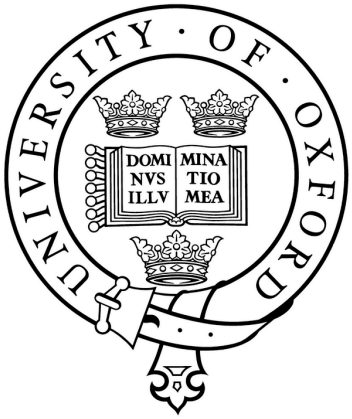
# Is More Different?
# Emergent Properties in Physics

Paul Mainwood
Merton College, University of Oxford




Submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy

Trinity Term, 2006

**Abstract**

This thesis gives a philosophical assessment of a contemporary movement, influential amongst physicists, about the status of microscopic and macroscopic properties. Although it is a recognisable version of older metaphysical theses of emergentism, these "New Emergentists" support their position by appealing to recent discoveries in condensed matter physics. The fountainhead for the movement was a short 1972 paper 'More is Different', written by the physicist Philip Anderson. Each of my chapters is concerned with themes mentioned in that paper, or subsequently expounded by Anderson and his followers.

In Chapter 1, I aim to locate Anderson's existence claims for 'emergent properties' within the metaphysical, epistemological and methodological doctrines that identify themselves as 'emergentist'. I argue, against the commentators' consensus, that the New Emergentists make claims about the metaphysical status of physical properties, and should not be read as concerned only with matters of research methodology for physics.

In Chapter 2, I look at the physical examples that the New Emergentists appeal to, and propose a way of formulating their main claims within modern analytic metaphysics. I argue that it is possible to view their thesis as an updated version of 'British Emergentism' a movement popular in the early years of the twentieth century. I support this contention by comparing examples of emergent properties put forward by the British and the New Emergentists.

Chapter 3 is a discussion of the significance of renormalisation techniques, a theme not present in 'More is Different', but prominent in Anderson's later work and emphasised by his followers. I concentrate on a set of claims by Robert Batterman, who presents renormalisation as an explanatory strategy unrecognised by the philosophy of science. I attack his thesis by separating several different methods of renormalisation analysis, and arguing that he has conflated them. I go on to examine some seemingly contradictory views of the significance of renormalisation methods, and present an interpretation that goes some way towards a reconciliation.

In Chapter 4, I discuss the theoretical representation of phase transitions in condensed matter physics; in particular, their appeal to the limit of an infinite system. While this theme is mentioned in Anderson's work, I focus on a debate which has arisen independently, within the philosophical literature. I examine and refute various claims to the effect that the ineliminability of the infinite limit in modelling phase transitions is of great metaphysical significance. I suggest a definition of phase transitions for finite systems that dissolves this illusion, and gives us reason to trust the results of the theories that only apply in the infinite limit. I then comment on the significance of these results for a suggestion of Laura Ruetsche, that quantum statistical mechanics should be used as a guide to the interpretation of quantum field theory.

Chapter 5 revisits the doctrines of the New Emergentists in order to place their views within some wider philosophical debates. I look at how their doctrines bear on issues in the interpretation of quantum mechanics, in the philosophy of mind and the philosophy of science. I close by returning to the context within physics, in which the New Emergentists originally made their presence felt: the controversy over whether elementary particle physics should enjoy greater status than other areas.

**Acknowledgements**

I would like to thank Merton College both for their hospitality during my time as a D.Phil student and for generous financial support. Their Domus Scholar programme has enabled me to conduct my studies in the most welcoming and intellectually stimulating environment. They also allowed me to take part in the Oxford-Princeton Link programme, and I am most grateful to The Graduate College of Princeton University for hosting me as a visiting scholar during the fall semester of 2004.

My intellectual debts to Jeremy Butterfield are vast. As supervisor for both my B.Phil and D.Phil, he has been outstandingly generous with his guidance, advice and encouragement. His influence on the development of my work has been profound, and I consider myself truly privileged to have benefited from his teaching. In addition to his academic role, his personal support and friendship have been invaluable to me, throughout the time that I have known him.

I am lucky enough to have spent my time as a graduate student in close contact with the Philosophy of Physics Group at Oxford. Members of the group have provided invaluable comments on early versions of material in this thesis, and I thank especially Harvey Brown, Simon Saunders, David Wallace and Oliver Pooley.

During my stay in Princeton University, I was privileged to be supervised by Bas van Fraassen, whose eminence in the subject is matched only by his generosity with advice and time for his students. I also benefited from conversations and teaching from Phillip Petit, Adam Elga, Frank Arntzenius and especially from Hans Halvorson.

Fellow graduate students to whom I owe special thanks for inspiration and conversation include David Baker, Anil Gomes, Hilary Greaves, Dennis Lehmkuhl, Julia Markovits, Ofra Magidor, Acer Nethercott, Justin Pniower, Richard Price, Chris Timpson and Alastair Wilson.

Lastly, thanks are due to my family: my mother and father for the opportunities they have provided, and together with my sister, for their loving support. Despite my choosing to undertake interminable academic studies in increasingly arcane subjects, this support has never wavered.

# Contents

# Chapter 1

# New Emergentism: Metaphysical Positions

Throughout the twentieth century, those scientists and philosophers of science willing to call themselves 'emergentists' have been engaged in a slow, but steady retreat across subject areas. The term was first used in its modern sense by Lewes (1877), and quickly adopted by a number of writers. Although they differed on many points of doctrine, they were united by a belief that certain large systems studied in the special sciences — in chemistry, biology, or psychology — exhibited fundamentally different properties to those of the very smallest systems studied in physics. In the first decades of the century, one could find emergentists in certain areas of large-scale physics; then the front line shifted to chemistry and then to biology, as a steady march of reductionist success pushed emergentists to less developed areas of science (McLaughlin, 1992). In the late twentieth century it seemed that psychology and the sciences of the mind were their last bastion: everywhere else, a reductionist programme had triumphed. The sciences of the very small — high energy particle physics in particular — were seen as the only areas engaged in a search for "fundamental" laws. All other physical sciences were

in a sense "derivative", in that their subject matter was, ultimately, the behaviour of physical systems, and this could be derived — in theory at least — from the laws and principles that governed the very small.

However, the last few decades have seen a strong emergentist comeback hailing from an unexpected quarter: physics itself, which had been thought to be the first to fall. The variety of fields claiming to study "emergent phenomena" has now reached such a point that rival unificatory frameworks have been proposed. In these first years of the twenty-first century these include complex adaptive systems theory, nonlinear dynamical systems theory, self-organised criticality, and the theories based around the study of cellular automata.[1] These frameworks are claimed to have applications well beyond physics, in biology, social theory, economics and population biology, amongst others.

Much of this startling 'resurgence of emergence' can be traced back to a short 1972 paper by the Nobel Prize winning physicist Philip Anderson, entitled 'More is Different'.[2] Anderson defended the view that the laws and principles he studied as a condensed matter physicist were emergent, in the sense that they were entirely different from, but had no lower status than, those studied by particle physicists. In contrast to some of the later developments, Anderson's own claims were based on well-understood physical theories. Nonetheless, his position was, and remains, a controversial one. Most famously, Anderson was opposed by Steven Weinberg, who claimed that the physics of the very small held a privileged position in a

[1] Classics of complexity studies include Holland (1992) and Kauffman (1996). The field is particularly associated with the Santa Fe Institute in California, which recently celebrated its twentieth anniversary. Waldrop (1992) gives a history of thinkers and ideas associated with the Santa Fe institute, which is useful in tracing its origins and connections to other "emergentist" thought. Nonlinear dynamical systems theory developed from theories associated with chaotic phenomena. Kellert (1993) provides a detailed philosophical study, which is largely deflatory as to its metaphysical significance. Newman (1996) takes the opposite view. Self organised criticality is associated expecially with the work of Bak et al. (1987), and theories of cellular automata have been brought to public prominence by Wolfram (2002).

[2] Anderson (1972) is traced as the fountainhead article for emergentist and complexity theories by several authors, (e.g., Solomon, 2005).

hierarchy of scientific explanation.

I shall postpone an examination of the details of the Anderson/Weinberg debate.[3] My aim, in this Chapter and the next, is to try to relate this new doctrine of emergence, and its opponent reductionism, to the metaphysical debates and doctrines that have long existed under the same names. This project comes in two parts. The first — covered in this chapter — aims to locate the basic metaphysical position taken by Anderson and his followers amongst those familiar from the philosophy of mind and of science. I shall argue that their new position does not fall easily into the most familiar metaphysical categories, and I shall make some suggestions as to why this might be so. These largely negative conclusions will then inform the second part of the project, which constitutes the next chapter. There, I argue that Anderson's own emergentist position is a metaphysical thesis of more originality, plausibility and importance than its critics have allowed.

## 1.1 Preliminaries

It is helpful to start by introducing some nomenclature. I shall refer to the position first articulated by Anderson as 'New Emergentism', a term that may serve to distinguish it from the bewildering variety of positions that have precipitated out from other emergentist debates. 'New Emergentism' is intended to cover a reasonable range of the positions held by those modern physicists who call themselves emergentists, but it would be a mistake to try to accommodate too many of the claims made since Anderson's original, concise presentation. Here I shall perhaps err on the side of caution and only appeal to a few authors to articulate the central position. As well as Anderson, his colleagues in condensed matter physics —

[3] In this thesis, I will return to it only in §5.2, considering an argument based on the concept of a *protectorate*, which is separate to the considerations I shall be engaging with in this chapter and the next.

Robert Laughlin, David Pines and Piers Coleman — will be quoted as representing New Emergentism position due to their explicit identification with Anderson's themes (Laughlin and Pines (2000), Coleman (2003), Laughlin (2005)). Ernst Mayr's discussions, and especially his position in exchanges with Weinberg, also put him squarely in the New Emergentist camp, though he is principally concerned with issues in biological sciences (Mayr, 1988).

It is also useful to reserve judgment on the exact branch of physics that may describe the very small, so for the moment I shall speak of "microphysics" and "microphysical laws" leaving it open whether these terms refer to the current Standard Model of particle physics, to a modification of it in some successful "grand unification", or to some advance as yet unimagined.

It is helpful to begin with a short exercise of undergrowth clearing, to address several points that appear to have a clear resolution, before engaging with more controversial issues.

### 1.1.1 Red Herring 1 — A Comprehensive Definition

John Holland, a guru of Complexity Theory, introduces a semi-popular exposition of emergence with some pessimism:

> Despite its ubiquity, emergence is an enigmatic, recondite topic, more wondered at than analyzed. What understanding we do have is mostly through a catalogue of instances ... It is unlikely that a topic as complicated as emergence will submit weakly to a concise definition. (Holland, 1998)

Claims that a field is dealing with "emergent phenomena" have become fashionable in the years since Anderson's article. The list of the last section could be extended a great deal further without even venturing beyond physics. Why, in all this variety, should we expect to find some core criteria, shared also by older "emergentist" positions in the philosophy of mind and other areas of science?

