Philosophers taking an antireductionist position have not addressed this question yet. Similarly, the literature defending the reduction of phase transitions has generally overlooked these new approaches.

5.1.4 Defining Equilibrium for Symmetry-breaking Phase Transitions

An interesting issue that has been addressed recently concerns the differences in Boltzmann and Gibbsian's approaches to give an account for the phenomenon of spontaneous magnetization (ferromagnetic phase transition). Werndl and Frigg (2018) point out that whereas in the Gibbsian framework of statistical mechanics there can be no spontaneous magnetization (because the magnetization is zero for any arbitrary value of the temperature and any arbitrary value of N), in the Boltzmannian framework, for any arbitrary N, the magnetization will be non zero at a certain temperature, which means that using Boltzmannian framework will allow us to define spontaneous magnetization also for finite N. The latter rises many interesting questions. The first, which is addressed by Werndl and Frigg 2017, is whether this means that Gibbsian and Boltzmannian frameworks lead to different empirical results. The second, that remains to be addressed, is whether this implies that

Botzmannian approach offers a more suitable framework to account for phase transitions that involve symmetry-breaking than the Gibssian approach. Furthermore, whether this is related with the the property of ergodicity-breaking that is associated with symmetry-breaking phase transitions.

5.1.5 Analogue Experiments

During the past years, physicists have tried to gain insight into domains of nature that are beyond experimental reach by testing the hypotheses at stake in systems that are analogous to, but not identical with, the target system. For instance, in order to study properties of black holes, which are empirically inaccessible, they have recently performed experiments in analogue systems, such as fluids, which have the methodological advantage of being manipulable in the laboratory. The philosophical question that arises then is: what does justify the confirmatory power of such indirect experimental procedures? A common justification that is found in the literature hinges on the notion of universality, according to which the target and analogue systems, despite their differences, instantiate the same macroscopic behavior (Unruh and Schützhold 2005, Dardashti, Hartmann, and Thebault 2015, Dardashti, Thebault, and Winsberg 2015. However, how to demonstrate the relevant universal behavior in this context remains an outstanding problem. Questions that deserve to be addressed in future research are: How can we demonstrate the existence of universal behavior, for example in the case of black holes? Is the notion of universality present in the case of black holes the same as the one that characterizes critical phase transitions? How can we compare the strength of different notions of universality? And in what sense this a affects the confirmatory power of analogous experiments?

5.2 Summing up

An important lesson from what has been discussed here is that, when we consider the case of phase transitions, the concept of emergence must be taken with a grain of salt. Although it is true that there are good reasons to believe that phase transitions are "emergent" in some sense, this does not necessarily imply that phase transitions undermine important notions of reduction that have been at stake in the philosophical literature. In particular, in Chapter 2, I have argued that the physics of classical phase transitions are not ad odds with a notion of inter-theory reduction that compares the values of the relevant quantities in two different theories. This notion of reduction may appear weak to some philosophers, but it is enough to justify the success of the thermodynamics of phase transitions and to establish a connection between thermodynamics and statistical mechanics. More importantly, it allows us to build a connection between the macroscopic behavior of phase transitions that we observe everyday with the cooperative behavior of interacting lower level entities.

Another claim that it was made here was that the justification of infinite limits is primarily an empirical task that can be achieved if it can be shown that the limit is controllable. Although in Chapter 2 and 3, I have defended the view that we are justified in using the thermodynamic limit in the theory of phase transitions, this does not imply that we have the same justification for the use other limits in statistical mechanics. As it was shown in Chapter 3, the infinite-time limit is particularly hard to justify, because generally one does not have control over how fast this limit approaches the experimental values. It would valuable to continue investigating the role of the infinite-time limit and the possibility of offering an empirical justification for it.

Finally, in Chapter 4 I have argued that the physics of phase transitions could actually help us provide reductive explanations for stock market crashes. It would be worth investigating whether similar interpretations can be given to other models that use physics to explain cooperative behavior in social sciences such as models for vehicular traffic (e.g. Chowdhury et al. 2000) and Galam models for the process of workers' strike in big companies (Galam et al 1982) .