A source of much difficulty is that both philosophers and scientists have typically been driven by their own motivations in searching for emergence, and have allowed these to influence their selection of criteria. Kim (1995) is concerned with issues of mental causation, and so it is natural for him to search for emergence as defined by some appearance of causal autonomy. But contrast his approach with that of Zurek (1991), who is concerned with issues of how classical behaviour emerges from the quantum domain, and looks for it in the vanishing of the interference terms from a decohering wavefunction. It is hard to see how these two can be covered by a single notion of emergence.

Similar parochial concerns can be expected to appear in any approach which first constructs a firm list of desiderata for emergence, then searches for examples which conform to it. There is no reason to think that a single definition will be suitable for all these hoped-for applications of emergence. In the philosophical literature in particular, background interests have shown through strongly, despite pleas to search for examples of emergence in physical science on their own merits, e.g., Humphreys (1997, 15). I claim that the New Emergentists deserve special attention for taking the opposite tack: looking within well-understood physical sciences for simple phenomena that seem to display a few core emergentist characteristics, illustrating fundamental laws, or principles operating on a large scale, and then examining the theoretical descriptions to see whether they might suggest criteria for a more general approach.

Thus, our aim is *not* to discover any neat package of necessary and sufficient conditions for emergence. Far less are we searching for a definition covering all those desiderata needed to claim that mental causation, the classical world or a direction of time are emergent. For these goals, Holland's scepticism appears well-founded. Neither is it my ambition to provide a framework for the new sciences of emergence, by defining such terms as 'complex' or 'self-organised'. Such a

clear framework is certainly needed, the field being often compared to pre-Carnot thermodynamics. (I shall make a few comments of the relationship between the New Emergentist position and Complexity, in §5.1.)

### 1.1.2 Red Herring 2 — Emergence of What?

Silberstein (2002) gives a careful survey of three separate emergentist debates in the philosophical literature. They concern the status of physical entities, of causal powers and of properties. A good first step would be to identify the issue with which the New Emergentists are concerned.

Fortunately, it is clear that the New Emergentists' interest is in property emergence (understood to encompass also those laws which govern the emergent properties). As regards "entity emergence" consider the following from Laughlin and Pines:

> The low-energy excited quantum states of these systems [crystalline solids] are particles in exactly the same sense that the electron in the vacuum of quantum electrodynamics is a particle: They carry momentum, energy, spin, and charge, scatter off one another according to simple rules, obey Fermi or Bose statistics depending on their nature, and in some cases are even "relativistic," in the sense of being described quantitatively by Dirac or Klein-Gordon equations at low energy scales. *Yet they are not elementary*, and, as in the case of sound, simply do not exist outside the context of the stable state of matter in which they live. (Laughlin and Pines, 2000, 28, italics mine)

This attitude is typical — the New Emergentists accept that there *may* be an ultimate microlevel of entities that can usefully be called elementary, or even 'fundamental', but they deny that this confers any privileged status on the properties of these elementary entities, or on the laws governing those properties.

While the New Emergentists can concede the issue of entity emergence for the sake of argument, they never even mention issues of causation. Their writings contain no discussions of overdetermination, downward causation, or of the

causal closure of the microphysical. In these two chapters we shall follow them in their silence, but in doing so we may be accused of bypassing a large portion of the philosophical debates on emergence, especially those within the philosophy of mind.

I think this is defensible, since a focus on causal emergence is largely driven by concerns particular to the philosophy of mind. As one of the foremost metaphysicians of the mental, Jaegwon Kim is able to look back over his career and write (with approval), that 'devising an account of mental causation has been, for the past three decades, one of the main preoccupations of philosophers of mind who are committed to physicalism in one form or another.' (Kim, 2005, 8). A physicalist will be keen to discover whether there is a conception of emergence that might imbue mental states with causal roles suitable for a robust conception of mental causation.[4]

Important and wide-ranging these debates may be: but we must leave them aside if we are to understand the New Emergentists, for issues of causation are by no means apparent in their principal examples. In particular, they place symmetry-breaking centrally in their discussion, a phenomenon with notorious causal peculiarities. Philosophers of physics have discussed both (i) general questions as to whether causal powers can be isolated cleanly in such processes (Liu, 2002), and (ii) particular issues such as apparent violations of Curie's Principle, which relates the symmetry characteristics of causes and effects (Brading and Castellani, 2003).

Meanwhile, the causation-concerned philosophers of mind can be accused of being overly influenced by "ball-bearing physics",[5] based on examples taken from

[4]Kim himself is the foremost proponent of the view that emergence (or at least the British version described in §1.5.2) was unavoidably committed to a doctrine known as 'downward causation', and that this is in contradiction with the causal closure of the physical realm (Kim, 1998, 100-104). A brief discussion of the sense in which British Emergence is committed to a version of downward causation, is found in Appendix A.

[5]Indeed they often *have* been so accused: (e.g., Wilson (1985, 2004), Norton (2003)). Suitable notions of causation are particularly elusive in physical theories built on statistical mechanical

elementary Newtonian mechanics. Once we take account of the difficulties of discerning causal processes in a more realistic physics, causal considerations become too elusive to serve as a useful guide on issues of emergence. At any rate, the New Emergentists steer well clear.

So, from Silberstein's list, we are left with property emergence. The New Emergentists themselves discuss the emergence of novel 'behaviour' of systems, and 'emergent phenomena' as well as referring directly to emergent properties, and emergent laws governing these properties. I shall try to avoid the rather vague terms 'phenomena' and 'behaviour', which could cover a whole range of token or type events, as well as failing to distinguish law-governed events from mere accidental regularities. I shall focus on emergent *properties*. Let us introduce a further useful term: a property of a system made up from many parts will be called a *systemic* property.[6]

As well as the status of systemic properties themselves, the New Emergentists are also concerned with the status of the laws that govern them. They are often called 'higher-level' laws, and take on their distinctions: 'emergent' and 'non-emergent'. This immediately leads to the question of whether the status of properties determines the status of the laws that govern them, or vice versa. In what follows, I shall assume that is an issue to be decided by the particular metaphysical account of laws that we may wish to adopt. For example, a Humean account would treat properties as primary, laws as derivative. One advocating necessary laws is likely to reverse this order of priority.

---

principles, and in this connection it may be significant that Anderson and his principal allies — Coleman, Laughlin and Pines — are all condensed matter theorists, an area in which statistical mechanics is used extensively.

[6]In keeping with common usage, the term 'properties' should also be understood to encompass relations. I shall concentrate on systemic properties which are intrinsic to that system, and assume that any relations appealed to will be non-external (in Lewis' 1983 sense).

### 1.1.3 Red Herring 3 — Physics and Metaphysics

One of the basic questions we encounter in interpreting the New Emergentist claims, is whether their thesis is meant to be:

a) one of metaphysics – about the metaphysical status of emergent properties and laws; or

b) one of epistemology and the methodology of science – about the status of experimental and theoretical sciences of large-scale systems as compared to microphysics.

In their canonical statements the New Emergentists are unequivocal: they are making full-blooded metaphysical claims; that is, they are concerned with the metaphysical status of laws and properties in the world, not just about the research that investigates them. Yet this can be obscured by the fact that they do not base their claims on explicitly metaphysical arguments, but on evidence from the practice of physics and other sciences. This leads the reader to wonder whether the New Emergentist claims have support only as epistemological and metholodogical points about the practice of science.

Anderson himself is an offender on these grounds. He starts *More is Different* with claims about the fundamental nature of properties and laws, rejecting the view that: 'the only people studying anything fundamental, are those studying microphysical laws.' But then he slides to discussing the fundamental nature of the *study*, and defends it as 'research which is as fundamental as any other'. Yet he finishes back with metaphysical lessons, speaking of the properties of states of large systems being 'not only more than, but very different from the sum of its parts'. Elsewhere, he speaks in wholly objective terms, without any mention of scientific research at all, describing 'an "emergent property": a property which is

manifested only by a sufficiently large and complex system by virtue of that size and complexity' (Anderson, 1989, 586).

These confusions are not helped by the New Emergentists' use of terms such as 'fundamental', 'basic' and 'elementary' to refer both to the status of laws and properties themselves, and to the theories that are supposed to describe them. The word 'fundamental' is particularly difficult, since as we have already seen, the New Emergentists happily grant that certain small entities can be called 'fundamental', even calling the properties that belong to them, 'fundamental properties'. Yet they deny that this implies a privileged status for the properties of very small systems. This is further complicated by the fact that they also attach the term 'fundamental' to *sciences* — the study of both entities and properties — and here they *do* imply an associated privilege, for they deny that the sciences of the very small enjoy a 'fundamental status'.

Worse, when speaking of 'fundamental laws', the New Emergentists do not seem consistent in their usage at all. Take for example Anderson's response to questioning by the congressional panel in the hearings on the SSC: 'These new laws don't contradict the laws of elementary particles people discover; they are simply independent of them, and I would argue that they in no way any less fundamental' (Cat, 1998, 265n). (Although this is a statement about the metaphysical status of laws, Anderson seems unable to express his views without mentioning the methodological issue of their discovery.) But in 'More is Different' he refers happily to microphysical laws as 'fundamental laws', without any hint that this this is in any conflict with his own position (Anderson, 1972, 393).

Similar remarks could be made about the terms 'basic' and 'elementary', for it is trivially true that we can associate these terms with the properties of any of the smallest objects in our ontology, if all others are made up from composites of them. The New Emergentists do not challenge this trivial sense, but rather

the inference that this imbues properties or laws of the smallest entities with a privileged metaphysical status.

As I shall discuss in §1.4, a popular reaction amongst philosophers is that the New Emergentists in general, and Anderson in particular, should not make metaphysical claims, and are simply confused when they do so. On this view, the only defensible interpretation of New Emergentism is as a fairly unremarkable position within the methodology and epistemology of science. It will be one of the main points of this chapter and the next to deny this response. I believe that the explicit statements of Anderson and his followers can — and should be — taken seriously as outlining a coherent metaphysical position. Of course this does not mean that they spell out their position in the form of metaphysical theses familiar to philosophers.

### 1.1.4 Prospectus and Statement of Purpose

With these issues cleared away, we should be ready to begin the task of setting New Emergentist claims in a clear philosophical framework. We shall try to locate their thesis as a fully metaphysical position, though one informed by physics. What I mean by that rather cryptic phrase is the following: while they are not committed to a detailed metaphysical position, they hold some doctrines that are unquestionably metaphysical in nature, and they do so for reasons stemming from physical theory. I shall follow them, not by seeking to force them into an over-detailed metaphysical position, but rather by attempting to disambiguate and interpret those claims that they are committed to. This project is a little more ambitious than it may appear. The main reason to interpret New Emergentist claims as methodological, is because they appear incoherent when interpreted as metaphysical doctrines. Part of the project will be to argue that these interpretations are based on misunderstandings.

By saying above that the New Emergentists' position is informed by physics, I do not mean just that they motivate their doctrines with examples drawn from their expertise as condensed matter physicists. Rather, their metaphysical claims are motivated by a desire to be faithful to the discoveries of our best physical theories, and they believe that to hold microphysical laws and properties as privileged, is in conflict with this practice. This attitude will come out strongly in Chapter 2 where we take an approach to properties directly influenced by the need to accommodate the enormous range admitted by physical theories.

The rest of the chapter will be organised as follows. In the next section (§1.2) I shall try to bring out the central claims of the New Emergentists, expressed as concise statements of philosophical doctrine. Next (§1.3—§1.5), I shall look at how each of these New Emergentist claims has been made precise in orthodox philosophical discussion. I shall then argue that these interpretations fail to give an adequate appreciation of their position (§1.6 & §1.7). Finally, in §1.8, I argue that while the New Emergentist position resembles one of the orthodox approaches in many respects, their explicit claims mean that they cannot occupy the very same ground. And I give a short argument to show that their position is rather tightly constrained by their other claims. This looks ahead to Chapter 2, which builds on these rather negative conclusions to construct a positive interpretation of the New Emergentists' metaphysical position.

## 1.2 New Emergentism: The Central Position

Selecting a few of the many claims made by physicists calling themselves "Emergentists" as "The Central Position", may seem an exercise in arbitrary selection. Yet I claim that those chosen here, expressing their acceptance of microphysicalism, and the promotion of some systemic properties as importantly novel, are both

spanning and minimal.

By "spanning" I mean that they are accepted by all — or at least the vast majority — of those marching under a New Emergentist banner. Rather stronger: I believe that they are central to the physicalist, yet inclusive metaphysical picture that tends to be tacitly assumed by physicists working in 'emergentist' areas. (Admittedly, it is hard to justify either of these claims without some sort of opinion poll; philosophical analysis can only go so far).

By "minimal" I mean that the claims are about as weak as can be made while still generating controversy. In particular the claims are rejected by self-declared opponents of the New Emergentists, who hail especially from fields such as particle physics and string theory. Steven Weinberg is probably the best-known example of a "New Reductionist",[7] but joining him will be anyone who accords a metaphysical priority to the discoveries of microphysics above those of the rest of physics.

### 1.2.1 Physicalism and Microphysicalism

The New Emergentist movement is rooted in physics and as such, its adherents subscribe to a physicalist world view. Although they make some claims that might be thought unorthodox, they will have no truck with anti-physicalist doctrines such as Cartesian dualism, vitalism or epiphenomenalism, which they would doubtless consider "spooky" and unscientific. The project of turning physicalism into a precise philosophical claim has of course been the focus of much discussion over the last few decades, but the New Emergentists help us to identify their position, by also accepting something stronger:

> The reductionist hypothesis may still be a topic for controversy among philosophers, but among the great majority of active scientists I think

[7] Weinberg is also the principal opponent of Anderson in public debates on the issue. In §5.3 I address Weinberg's position and the New Emergentists' criticisms of it.

> it is accepted without question. The workings of our minds and bodies, and of all the animate or inanimate matter of which we have any detailed knowledge, are assumed to be controlled by the same set of fundamental laws, which except under certain extreme conditions we feel we know pretty well. (Anderson, 1972, 393)

Anderson later lists those who study such laws, and it becomes clear that he is speaking of *microphysical* laws, those that govern the tiny elementary building-blocks which we believe make up our physical world.[8]

Ernst Mayr is even clearer than Anderson when he accepts what he refers to as 'constitutive reductionism' in the context of biology:

> ... the material composition of organisms is exactly the same as that found in the inorganic world. Further ... none of the events and processes encountered in the world of living organism is in any conflict with the physico-chemical phenomena at the level of atoms and molecules. (Mayr, 1982, 59)

It is rather startling to hear this thesis accepted by those calling themselves emergentist, and even more startling to hear it described as 'reductionism', a doctrine usually understood as a stronger claim about the definability in a certain sense of one theory in terms of another; (this sense of 'reduction' will be the focus of §1.5.1). But this seems a minor matter of a mismatch in terminology; in metaphysics, the claim endorsed by Anderson and Mayr is usually called *microphysicalism*.

> [M]icrophysicalism ... is the doctrine that actually (but not necessarily) everything non-microphysical is composed out of microphysical entities and is governed by microphysical laws. (Pettit, 1994)

Pettit also brings out the two components which constitute microphysicalism.

> The first is that everything in the empirical world — every particular in that world — is composed in some sense out of subatomic materials. And second is that everything that happens in the empirical world

[8] What should we make of Anderson describing these as 'fundamental', which he repeats several times? As outlined in §1.1.3, I think he means no more than to point out that they are the laws obeyed by the fundamental microphysical entities.

> happens, ultimately, under the controlling influence of subatomic forces and laws. (Pettit, 1996, 342-3)

As we have seen, the New Emergentists embrace *both*.

### 1.2.2 The 'Novelty' of Emergent Properties

While accepting the core claims of microphysicalism, New Emergentists deny an 'at first sight obvious corollary':

> ... that if everything obeys the same fundamental laws then the only scientists who are studying anything really fundamental are those who are working on those laws. (Anderson, 1972, 393)

Just how obvious this may seem is shown by Pettit's immediate continuation from his statement of microphysicalism quoted above:

> There are higher level laws but they do not represent forces independent of those associated with the subatomic. Thus the higher level laws do not come into conflict with the subatomic, they do not reinforce the subatomic in the fashion of another shoulder at the wheel, and they do not take up any indeterminacy or slack that the subatomic laws leave. (Pettit, 1996, 343)

These claims, the New Emergentists would *not* accept, nor Pettit's characterisation of microphysicalism as involving a belief that 'the microphysical levels — are more fundamental' (342). Anderson holds in opposition, '[i]nstead at each level of complexity *entirely new* properties appear' (Anderson, 1972, 393, italics his).

The requirement that emergent properties must be in some interesting sense 'entirely new' appears as a central claim in almost every field in which emergentism is defended. In both his identification of the intuition and his immediate response, I agree with Tim Crane:

> The intuitive idea of an emergent property is the idea of a "novel" property of a whole or complex that emerges from the parts of the whole and the way the parts are put together. But how should we make this idea more precise? (Crane, 2001, 211)

For if we try to pin down what we mean when we call certain systemic properties "emergent" in virtue of their novelty, it appears difficult to negotiate the twin pitfalls of triviality and unreasonable strength.

For example, take this statement from John Wheeler:

> When you put enough elementary units together, you get something that is more than the sum of these units. A substance made of a great number of molecules, for instance, has properties such as pressure and temperature that no one molecule possesses. It may be a solid or a liquid or a gas, although no single molecule is solid or liquid or gas. (Wheeler and Ford, 1998, 341)

To propose that a systemic property qualifies as "novel" because the property is not possessed by any of the parts is far too weak for a metaphysically interesting sense of emergence: for under Wheeler's definition an "emergent" property could be as trivial as the total mass of a composite system. For that property is not possessed by any of its (proper) parts, but is possessed by the whole. To exclude these most trivial examples Wheeler, (in common with many other physicists associated with the New Emergentist camp)[9] claims here that the whole must be 'more than the sum of its parts.' He cannot mean "sum" (or "more") literally, for the total mass of a composite system with any energetic interaction is more than the sum of the masses of the parts; presumably then, he means the term to cover any simple operation of composition. But how simple should we allow the operation to be?

A natural strengthening of this criterion is to hold that to be a candidate for emergence, a property of a system must be not just a new property, but a new determinable. For example, Spencer-Smith fills this out in detail: 'a property $P$ is *novel* in $x$ if $x$ has $P$, and there are no determinates $P'$ of the same determinable as $P$, such that any constituents of $x$ have $P$' (Spencer-Smith, 1995, 117). He thinks that novelty defined in this way will be necessary for emergence, but does not claim

[9] e.g., Holland (1998), Morowitz (2002), as well as a mention in the original (Anderson, 1972, 395).

that it is sufficient. For Wheeler's examples of pressure and temperature pass this test of novelty, yet statistical mechanics shows that they may be defined in terms of fairly simple statistical properties of the relevant collection of molecules. In any case, I do not think that Spencer-Smith's criterion is even suitable as a necessary condition. For it rules out the possibility that systemic properties such as total mass could be an emergent property, which it surely would be if it could be shown that there was a system whose total mass was entirely independent of the masses of its parts.

Anderson suggests a hierarchy of areas of study, in which the elementary entities of science $X$ obey the laws of science $Y$, yet the novelty is such that $X$ is in no sense 'just applied $Y$'. His table starts like this:

| Science $X$ | Science $Y$ |
|---|---|
| solid-state or many-body physics | elementary particle physics |
| chemistry | many-body physics |
| molecular biology | chemistry |
| cell biology | molecular biology |
| ... | ... |

Anderson concedes that this table becomes more speculative as we move down the rows, though others who ally themselves with New Emergentism have confidently continued similar lists with as many as twenty-eight stages (Morowitz, 2002). Nevertheless, at least for these first few rows, Anderson presents clear candidates for the processes by which the systemic properties mentioned in each science become "novel enough" to demand a new set of principles and laws for the science of that domain. In particular he is confident about the first row of the table: the large-scale properties studied in many-body physics (now usually referred to as solid-state, or condensed matter physics) should be considered as having no inferior status to the laws and principles of elementary particle physics.

As we shall see, the New Emergentists have a clear set of examples of physical phenomena which they hold to be genuinely novel (and thus for the associated properties to be judged as emergent). Among these they mention qualitative differences in dynamics, and difficulties in treating the phenomena in terms of microphysical theories. But they do not list necessary and sufficient conditions for novelty, for they do not fall into the trap we identified in §1.1.1; (namely, of claiming a general definition of emergence motivated by the particular examples that one is interested in capturing). I will follow their practice, and in Chapter 2 I will look to their physical examples to guide how to interpret their criteria for novelty.

### 1.2.3 The Central Claims of New Emergentism

Accordingly, we end up with two simple claims that together make up New Emergentism, each of which which will be examined and compared to existing accounts in what follows.

1) *Microphysicalism* (both property and entity varieties) – All entities are composed of microphysical entities, and all their properties are fixed by the microphysical properties, which evolve according to microphysical laws.

2) *Novelty* – Some systemic properties are importantly *novel*: so different to the microphysical that they and the laws that govern them can be recognised as having a metaphysical status in no way inferior to the microphysical.

Although we shall concentrate on the nature of the second claim, we must also be clear as to the apparently harmless first claim of microphysicalism. (Looking ahead, one of the major claims of this chapter (§1.7.2) will be that there is a problematic ambiguity in labelling a property as 'microphysical'.)

When we compare New Emergentism to existing philosophical accounts, we must grant that they have not always set out their analyses of emergence in a similar form to the claims presented here. But by forcing the philosophical discussions into a similar framework, it may become possible to see how they might relate to those of the New Emergentists. So, our task is to see how these two claims can be interpreted as clear metaphysical doctrines, to arrive at a clear interpretation of the New Emergentist position. §1.3 will analyse approaches to microphysicalism, §1.4 & §1.5 will look at ways of making out the criterion of novelty.