As we saw, the topic of phase transitions raises many foundational questions that are of interest for both physicists and philosophers. Fortunately, in the last years philosophers have begun paying attention to some of these issues, but it is clear that more work needs to be done in the future.

References

Anderson, P. W. (1972). More is different. *Science*, *177(4047)*, 393-396.

Bachelier, L., & Samuelson, A. (2011). *Louis bachelier's theory of speculation: The origins of modern finance*. Princeton University Press.

Ball, R. (2009). The global financial crisis and the efficient market hypothesis: What have we learned? *Journal of Applied Corporate Finance*.

Bangu, S. (2009). Understanding thermodynamic singularities: Phase transitions, data, and phenomena. *Philosophy of Science*.

Bangu, S. (2011). On the role of bridge laws in intertheoretic relations. *Philosophy of Science*, *78(5)*, 1108-1119.

Barenblatt, G., & Zeldovich, Y. (1971). Intermediate asymptotics in mathematic physics. *Russian Math. Surveys*, *26*, 45-61.

Batterman, R. W. (2001). *The devil in the details: Asymptotic reasoning in explanation, reduction, and emergence.* Oxford University Press.

Batterman, R. W. (2005). Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Science Part B: Studies in History and Modern Physics*, *36(2)*, 225-244.

Batterman, R. W. (2011). Emergence, singularities and symmetry breaking. *Foundations of Physics*, *41(6)*, 1031-1050.

Batterman, R. W. (2016). Intertheory relations in physics. *The Stanford encyclopedia of philosophy (2016 Edition).*

Batterman, R. W., & Rice, C. (2014). Minimal model explanations. *Philosophy of Science*.

Blanchard, O. (1979). Speculative bubbles, crashes and rational expectations. *Economics Letters*, *3*(4), 387-89.

Butterfield, J. (2011). Less is different: emergence and reduction reconciled. *Foundations of Physics*, *41(6)*(1065-1135).

Butterfield, J., & Buoatta, N. (2011). Emergence and reduction combined in phase transitions. *Frontiers of Fundamental Physics*, *FFP11*.

Callender, C. (2001). Taking thermodynamics too serously. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *32(4)*, 539-553.

Calvet, L. E., & Fisher., A. J. (2008). *Multifractal volatility: Theory, forecasting, and pricing.* Amsterdam; Boston: Academic Press.

Cardy, J. (2004). *An introduction to survival analysis using stata.* Taylor & Francis.

Casetti, L., & Kastner, M. (2006). Nonanalyticities of entropy functions of finite and infinite systems. *Physical review letters*, *97*(10).

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *http://www.tandfonline.com/doi/abs/10.1080/713665670.*.

Cottrell, A., P., C., G., M., Wright, I., & V., Y. (2009). *Classical econophysics.* Routledge.

Dardashti, R., Hartmann, S., & Thebault, K. (2015). Confirmation via analogue simulation: A bayesian analysis. *Manuscript*, Available at arXiv:1604.05932.

Dardashti, R., Thebault, K. P., & Winsberg, E. (2015). Confirmation via analogue simulation: what dumb holes could tell us about gravity. *British Journal for the Philosophy of Science*, *1*(68), 55-89.

Domb, C. (2000). *Phase transitions and critical phenomena* (Vol. Vol. 19). Academic Press.

Earman, J., & Rédei, M. (1996). Why erogic theory does not explain the success of equilibrium statistical mechanics. *The British Journal for* *the Philosophy of Science*.

Ehrenfest, P., & Ehrenfest-Afanassjewa, T. (1907). Über zwei bekannte einwände gegen das boltzmannsche h-theorem. *Hirzel*.

Emch, G., & Liu, C. (2013). *The logic of thermostatical physics.* Springer Science & Business Media.

Fama, E. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1).

Fisher, M. E. (1998). Renormalization group theory: Its basis and formulation in statistical physics. *Reviews of Modern Physics*, *70*(2).

Fletcher, S. C. (2015). Similarity, topology and physical significance in relativity theory. *The British Journal for the Philosophy of Science*.

Franklin, A. (2017). On the renormalisation group explanation of universality on the renormalisation group explanation of universality on the renormalization group explanation of universality. *Philosophy of Science*.