## 1.3 Metaphysical Approaches to (Micro)physicalism

Although the New Emergentists accept *micro*physicalism, we will start by discussing metaphysical approaches to the apparently weaker claim of *physicalism*, and their links to their stronger cousin. First, we should distinguish the modern sense from an older one. When the term 'physicalism' first entered the philosophical mainstream, it formed part of the logical positivist programme of "unified science", which aimed to provide a single set of scientific methods of controlled observation and systematic generalization (Neurath et al., 1955). All sciences would proceed with similar methods to those of the physical sciences, which the positivists held to be (roughly) the application of rules of induction to sentences expressing observation reports. Ontological questions such as whether "all is physical" was of little interest to them, if indeed they judged it meaningful at all.

More recently, this emphasis has reversed. Now, physicalists hold that all properties mentioned in scientific study have the same ontological status — they are all physical — but they allow that the methods of the different sciences may vary widely. The much-discussed fragmentation of sciences into distinct domains and the rise of analytic metaphysics with its practice of conceptual analysis has brought

us a modern view of physicalism as having no direct methodological implications. Indeed, those formulations that can be argued to imply some universal "scientific method" are sometimes criticised on precisely those grounds. Methodological variety amongst the sciences is treated as a datum, and accounts are designed to accommodate it. It is these modern versions of physicalism — closer to the old doctrine of *materialism* — that we shall be concerned with.

### 1.3.1 Supervenience Physicalism

Following our discussion of §1.2.1, we can focus on capturing *property* physicalism: the claim that all properties are, or are determined by, physical properties. Though there is still controversy, it is fair to say that an orthodox view has arisen over the last few decades. The most promising approach to capturing the 'determination' inherent in property physicalism is through a relation of *supervenience*.[10]

Take a set of objects $O$, let their *physical* properties form a set $P_A$, and *all* their properties (both physical and ostensibly non-physical) form a set $P_B$. To capture physicalism via supervenience, we demand that the extensions of elements of $P_B$ are fixed by specifying the extensions of the elements of $P_A$. The simplest way of accommodating any such $n$-adic relations amongst objects $o_1, o_2, ...o_n$ is to treat it as a one-place property of the larger object composed of all of these. So we need to close the set of objects $O$ under the operation of taking ordered $n$-tuples.

A little more formally: the properties $P_B$ *supervene* on $P_A$ with respect to a set of objects $O$, on which both are defined, iff any two objects in $O$ that match with respect to all properties in $P_A$, also match with respect to all properties in $P_B$. The central claim of property physicalism can then be expressed very simply: all properties supervene on physical properties.

[10]For example, Jackson (1998) and Loewer (2001) both give supportive surveys of the modern supervenience approach to physicalism.

Although it is widely accepted, supervenience physicalism has its share of difficulties. For one thing, the scope of the whole claim has to be carefully restricted if it is to be plausible. Our set of objects $O$ is conventionally assumed not to encompass all possibilia. (For example, properties of abstract objects can provide controversial cases.) Neither are the property sets unrestricted, for some accounts of dispositional properties cause difficulties. It is also often thought necessary to distinguish several grades of contingency, leading to subtleties such as the distinctions between weak and strong, or local and global supervenience (e.g., Kim, 1993, §§4-5). But since we are dealing with concrete objects and occurrent properties suitable to a broadly scientific investigation it is possible to set aside most of these problems.

Early discussions of supervenience physicalism (such as Hellman and Thompson (1975)) took the property sets $P_A$ and $P_B$ to be provided by the theoretical descriptions of the set of objects $O$. More recent authors have seen a significant advantage in applying supervenience physicalism directly to properties themselves, as opposed to those given by a particular theoretical description.[11] We shall follow the New Emergentists in adopting a fairly naïve realist view of the properties mentioned in physical theories, and so not worry much about the distinction. Unless there are pressing reasons to the contrary, the descriptions given by physical theory will be assumed to be our best authority as to the ontological furniture of the world.

The most obvious difficulty is to determine which properties should be counted as physical, beyond such unhelpful definitions as 'those properties that physicists study'. (The rise of fields such as econophysics suggest that this proposal is not only glib, but also inadequate.) As often noted, if we approach the question

[11] Among the surveys already mentioned, explicit claims of these advantages are made by Howard (2005) and Silberstein and McGeever (1999). Wilson (2005) seems to accept a more theory-dependent view.

naturalistically we face a problem, sometimes labelled the 'Hempelian Dilemma',[12] which can be stated as follows. We cannot use present physics to define what is a physical property, for almost no-one thinks present physics is complete. But if we appeal to the properties mentioned in some future or an ideal physics, then the proposal becomes trivial, for we cannot forsee what this restriction would amount to.[13])

There have been a variety of solutions proposed to the Hempelian Dilemma, the most promising being that the first horn is blunt, and it is possible to formulate physicalism using present physics (Hardin and Rosenberg, 1982; Melnyk, 1997). But perhaps the popular approach to this problem looks to supervenience again, and seeks to define physical properties themselves, by a further dependence on a *micro*physical set.

### 1.3.2 Supervenience Microphysicalism

The thought behind the microphysical approach to demarcating the physical is a simple one. Although any attempt to delimit the scope of future physics might be questioned, it seems plausible that any sciences of the very small qualify as 'physical'. If, like Anderson, we are sure that all physical properties on larger scales supervene upon them, we have a promising approach.

Thus, Latham suggests that 'a property is physical if and only if it logically

[12]According to Stoljar (2005), the problem was dubbed the 'Hempelian Dilemma' in honour of a discussion of a similar problem by Hempel (1969). The difficulty is discussed in more detail by Crane and Mellor (1990).

[13]Nonetheless, there have been several direct proposals for such restrictions, such as Jessica Wilson's 'No Fundamental Mentality' account. Elaborating the suggestions of Papineau (1993), she argues that whatever else future physics may involve, it cannot accommodate properties with a fundamental component of mentality (Wilson, 2005). Even if we set aside as unorthodox those proposals that invoke a mental aspect in the collapse of quantum mechanical wavefunctions, Wilson's proposal appears to be motivated by factors specific to the philosophy of mind, i.e.: a desire for a 'physicalism' which will be useful for setting up clear debates on the mind-body problem.

supervenes on microphysical properties.' (2001, 152). If we take this approach to demarcating physical properties, then physicalism collapses to *microphysicalism*. Microphysicalism is the thesis that the supervenience basis consists of properties mentioned by theories of the 'very small', and thus promises to reduce the problem of demarcating the physical to the use of a ruler. There is obvious appeal in this deal: we obtain a clearly specified supervenience basis by paying the price of defending a rather stronger thesis. And for our purposes, aiming to capture the physicalist doctrines of the New Emergentists, it looks especially attractive. For as we saw in §1.2.1, the New Emergentists already accept microphysicalism: there appears to be everything to gain, and the only cost, one we have to pay anyway!

Thus, to capture the sort of property microphysicalism that the New Emergentists accept, we can use the same supervenience framework as for physicalism. The objects $O$ are taken as comprising those objects subject to scientific investigation (or at least those 'of which we have detailed knowledge', to reproduce Anderson's hedging as quoted on page 16)[14] and $P_A$ as some level of uncontroversially microphysical properties governed by microphysical laws.[15]

This approach looks promising, but I shall make just a few points of caution. First, paradoxical as it might sound, not all properties mentioned by theories of the very small are properties only of very small systems. Although theories such as elementary particle physics use terms referring to properties such as the spin, colour, mass and charge of very small elements, the theories may also be used to describe systems made up from many of the elements together, so can include the systemic properties of these far larger composites. And it is part of the New

[14]We may need to cover more under $O$ than actual objects, for there may not be two objects that match in respect of all properties in $P_A$, in which case the supervenience requirement will be trivially satisfied. This is likely to involve some modal commitment, but notwithstanding the concerns expressed in §1.3.1 I shall assume that it is unproblematic.

[15]Again, there is no requirement that this be associated with any ultimate ontological micro-level. Anderson even accepts that the level $P_A$ could be properties covered by the Standard Model of contemporary particle physics.

Emergentist thesis that these systemic properties might be very different to the spin, colour, mass and charge of the elementary particles. This point is clear if we consider the set of objects $O$; it encompasses objects of *all* sizes. We have called the properties belonging to these objects 'microphysical', not because they are all properties of small systems — they are not — but because they have been assigned by a theory we call "microphysical", such as the Standard Model.

Second, it should perhaps give us pause that even if we did restrict attention to properties of the very small, we would fail to exclude such non-physicalist theories as panpsychism and Cartesian dualism from the realm of the physical. To qualify, such a theory would merely have to associate a small spatial extent with their non-physical properties. As a matter of fact, existing dualist and panpsychist theories tend to postulate non-spatiotemporal properties, and so are correctly categorised by the criterion, but this is by no means a necessary requirement.

These reservations notwithstanding, our characterisation of microphysicalism by way of supervenience should ensure that we shall be concerned with properties that really do 'emerge' from the microphysical, as opposed to examples such as dualist properties, which could be said to 'float' entirely free of them. We can now move to the main hunt: for the precise sense of this 'novelty' which is meant to separate out the emergent properties from amongst the systemic.

## 1.4 Novelty — A Merely Practical Matter?

There is a clear candidate for a distinction of novelty in Anderson's article, which he calls 'anti-constructionism'. But once it is spelled out, those metaphysicians who have discussed the New Emergentist views tend to express doubts. For it begins to look as though Anderson has failed to notice a clear distinction in the sense of the term "emergent". Namely, the distinction between deductions or

predictions being possible in practice, versus it being possible in principle.

Almost all conceptions of emergence characterise novelty by appealing to some conception of *predictability* or *derivability*, questioning whether it is possible to obtain one set of properties from another by some mechanical process of logical derivation.[16] Kim's 1995 definition is typical from one concerned with the philosophy of mind: 'a property of a complex system is said to be 'emergent' just in case, although it arises out of the properties and relations characterizing simpler constituents, it is neither predictable from, nor reducible to, these lower-level characteristics.' The thought is that if it is possible to derive systemic properties from those of their components, these properties cannot be truly novel — for they were "there all along" in the parts, and it was merely a matter of careful analysis to make this fact manifest. This leads naturally to a distinction between this 'prediction' or 'reduction' being one that could be carried out in practice, and one that is possible only in principle.

This distinction is often the very first move made in discussions of emergentist doctrines. The survey article of Silberstein and McGeever (1999) uses a widespread terminology when they distinguish 'ontological emergence' (a failure in principle), and 'epistemological emergence' (a failure in practice). They are also typical in dismissing the latter as meriting little interest, at least for those who are searching for a sense of emergence which carries metaphysical significance.[17] In this section

[16]There are clearly differences between a property being 'predictible' and being 'derivable', each of which are appealed terms are used in discussion, along with 'reducible', and 'determinable'. In particular, 'predictable' has a temporal connotation, which suggests that we may be want to draw a distinction between the current systemic properties, and the ones to which they may evolve in the future. Distinctions between 'synchronic' and 'asynchronic' emergence have been discussed (e.g., Stephan, 2002), but we can ignore these distinctions, since we shall be concerned with the systemic properties of our system at a particular time.

[17]O'Connor and Wong (2005) use the same terminology as Silberstein and McGeever. Chalmers' 'weak' and 'strong' emergence appears to express the same distinction (Chalmers, 1996, 39), and Bedau (2002) uses the same terms, with the addition of further categories. Though they are accused of confusion, the New Emergentists are also aware of the difference, see for example the distinctions discussed in Mayr (1982, 59-64).

we shall look more closely at the most popular way of reading the New Emergentist claims, which appears to accuse them of confusion, between the 'epistemological' and 'ontological' senses. While I concede that it is one way to interpret some of their claims, I contend that it makes their claims so mysterious, and attributes to them such simple-mindedness, that it is uncharitable to accuse them of such errors.

### 1.4.1 Anti-constructivism

Anderson accepts that all matter is composed of microphysical parts, that all facts supervene on microphysical facts, and that all are governed by microphysical laws. We argued back in §1.2.1 that his reference to this as 'reductionism' is at odds with the metaphysicians' usage: they would call it 'microphysicalism'. This should be kept in mind when we read Anderson making out the following claim, usually understood as setting out his Emergentist stall:

> ... the reductionist hypothesis does not by any means imply a constructionist one: The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe. (Anderson, 1972, 393)

But despite the prominence with which this statement of anti-constructivism has been quoted in the discussions of Anderson's claims, his immediate attack on 'constructionism' is then based on practical difficulties:

> The constructionist hypothesis breaks down when confronted with the twin difficulties of scale and complexity. (Anderson, 1972, 393)

As is the attack from Laughlin & Pines:

> However, it [the non-relativistic Schrödinger equation] cannot be solved accurately when the number of particles exceeds about 10. No computer existing or that will ever exist, can break this barrier because it is a catastrophe of dimension ... the triumph of the reductionism of

> the Greeks is a pyrrhic victory: We have succeeded in reducing all of ordinary physical behaviour to a simple, correct Theory of Everything only to discover that it has revealed exactly nothing about many things of great importance. (Laughlin and Pines, 2000, 28)

The thesis of anti-constructionism then looks an uncontroversial one. It is commonplace that the laws of particle physics are practically useless once we attempt to apply them to systems beyond a certain size and complexity. No-one in their right minds would attempt to use the Standard Model to determine the behaviour of macroscopic objects.[18] And reductionists such as Weinberg accept this — all they maintain that it is possible *in principle*, a claim that the New Emergentists do not challenge.

### 1.4.2 The Orthodox Gloss

The usual reading of the New Emergentist claims is that we should take the claims of anti-constructionism as the New Emergentist's criterion of 'novelty', i.e. the criterion which distinguishes the 'trivial' systemic properties from the emergent ones. That is, systemic properties are novel, if and only if it is *practically* impossible to derive them from the microphysical properties mentioned in microphysical supervenience.

If we follow these suggestions we can see why the reaction to the New Emergentist claims from philosophers has so often been one of faintly puzzled dismissal. For Anderson makes claims such as this:

> The essential idea is that in the so called $N \to \infty$ limit of large systems (on our own, macroscopic scale) it is not only convenient but essential to realize that matter will undergo mathematically sharp,

[18] There are interesting exceptions. Superconductivity and superfluidity are not exactly everyday phenomena, but they are on macroscopic scales, and require a treatment using principles very similar to those in quantum field theory. In fact, many-body quantum mechanics has wide-ranging and suggestive parallels with techniques exploited in the Standard Model, most famously in the Higgs-Anderson mechanism — a link we shall not pursue here.

> singular "phase transitions" to states in which the microscopic symmetries, and even the microscopic equations of motion, are in a sense violated. (Anderson, 1972, 395)

It looks as though Anderson moves illegitimately from his premise that it is *practically* essential to understand higher-level properties associated with phase transitions, to an ontological conclusion about the microphysical laws (=equations of motion) that govern the properties of such systems. But we have also seen him concede that all systemic properties, even at our own macroscopic scale, continue to be determined by microphysical properties. And Anderson has given us no reason to think that microphysical laws will be violated — only that it is practically impossible to use them.

The New Emergentists are thus accused of illegitimately citing evidence for epistemological emergence (a practical difficulty in deriving the properties of a large system) as support for claims of ontological emergence (there exist emergent properties that are "novel" in a metaphysically important sense).

Leaving aside specific quotes, it is hard to see in general how any criteria for novelty as a "practical" matter could support the full-blooded metaphysical claims made by the New Emergentists. If criteria for novelty considerations are sensitive to the availability of physical theories, our interpretations of them, and to our levels of intellect and ingenuity, how can they could support conclusions about the status of physical properties and laws? Of course, the interpreters taking this tack make these very objections: this is the very reason that they they judge 'epistemological' emergence to be uninteresting.

Several philosophers have concluded that New Emergentism must be interpreted along these lines, and that it does not stand up to detailed scrutiny as a metaphysical thesis. This is the conclusion of O'Connor and Wong (2005), who explicitly accuse Laughlin and Pines confusing practical and in-principle derivations, and of Don Howard who finds no grounds for their claims of emergence

(Howard, 2005, 17). Similar scepticism is expressed by Chalmers (1996, 129-30), and it is also implicit in Cat (1998), who is slightly more charitable in his interpretation, but ends by judging that Anderson's commitments are unclear between an ontological or epistemological viewpoint.

These dismissive reactions from philosophers are certainly understandable given the difficulties of relating the New Emergentist positions to existing frameworks for emergence. But rather than rushing to condemn, it is worth looking at Anderson's thesis more carefully. In the next sections, we will look at some existing metaphysical approaches to the novelty of emergent properties, and ask whether the New Emergentist theses can be related to them.

## 1.5 Novelty — Three Metaphysical Approaches

Let us assume that the dismissive reaction of the last section should be resisted, and look within metaphysics for fully objective conceptions of emergence. Again, we seek a clean distinction amongst the systemic properties, to separate the 'trivial' and 'non-trivial': those that are straightforwardly obtainable from the microphysical properties belonging to the parts of a system, and those that are importantly *novel*. In §1.5.1 – §1.5.3, we examine three proposals; novelty is shown by: 1) a failure of inter-theoretic reduction; 2) an impossibility of deducing the systemic properties from the properties of the parts; or 3) a failure of mereological supervenience.

These three approaches have rather different histories: the first and last have been the focus of much recent philosophical discussion, while the second is part of a rather older tradition, known as "British Emergence". The main claim I shall argue in this section is that the approaches appeal to three entirely separate distinctions between the properties of parts and wholes; thus there are three entirely different

sets of criteria for emergence.

### 1.5.1 Failure of Inter-theoretic Reduction

A popular conception of emergence is made out in terms of the possibility of the *reduction* of one theory to another (e.g., Hellman and Thompson (1975), Howard (2005), Butterfield and Isham (1999), Batterman (2002a)). Inter-theoretic reduction is usually discussed using the syntactic conception of theories, as opposed to the semantic conception. (I shall assume that nothing hangs on the choice, except that it is easier to make a clear definition of reducibility in syntactic terms.) In the syntactic conception the postulates of a theory are treated as sets of sentences, and in these terms, when we succeed in showing that one theory *reduces* to another, we show that is possible to deduce the postulates of one theory from those of another, by adding judiciously chosen definitions of its non-logical vocabulary. The idea is that if a reduction is available, the content of the first theory has been shown to be completely accommodated within the second, though this fact might have been obscured by differences in terminology, or a lack of appropriate definitions.

Like supervenience, reduction was originally studied to capture the logical positivist thesis of physicalism as a thesis of the unity of science. As we mentioned in §1.3, such conditions have been found too strong to capture the modern doctrine of physicalism. But this disadvantage can be turned around, and exploited in our search for criteria for *novelty* amongst the systemic properties. A natural suggestion is that if a theory of the behaviour of these systemic properties can be shown to *fail* to reduce to microphysical theory, this constitutes an important degree of novelty in those properties. Thus, a simultaneous success of microphysical supervenience, with a failure of reduction to a microphysical theory, can be put forward as a definition of emergence.

So, the suggested analysis goes as follows: Take our common domain of ob-

jects $O$, with properties $P_A$ described by a microphysical theory theory $T_A$, and properties $P_B$ described by a higher-level theory $T_B$. If the properties $P_B$ described in theory $T_B$ *supervene* on the properties $P_A$ addressed in theory $T_A$, yet it is impossible to *reduce* theory $T_B$ to $T_A$, we have emergence. We thus require precise conditions for theoretical reduction to add to our precise supervenience conditions for microphysicalism. We define emergent properties as "squeezing" into the logical space where the second holds, but the first does not.[19]

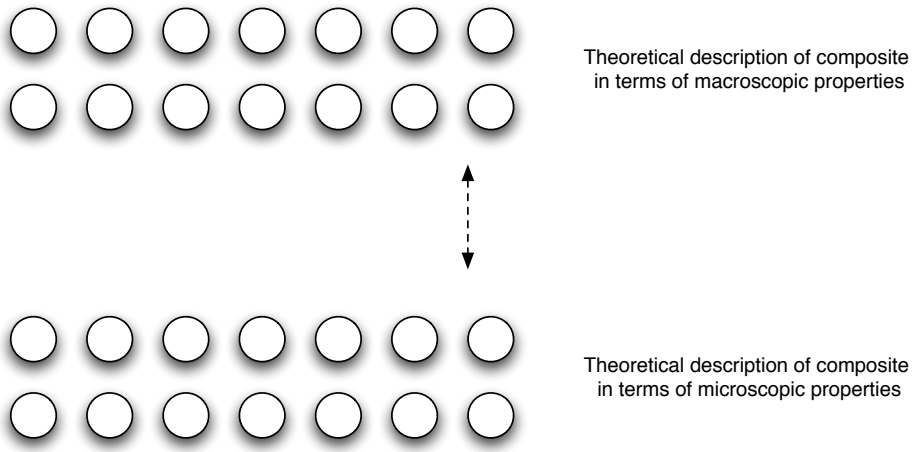



Figure 1.1: The relationship relevant to emergence understood as a failure of Inter-theoretic Reduction. It is between two descriptions of the same system in microphysical and macrophysical terms.