Franzosi, R., Pettini, M., & Spinelli, L. (2007). Topology and phase transitions i. *Nuclear Physics B*, *782*(3), 189-218.

Frigg, R. (2008). A field guide to recent work on the foundations of statistical mechanics. In *The ashgate companion to contemporary philosophy of physics.* Ashgate.

Gallavotti, G. (1999). *Statistical mechanics: A short treatise.* Springer-Verlag Berlin Heidelberg.

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

Goldstein, M. A., & Kavajecz., K. A. (2004). Trading strategies during circuit breakers and extreme market movements. *Journal of Financial Markets*, *7*(3).

Hsiao, P., & Monceau, P. (2002). Critical behavior of the three-state potts model on the sierpinski carpet. *Physical Review B*, *65(18)*(184427).

Hüttermann, A., & Love, A. (2016). Reduction. In P. e. Humphried (Ed.), *The oxford handbook of philosophy of science. new york.* Oxford Universoty Press.

Ilinski, K. (1999). Critical crashes? *International Journal of Modern Physics C*, *10*(4).

Itakura, M. (2003). Monte carlo renormalization group study of the heisenberg and the xy antiferromagnet on the stacked triangular lattice and the chiral ϕ 4 model. *Journal of the Physical Society of Japan*.

Jhun, J., Palacios, P., & Weatherall, J. O. (in press). Market crashes as critical phenomena: Explanation, idealization, and universality in econophysics. *Synthese*.

Johansen, A., Ledoit, O., & Sornette, D. (2000). Crashes as critical points. *International Journal of Theoretical and Applied Finance*, *3*(2).

Jones, N. (2006). *Ineliminable idealizations, phase ineliminable idealizations, phase transitions, and irreversibility* (Unpublished doctoral dissertation). The Ohio State University.

Joshi, M. S. (2008). *The concepts and practice of mathematical finance.* Cambridge; New York: Cambridge University Press.

Kadanoff, L. P. (2009). More is the same: Phase transitions and mean field theories. *Journal of Statistical Physics*, *137(5-6)*.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6).

Landsman, N. P. (2013). Spontaneous symmetry breaking in quantum systems: Emergence or reduction? *Studies in History and Philosophy of Science Part B: Studies in History and Modern Physics*, *44(4)*(379-394).

Lanford, O. (1973). Entropy and equilibrium states in classical statistical mechanics. In A. Lenard (Ed.), *Statistical mechanics and mathematical problems.* Springer-Verlag.

Le Bellac, M., Mortessagne, F., & Batrouni, G. (2006). *Equilibrium and non-equilibrium statistical thermodynamics*. Cambridge University Press.

Lebowitz, J. (1999). Statistical mechanics: A selective review of two central issues. *Reviews of Modern Physics*.

Lo, A. W., & Mueller., M. T. (2010). Warning: Physics envy may be hazardous to your wealth! *SSRN Scholarly Paper ID 1563882*.

Ma, S.-K. (1976). Renormalization group by monte carlo methods. *Physical Review Letters*, *37(8)*(461).

Malament, D., & Zabell, L. (1980). Why gibbs phase averages work: the role of ergodic theory. *Philosophy of Science*.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(4), 394–419.

Mantegna, R. N., & Eugene, S. H. (1999). *Introduction to econophysics: Correlations and complexity in finance*. Cambridge University Press.

McCauley, J. L. (2004). *Dynamics of markets: Econophysics and finance.* Cambridge University Press.

Menon, T., & Callender, C. (2013). Ch-ch-changes philosophical questions raised by phase transitions. *The Oxford handbook of philosophy of physics*.

Morrison, M. (2012). Emergent physics and micro-ontology. *Philosophy of Science*.

Nickles, T. (1973). Two concepts of intertheoretic reduction. *The Journal of Philosophy*, *70(7)*(181-201).

Norton, J. D. (2012). Approximation and idealizations: Why the difference matters. *Philosophy of Science*, *79(2)*, 207-232.

Norton, J. D. (2013). Confusions over reduction and emergence in the physics of phase transitions'. *available on Norton's website at: http://www.pitt.edu/jdnorton/Goodies/reduction emergence/red em.html*.