### Inter-theoretic Reduction

In keeping with their pedigree, the classic analyses of theoretical reduction are based around a logical operation, known as 'definitional extension' (e.g., Hempel and Oppenheim (1948) and Nagel (1961, 366-74)). Definitional extension consists in adding to the reduc*ing* theory $T_A$ some further definitions, which aim to capture

[19]Davidson (1980) seems to want his 'Anomalous Monism' or 'Non-reductive Physicalism' to share this space with emergence. Crane (2001) and Robinson (2001) are among those who explore the question of whether the two can be distinguished.

the terms used in the reduc*ed* $T_B$. Call this augmented theory $T_A^*$. If we then find that it is possible to deduce all the postulates of $T_B$ from within $T_A^*$ then intertheoretic reduction has been achieved.[20] Despite its age, and clear ancestry in logical positivist thinking, definitional extension has remained a popular way of viewing inter-theoretic reduction. For example, Don Howard appeals to it in his examination of emergence:

> Intertheoretic reduction is a logical relationship between theories. In the classic formulation owing to Ernest Nagel, theory $T_B$, assumed correctly to describe or explain phenomena at level $B$, reduces to theory $T_A$, assumed correctly to describe or explain phenomena at level $A$, if and only if the primitive terms in the vocabulary of $T_B$ are definable via the primitive terms of $T_A$ and the postulates of $T_B$ are deductive. (Howard, 2005, 3-4)

Nagel's 'classic formulation' is found in *The Structure of Science*, published in 1961, but we should bear in mind that a bare-bones requirement of definitional extension was only the formal core of a detailed set of proposals, which themselves inspired an extensive literature on further refinements. Even in his 1961 discussion, Nagel added further conditions of 'connectability' and 'derivability', (Nagel, 1961, 353-4) which place some conditions on the relations between the terms defined in both theories, and discussed the nature of the definitions that can be introduced to $T_A$ to effect the reduction. (See Belot (2005) for further discussion of Nagel's own treatment of the problems). However, it seems fair to say that definitional extension is the centrepiece of his position, and has remained central to the vast literature it has inspired.

The clearest examples of a reduction achieved by definitional extension come from mathematics, and the jewel is the success of Frege, Russell, Whitehead, Zermelo et al, in reducing classical pure mathematics to set theory. Astonishingly,

[20] To prevent confusion arising from accidental homonymies, we assume here that $T_A$ and $T_B$ have disjoint vocabularies.

it seems to be possible to construct definitions of all the terms of classical pure mathematics, and deduce its theorems, using just the resources of set theory, cast in first order predicate calculus. This stunning result is probably the principal reason that definitional extension became so popular as an analysis of inter-theoretic reduction.

The most popular objections to Nagel's proposal (or rather, to this caricature of it), has been that once we move to reductions in areas beyond pure mathematics, full definitional extension is far too rigid a requirement. Feyerabend (1962) mounted a famous attack on the example that Nagel himself used: the reduction of thermodynamics to equilibrium statistical mechanics. Feyerabend points out that the concept of temperature is inherently statistical in the latter theory, non-statistical in the former, and casts doubt on the possibility of constructing definitional extensions that could accommodate this fact.

Such objections, and more general concerns that strict requirements of definition and logical deducibility are unsuitable to capture relationships between pairs of scientific theories, have led to more sophisticated variations on Nagel's approach. Relationships such as approximation, limiting relations and analogy have been suggested (the discussions of Nickles (1973) and Schaffner (1976) are particularly influential), but it is difficult to make them precise, and the success of these more nuanced analyses remains questionable. It is telling that the "reduction" of thermodynamics to statistical mechanics, has moved in thirty years from being the archetypal example of a successful inter-theoretic reduction (Nagel, 1961, Ch.11), to providing the archetypal example of the problems inherent in capturing the concept (Sklar, 1993, Ch.4).

**Emergence**

Fortunately for our purposes, the question of how we should capture a notion of emergence can be torn free from the tangled issues of whether definitional extension does justice to the concept of inter-theoretic reduction. For even if definitional extension is unsuitable as an analysis of reduction, we could still require its failure as a necessary condition for property emergence. If it is to play this role there must merely exist sufficient logical "space" between it and supervenience, that is, there must be clear cases in which supervenience holds but a definitional extension is not possible.[21]

Unfortunately, such a retrenchment does not help. The widespread criticism of definitional extension as an unreasonably strong requirement for reduction might be thought to guarantee such clear logical "space"; but disappointment awaits. Butterfield and Isham (1999) call attention to the fact that inter-theoretic reduction may be surprisingly weak, and its strength compared to supervenience may be illusory in many cases. This is shown by a result known as Beth's theorem, which proves that the two are actually equivalent under a wide range of circumstances.[22] We can repair this by imposing restrictions on the resources we allowed to construct the definitional extensions. Howard's formulation mentions 'definability' and 'deducibility', which implicitly calls for the resources of some logic. Tradition-

[21] At first sight the distinction appears as clear as that between infinite and finite definitions. For suppose that supervenience holds. Then for each property in the set $P_A$, one can construct a definition of it in $T_B$'s terms, by taking the disjunction of the complete descriptions in terms of properties $P_B$ of all the objects in O that have that $P_A$ property. Thus there is a definitional extension of $T_B$ that covers the properties in $T_A$. If there are infinitely many ways that an object can have a $P_A$ property in virtue of possessing $P_B$ properties, the disjunction will be infinite, and this is not usually admitted as an acceptable definitional extension. But whether those disjunctions are finite or infinite, supervenience will ensure that objects matching in their complete $P_B$ descriptions match as regards that $P_A$ property. For this reason, supervenience is sometimes called 'infinistic reduction'.

[22] They credit the first application of Beth's theorem to issues of supervenience and reduction to Hellman and Thompson (1975), who also pointed out that it holds for first-order languages only. Accordingly Hellman and Thompson suggest that it can be evaded by holding that realistic scientific theories are typically formulated in higher-order languages.

ally, the reduction has followed the example of that celebrated result mentioned earlier: most of pure mathematics successfully reduced to set theory, formulated using just first order logic. But when we consider examples of reduction in physics it is natural to add at least some extra mathematical operations to the basic logical ones. Even very simple theories demand mathematical operations such as differentiation and integration, which appear rather 'high up' in the set-theoretic hierarchy.

However, once we start to allow further resources to enter our toolkit for building definitional extensions, many controversial examples of reduction are swept away by the sheer power of modern mathematics. For example, equilibrium classical statistical mechanics of isolated systems becomes a definitional extension of Newtonian micro-mechanics once the calculus of many variables and measure theory are employed (see Butterfield and Isham (1999, 119-120) for more details).

Such power should make us suspicious, and indeed we find that the questions of which resources should be admitted are often the same questions at issue in controversies of inter-theoretic reduction. Are the resources of measure theory alien to Newtonian mechanics? Certainly it was alien to Newton and his contemporaries (or even to Boltzmann, and his); but at what point do the more powerful mathematical developments stop being "mere" aids to construct definitional extensions, and start defining properties that were not present in the original theory?

For a more contemporary example, consider the recent controversial claims by (Batterman, 2002a), that there are physical phenomena which can be explained within $T_B$ and not $T_A$, though a successful reduction of $T_B$ to $T_A$ can be carried out. The subsequent argument (Redhead (2004), Belot (2005), Batterman (2005)) has centred on the mathematical resources one is allowed to appeal to, both in reduction and explanation. Distinguishing a reasonable from an unreasonable set of definitional resources presupposes answers to the very questions that motivated

us to look to inter-theoretic reduction as an analysis of emergence: the extent to which the properties of the upper-level theory are novel compared to the old.

We have seen that a criterion of inter-theoretic reduction may be made so weak as to be equivalent to supervenience, or so strong as to admit of no clear examples in the physical sciences, simply by altering the resources allowed to construct the extensions. It is hard to disagree with Butterfield and Isham, who conclude that the criteria of inter-theoretic reduction made out in terms of definitional extension are formulated at too high a level of generality to provide a useful analysis of emergence. Let us turn to our second set of criteria, which has an equally distinguished history to draw on.

### 1.5.2 British Emergentism

The British Emergentists constitute a movement active in the early decades of the twentieth century. They have attracted renewed attention in the last decade, inspired largely by a careful philosophical history by McLaughlin (1992), which presented C. D. Broad's views as particularly interesting and worthy of detailed study.[23]

Broad was concerned with the relationship between the properties exhibited by composite systems and the properties exhibited by their parts when separated

[23]Other writers also deserve mention, though I shall not discuss their views in detail. I have already mentioned that Lewes (1877) is thought to have first used the term 'emergence' in its modern sense. Lloyd Morgan (1923) and Alexander (1920) are each at least of equal influence as was Broad within the British Emergentist movement, particularly to subsequent developments in biology (Mayr, 1982, 1997). One reason that McLaughlin (and I) have chosen to focus on Broad is that Lloyd Morgan and Alexander's writings are rather opaque to interpretation, and appear to reject any scientific investigation of emergent properties, making them less relevant to the New Emergentist thesis. Lloyd Morgan appeals to extra forces that are 'in some sense extra-natural' (Lloyd Morgan, 1923, 8), admitting that this may involve an appeal to God as supplementary to scientific explanation. And Alexander rejects out of hand the idea that there could be any comprehensible physical process behind emergence: 'The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the natural piety of the investigator. It admits no explanation.' (Alexander, 1920, 47).

from those composites.[24] His central definition of emergence is expressed in the following passage:

> Put in abstract terms the emergent theory asserts that there are certain wholes, composed (say) of constituents, $A$, $B$ and $C$ in a relation $R$ to each other; that all wholes composed of constituents of the same kind as $A$, $B$ and $C$ in relations of the same kind as $R$ have certain characteristic properties; that $A$, $B$ and $C$ are capable of occurring in other kinds of complex where the relation is not of the same kind as $R$; and that the characteristic properties of the whole $R(A, B, C)$ cannot, even in theory, be deduced from the most complete knowledge of the properties of $A$, $B$ and $C$ in isolation or in other wholes which are not of the form $R(A, B, C)$ (Broad, 1925, 61)

In Appendix A, I separate two quite different ways in which this deduction of systemic properties might be impossible. Broad's first suggestion, that there might be '*non-behaviourably analysable* properties', is not totally clear, but it looks as though they may fail to supervene on the microphysical, and so are less relevant to the present debate. More interesting for our purposes is Broad's second suggestion which he introduces via *specific laws of composition*, a term which can be explained as follows. Broad held that to derive the properties of a composite system from the properties of its components, the following is required:

> (a) We need to know how the parts would behave separately. And (b) we need to know the law or laws according to which the behaviour of the separate parts is compounded when they are acting together in any proportion and arrangement. (Broad, 1925, 61)

Broad then distinguishes two possibilities for the type of the laws covered under clause (b). These laws can be *specific* to a situation or *general*. The composition of Newtonian forces by vector addition was taken as the paradigmatic example of

[24] In discussing Broad's views, we will be forced to depart somewhat from our policy of focussing on emergent properties to the exclusion of emergent laws. Also, there are many points of interest in his account, discussion of which would take us far from our central parts. A discussion of a few of the more relevant of these, I have relegated to Appendix A.

a general law of composition.[25] When we want to know how a large number of forces combine to produce a systemic effect, the rule is the same whether there are two forces or a thousand: we combine them by the simple rules of vector addition. Thus, under good conditions, it should be possible for a sufficiently skilful calculator to derive the systemic properties of such a composite from knowledge of properties of its components when alone, and in composites other than the one under study.

In contrast, when the laws of combination are *specific*, this derivation might not be possible without study of the system itself. Broad's favoured examples were drawn from chemical phenomena: for at the time he wrote, it seemed that the properties of chemical compounds could not be derived from the properties of their constituent parts.

> ... if we want to know the chemical (and many of the physical) properties of a chemical compound, such as silver-chloride, it is absolutely necessary to study samples of that particular compound. [...] The essential point is that it would also be useless to study chemical compounds in general and to compare their properties with those of their elements in the hope of discovering a general law of composition by which the properties of any chemical compound could be foretold when the properties of its separate elements were known. So far as we know, there is no general law of this kind. It is useless even to study the properties of other compounds of silver and of other compounds of chlorine in the hope of discovering one general law by which the properties of silver-compounds could be predicted from those of elementary silver and another general law by which the properties of chlorine-compounds could be predicted from those of elementary chlorine. (Broad, 1925)

Broad's suggested explanation of this failure was striking — it could be that the law that determined how the properties of silver and chlorine combine to give the properties of silver chloride, was a law *specific* to that situation. Broad's main

[25]For simplicity, we can narrow the example to Boscovitchian point particles, interacting via well-behaved potentials, and ignore esoteric possibilities such as "space-invader" cases. The non-deterministic aspects of Newtonian mechanics, though fascinating, are not our concern here.

examples of specific laws were ones specific to a particular configuration of atoms, which he called 'configurational laws'.

To emphasise his point, Broad introduced a 'Mathematical Archangel', blessed with both infinite calculational ability, and 'gifted with the further power of perceiving the microscopic structure of atoms as easily as we can perceive hay-stacks.' This being 'could no more predict the behaviour of silver or of chlorine or the properties of silver-chloride without having observed samples of those substances than we can at present' (Broad, 1925, 71). When silver and chlorine formed the particular configuration which constitutes silver chloride, an entirely new law of composition "kicked in" to produce the chemical properties. Since one would need to know this specific configuration law to deduce the systemic properties from those of the components, and since this law (by assumption) shows no effect except in its influence on silver-chloride itself, even the Archangel will fail to deduce its systemic properties. Broad labelled both the underivable chemical properties, and the laws that govern them, 'emergent'.

Broad's criteria for emergence thus makes precise those central emergentist intuitions of novelty and unexpectedness. He recognised that the properties of a composite system are always new *in some sense*: that is just what we mean by recognising systemic properties. As we emphasised in §1.2.3, the point is to make a distinction between a trivial and non-trivial examples. Putting the same point in a different way, if we follow Wheeler with his slogan that the properties of the whole are greater than the 'sum of the parts', we must distinguish a trivial from non-trivial "summing" operation. Here we can use common terminology and refer to a whole system that would be created if we brought together two isolated systems $S_1$ and $S_2$ into a certain configuration as the *composite* of $S_1$ and $S_2$. For Broad, the general laws of composition express trivial "summing" operations, but when specific laws come into effect the operation is non-trivial, and the resulting

composite system can then be said to possess emergent properties.

To compare British Emergence to the other conceptions considered here, it is important to recognise three closely related points. First, there is a clear distinction between specific laws of composition on one hand, and emergent properties and laws on the other. A law of composition is a rule for combining properties of components to give properties of a composite. It is specific if it applies only to a single configuration, general if it applies to many. When specific laws apply, the properties of that configuration may be called emergent, as may the laws that govern them. Thus emergent properties and laws may be derived from the general *and* specific laws of composition together, but not from the general laws alone.

A second point is the status of a specific law such as a configurational law. Certainly it is a physical law, but we might wonder whether to call it *micro*physical. On the one hand, it is relevant only to the systemic properties of large systems, so we might not consider it microphysical. On the other, the law can be expressed using only microphysical terms. That is: we need not refer to higher-level terms such as 'silver chloride' when we specify the law: we can talk about such-and-such arrangement of silver atoms and chlorine atoms and their effects,[26] likewise in microphysical terms. So if we added this law to the existing set of microphysical ones, the augmented theory seems to remain microphysical; yet it could account for the properties studied in chemistry. Thus there seems an ambiguity in the term "microphysical", both when applied to properties and to laws: an issue we shall return to later.

This leads to a third point: the existence of a specific law of composition, and thus the idea of British emergentism, can be made either (1) entirely independent of emergence understood as a failure of inter-theoretic reduction, or (2) equivalent to it, depending on the properties mentioned by theories $T_A$ and $T_B$.

[26]Or, when the microphysical theory is taken to be the Standard model, we would speak of arrangements of protons and electrons, or quarks and electrons.

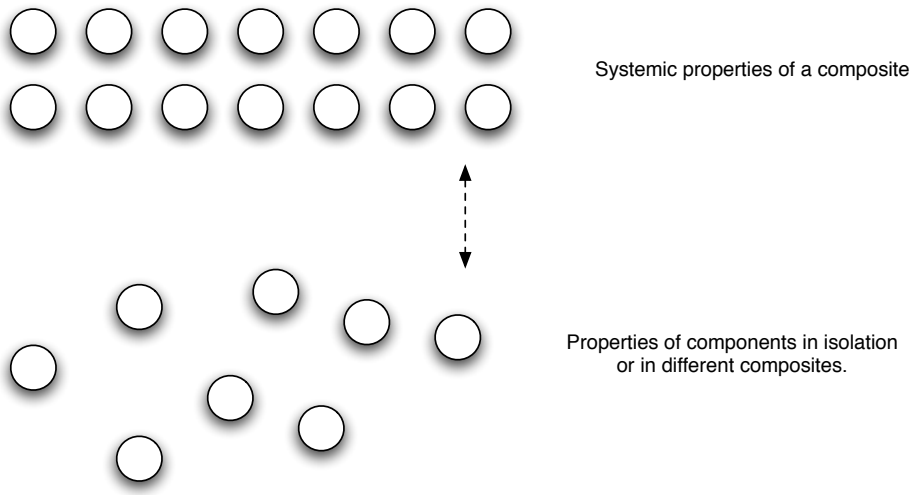



Figure 1.2: A schematic representation of the relationship relevant to Broadian, or British emergence. The novelty is to come between sets of properties belonging to two different systems: properties of a composite system, and those of its component parts separated from that composite.

To see this let us start by considering how we might express British Emergentism in terms of a failure of inter-theoretic reduction. We could dictate that $T_A$ be a theory mentioning only properties obtainable under general laws of composition, and $T_B$ a theory differing from $T_A$ only in that it also mentions properties obtainable with any specific laws. If a configurational law of composition is in effect, its effects would be taken into account by $T_B$, but not $T_A$, so reduction would fail.

But these two conceptions of emergence only become equivalent when we engineer them as such. We tend to assume that it must be a microphysical theory $T_A$ that would neglect properties obtainable with the configurational law, for we tend to associate a theory that describes systems in microphysical terms with a restriction to the analysis of small systems. But we could just as easily dictate that it was the macrophysical theory $T_B$ that omitted them. Further, if we did find that neglect of a specific law of composition was the reason for the failure of reduction, either theory could easily be repaired by adding it to the existing laws.

Again, $T_A$ would handle the addition of the specific law by referring to microphysical properties and relations, $T_B$ would use the macrophysical, but this difference does not affect whether each theory could accommodate it.

An inter-theoretic reduction could also fail for reasons entirely independent of British Emergentism: due to the set of resources being used to construct the definitional extensions being inadequate for the task, or any of the general difficulties with inter-theoretic reduction outlined in the last section. So while failure of inter-theoretic reduction might be taken as *evidence* for the existence of British emergence, the diagnosis is just one amongst many possibilities. Reduction can fail even when only general laws are in effect.

### 1.5.3 Failure of Mereological Supervenience

Finally, we turn to the third proposal — that a novelty of systemic properties, and thus emergence, is shown by a failure of mereological supervenience. While this proposal shares some of the "logical technology" of supervenience, it expresses a very different condition to the doctrines of physicalism and microphysicalism discussed in §1.3.1. We look at the relation between parts and wholes by looking whether the properties of wholes supervene on those of their parts. This is quite different to our previous question: whether a set of properties assigned by a macrophysical description of a whole supervene on those assigned by a microphysical description, also of the whole.

Don Howard, who judges this conception of emergence the only clear one recognised by modern physics, defines it thus:

> Supervenience is an ontic relationship between structures. A structure, $S_x$, is a set of entities, $E_x$, together with their properties and relations, $PR_x$. A structure, $S_B$, characteristic of one level, $B$, supervenes on a structure, $S_A$, characteristic of another level, $A$, if and only if the entities of $S_B$ are composed out of the entities of $S_A$ and the properties

> and relations, $PR_B$, of $S_B$ are wholly determined by the properties and relations, $PR_A$, of $S_A$. (Howard, 2005, 4)

To define microphysical supervenience, we examined a supervenience relation between *all* the properties available in a description given by microphysics and all the other properties. To define mereological supervenience, we take the supervenience relation as between those properties that can be assigned to the composite system taken as a whole, and those that can be assigned to the parts of the composite system (so we can think of structure $S_A$ as the parts, and $S_B$ as the whole). It is particularly important to realise that these two sets of properties could be two sets provided by different theories, or the two sets provided by the same theory in each case — the difference is simply in that the first set can be meaningfully ascribed to the system's parts, the second to the system as a whole.

Howard then argues that there are uncontroversial situations in physics for which mereological supervenience does fail: when the parts make up an entangled quantum system (Howard, 2005, 12). Consider two electrons with their pure spin states representable by vectors in Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. There exist entangled states of the compound system of two electrons, which cannot be written as the simple tensor product of any two pure states from $\mathcal{H}_1$ and $\mathcal{H}_2$. For example, ignoring the spatial degrees of freedom, consider the singlet state:

$$\frac{1}{\sqrt{2}} |\uparrow_1\rangle |\downarrow_2\rangle - \frac{1}{\sqrt{2}} |\downarrow_1\rangle |\uparrow_2\rangle \tag{1.1}$$

(where $|\uparrow_1\rangle$ represents the pure spin-up state for the $z$-direction, say, and $|\downarrow_1\rangle$ the pure spin-down state of electron 1, similarly for electron 2). This state cannot be written as the tensor product of any pure state in $\mathcal{H}_1$, with any pure state in $\mathcal{H}_2$. And if we try to describe the electrons with *reduced* density matrices (the orthodox way to express such separated states) we find that the description fails to capture

(measurable) correlations between the states of electron 1 and 2.[27] The state of the whole (structure $S_B$ in Howard's terminology) is not wholly determined by the states of the parts (structure $S_A$).

In the case of these entangled quantum systems, mereological supervenience fails, while microphysical supervenience holds. Mereological supervenience fails because it restricts its supervenience base to only those properties that we can fully assign to a part of a system (in our example, density matrices of single electron systems). The properties assignable to the components of a system are not sufficiently rich to determine all the properties of the composite. Microphysical supervenience allows as its base the whole range of properties and relations available in the description used by a microphysical theory, (in our example, elementary quantum mechanics) and amongst these is the singlet state of Equation 1.1, so microphysicalism holds.

Opinions differ on how we should interpret the metaphysical significance of these states. It is especially controversial whether it is possible to separate the significance of the 'fused' states from other issues of the interpretation of quantum states.[28] Nevertheless, it is generally accepted that they bring out fundamentally non-classical features of composite quantum systems.