Oatley, T., Winecoff, W., Pennock, A., & Danzman, S. B. (2013). The political economy of global finance: A network model. *Perspectives on Politics*, *11*(1), 133-53.

Osborne, M. F. M. (1959). Brownian motion in the stock market. *Operations Research*, *7*(2), 145-73.

Pawley, G., Swendsen, D., R.H. Wallace, & Wilson, K. G. (1984). Monte carlo renormalization-group calculations of critical behavior in the simple-cubic ising model. *Physical Review*, *29*(4030).

Petter, H., Kim, B. J., Yoon, C. N., & Han., S. K. (2002). Attack vulnerability of complex networks. *Physical Review E*, *65*(5).

Pisani, B. (2013). New trading curbs: Limit up, limit down is finally here. *CNBC. April 4. http://www.cnbc.com/id/100617758.*.

Reutlinger, A. (2014). Why is it there universal macrobehavior? renormalization group explanaton as noncausal explanation. *Philosophy of Science*, *81*(5).

Rohrlich, F. (1988). Pluralistic ontology and theory reduction in the physical sciences. *The British Journal for the Philosophy of Science*, *39.3*, 295-312.

Rueger, A. (2000). Physical emergence, diacronic and synchronic. *Synthese*, *124*, 297-322.

Samuelson, P. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, *6*(2), 41-49.

Santos, M. S., & Woodford, M. (1997). Rational asset pricing bubbles. *Econometrica*.

Scheibe, E. (1998). *Die reduktion physikalische theorien: Ein beitrag zur einheit der physik*. Springer-Verlag Berlin Heidelberg.

Shiller, R. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, *15*(2).

Sklar, L. (1995). *Philosophical issues in the foundations of statistical mechanics*. Cambridge University Press.

Sornette, D. (2000). *Critical phenomena in natural sciences*. Springer-

Verlag.

Sornette, D. (2003). *Why stock markets crash: Critical events in complex financial systems*. Princeton University Press.

Sornette, D., Johansen, A., & Bouchaud, J.-P. (1996). Stock market crashes, precursors and replicas. *Journal de Physique*, 167–75.

Sornette, D., & Malevergne, Y. (2001). From rational bubbles to crashes. *Physica A: Statistical Mechanics and Its Applications, Application of Physics in Economic Modelling*, *299*(1-2), 40–59.

Sornette, D., Woodard, R., Yan, W., & Wei-Xing, Z. (2013). Clarifications to questions and criticisms on the johansen–ledoit–sornette financial bubble model. *Physica A: Statistical Mechanics and Its Applications*, *392*(19), 4417-28.

Uffink, J. (2007). Compendium of the foundations of classical statistical physics. In J. Butterfield & J. Earman (Eds.), *Philosophy of physics.* Amsterdam: North Holland.

Unruh, W. G., & Schützhold, R. (2005). Universality of the hawking effect. *Physical Review D*, *71*(1).

van Lith, J. (2001). *Stir in sillness: A study in the foundations of equilibrium statistical mechanics.* (Unpublished doctoral dissertation). University of Utrecht.

v. Bothmer, G., & Meister., C. (2003). Predicting critical crashes? a new restriction for the free variables. *Physica A: Statistical Mechanics and Its Applications 320*, 539-47.

Wallace, D. . (2014). Probability in physics: stochastic, statistical, quantum. In A. Wilson (Ed.), *Chance and temporal asymmetry.* Oxford University Press.

Wallace, D. (2015). The quantitative content of statistical mechanics. *Studies in the History and Philosophy of Modern Physics*, *52*, 285-293.

Weatherall, J. (2013). *The physics of wall street: A brief history of predicting the unpredictable.* Boston, MA: Houghton Mifflin Harcourt.

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*.

Werndl, C., & Frigg, R. (2017). Mind the gap: Boltzmannian versus gibbsian equilibrium. *Philosophy of Science*, *84*(5).

Wilson, K., & Kogut, J. (1974). The renormalization group and the e expansion. *Physics Reports*, *12(2)*, 75-199.

Woodward, J. (2003). *Making things happen: A causal theory of explanation.* Oxford: Oxford University Press.

Zinn-Justin, J. (2007). *Phase transitions and renormalization group.* OUP Oxford.