In any case, we have here a third possible analysis of emergence. For a systemic property to be emergent in this sense, we require that microphysical supervenience to hold, but mereological supervenience to fail.[29]

[27]There is a large literature on the failure of mereological supervenience in entangled quantum systems. Good starting points would be any of the sources in note 28 and those referred to in Howard's own discussion.

[28]Taking a pair of positions from each side of the debate: Maudlin (1998) and Humphreys (1997) emphasise its significance for a form of holism, regardless of a wider position on the interpretation of quantum mechanics. In contrast, Healey (1991) and Dickson (1998) hold that its significance should be interpretation-relative.

[29]Many authors consider mereological supervenience to be the only context in physical science for which a clear and physically realised case of 'emergence' can be demonstrated. Of those already mentioned, Silberstein and McGeever (1999) and Howard (2005) are strong advocates of this position.

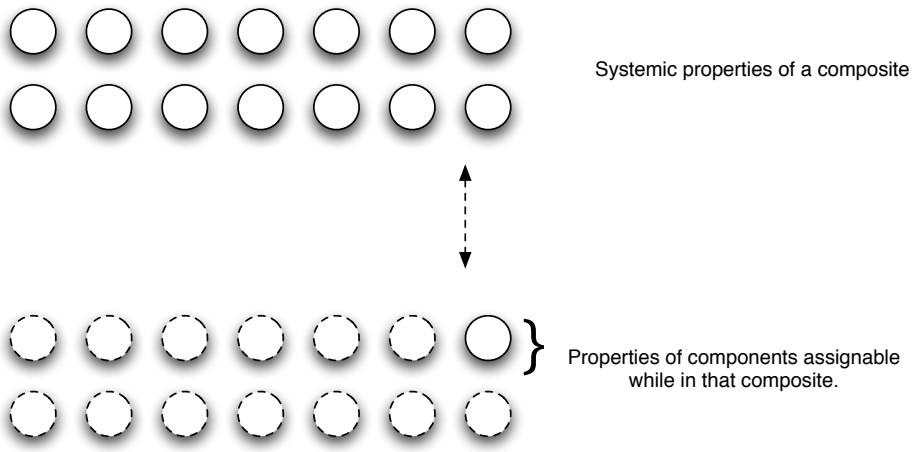



Figure 1.3: A schematic representation of the relationship relevant to emergence understood as a failure of mereological supervenience

## 1.6 Evaluation of the Three Metaphysical Approaches

The most striking feature of the three conceptions examined is that they judge emergence as occurring under quite different circumstances. While many differences have been recognised amongst the senses of 'emergence' recognised in the literature, I am not sure that they bring out the most fundamental distinctions. In §1.6.1 I shall illustrate the diversity of the criteria, by presenting a variety of situations for which they disagree as to whether emergent properties appear. In §1.6.2, I shall present what I take to be the most important difference between them: the interpretation of what counts as a 'part' of a composite system. Finally, in §1.6.3, I shall argue that none of them can capture the position held by the New Emergentists.

### 1.6.1 Their Diversity

It turns out that the three sets of criteria of emergence discussed in §1.5.1 – §1.5.3 hold in very different situations. In §1.5.2 we saw that while a case of British Emergence might be seen as a reason for the failure of inter-theoretic reduction, the two notions can also come apart, simply because we are free to include or omit specific configuration laws in both the reducing and the reduced theory. Let us move next to consider a situation in which a failure of mereological supervenience occurs, but without British Emergence.

When we combine quantum mechanical parts into a composite, such as the singlet state of Equation 1.1, we do so using a set of *general* rules of composition given by the tensor product operation. (Any composite quantum system may be represented in a Hilbert space formed by taking the tensor product of those representing the individual parts. And any interaction between the parts can be described by a Hamiltonian for the whole system, whose eigenstates are in this larger space.) Whatever the configuration of the parts, this procedure does not change, and there are no specific rules of composition.[30] The fact that we cannot assign pure states to the individual components of an entangled composite system (what Humphreys (1997) calls a "fused" system) is irrelevant.

A situation illustrating the converse possibility — mereological supervenience obtaining, yet British emergence — is provided by Broad's own favourite example. His silver chloride exhibits British emergence, due to the presence of a specific law of composition, but it would satisfy mereological supervenience. Broad wrote before the logical apparatus of supervenience was brought to bear on the questions of emergence, but one can infer his position from passages such as the following:

> No doubt the properties of silver-chloride are completely *determined* by

[30] Although it may be argued that astonishingly novel properties appear when quantum mechanical components come together in any configuration covered under the description: "measurement apparatus".

> those of silver and chlorine; in the sense that whenever you have a whole composed of these two elements in certain proportions and relations you have something with the characteristic properties of silver-chloride, and that nothing has these properties except a whole composed in this way. (Broad, 1925, 64, italics his)

Leaving aside Broad's stronger claim of uniqueness (irrelevant to our concerns), it is clear that he allows that the properties of a composite system exhibiting "British" emergence could nevertheless supervene on the properties of its parts, in that it could be 'completely determined' by them. His point is just that a specific law of composition may be involved in this determination.

The most elusive cases are those involving inter-theoretic reduction. In the main, this is due to the power and flexibility of that framework. In §1.5.2 we saw that it is possible both to express British Emergence as a particular case of a failure of inter-theoretic reduction, and as holding under completely different circumstances; simply by making appropriate choices of the properties allowed in the reducing theory $T_A$ and in the reduced theory $T_B$, or by altering the resources allowed for the construction of definitional extensions. The same trick can also be worked to express mereological supervenience as a special case of inter-theoretic reduction: let $T_A$ be a theory mentioning only properties assignable to separate parts of the system, and let $T_B$ also allow properties assignable to the system as a whole. (Naturally, supervenience will also fail in this case.) The very fact that the framework of inter-theoretic reduction can be so used, indicates that it will help little in analysing a detailed notion of emergence.

Undaunted, Howard (2005, 4-6) claims that inter-theoretic reduction and mereological supervenience are totally independent of one another. His examples of mereological supervenience without reduction are situations with no exceptionless laws governing the supervening level. Unfortunately, such examples rely on some fairly contentious additions to the basic framework of definitional extension. In

particular, Howard requires some unstated standard of what to count as a true exceptionless law. Otherwise, the fact of mereological supervenience guarantees that we can always construct laws that are in a sense exceptionless, (they may contain infinitely long disjunctive statements), along the same lines as outlined in Note 21.

Howard also gives a rather puzzling set of examples to illustrate the opposite possibility. He claims that a success in theoretic reduction does not imply mereological supervenience if boundary conditions become important. He mentions as examples 'global metrical structure in big-bang cosmology or edge state excitations in the fractional quantum Hall effect.' Judging from these examples, Howard seems to assume that when non-nomic circumstances have to be added to the dynamical laws in order to give a successful reduction, then mereological supervenience fails. But the need for some boundary conditions is a fairly general feature of physical theories which incorporate dynamical laws and one which comes in to play in the majority of realistic reductions; so by his standards a failure would be endemic.

### 1.6.2 Properties of Parts — Three Readings

I suggest that the main reason for the diversity of these analyses of emergence lies not in any fundamental difference as to "novelty", but in where it is looked for. For while each of the metaphysical approaches searches for a clear distinction between novel systemic properties of a large system and the properties of its parts, they interpret the entities involved in this relationship in three separate ways. We can sum up these differences as follows (see also Figures 1.1 – 1.3):

**Emergence understood as a failure of Inter-Theoretic Reduction** is concerned with the relationship between the properties appearing in two different descriptions of the *same* composite system — one description refers to microphysical

properties, the other to macrophysical. Thus, when we speak of "properties of parts" and "properties of wholes" in this sense, we are referring to two different theoretical descriptions of the same composite system.

**British Emergence** is concerned with the relationship between the properties of *different* systems. On the one hand: the systemic properties of the composite system; on the other: the properties of their components, in isolation and in other composites than the one in question. Thus "properties of parts" refers to those properties that are exhibited by components when separated from their places in the composite. "Properties of wholes" refers to systemic properties that are exhibited by the whole composite system.

**Emergence understood as a failure of Mereological Supervenience** is concerned with the relationship between the systemic properties of the composite systems, and those that can be assigned to individual components whilst part of that composite system. Thus "properties of parts" refers to those that may be predicated of the components while in their places within the composite system. "Properties of wholes" refers to those that may be predicated of the whole composite.

### 1.6.3 Their General Unsuitability

It appears that these metaphysical pictures fail not only as an analysis of the New Emergentist position, but are also inappropriate for anyone with broadly physicalist outlook who wants to take an emergentist position. The reasons should have already become clear from the discussion, but a brief summary may be useful.

In regard to inter-theoretic reduction, we have already discussed its sensitive dependence on the resources that we allow to construct definitional extensions, as

well as its suspiciously large flexibility. Indeed, it can be used to express almost any notion of emergence, simply by choosing an appropriate set of properties to be covered by the "reduced" and "reducing" theories. If any clear conception of emergence can be made out using the general framework of theory reduction, it seems that all the work in articulating emergentist criteria will appear in the details — either in some restriction in the resources allowed for definitional extensions, or in some addition to those basic requirements. So, as it stands, the general framework of reduction is *too* general to be useful to us.

In regard to British Emergence, the difficulty is empirical. Molecular chemistry has allowed made much progress on the prediction of chemical properties from general laws of composition which Broad noted to be impossible in his time.[31] The motivation for postulating specific laws withered with each new success, and the attraction of the whole position wilted with it. As McLaughlin argues (1992), the reasons for the failure of the British Emergentist movement lay not in any incoherence in Broad's hypotheses, but in the empirical fact that no specific laws of composition were found.

Finally, a failure of Mereological Supervenience is both well-defined and empirically realised in many systems. In fact it is realised *too* frequently. Any everyday macroscopic system exhibits entanglement, both internally, and with its environment. On this criterion, all large-scale systemic properties are emergent (with the possible exception of a few prepared in astonishingly delicate experimental situations).

More seriously, making out emergence as a failure of mereological supervenience makes no attempt to capture the emergentist intuitions of novelty or unexpectedness between micro- and macro-physical properties.[32] The systemic properties can

[31]The progress has not been so simple nor so successful as may be thought. It has been argued that many principles taken as basic to chemistry are still underivable from atomic physics, even today (e.g., Scerri and McIntyre, 1997).

[32]I would make a similar criticism of many suggestions in the "Complex Adaptive Systems"

be entirely predictable and derivable from those of the component parts (as indeed they are, in the examples of entangled electrons). Instead, this analysis looks far closer to the notion of *Holism*: it formalises the idea that it may be impossible to capture all the facts about the state of a composite system by considering properties of parts in isolation from their status as parts of a whole.

So, to summarise a summary: our first criterion is too general, the second never fulfilled, the third fulfilled too often.

These shortcomings do not just disqualify the three approaches from capturing the precise notion of emergence that concerns Anderson and his immediate followers; the failure is more general. Workers across complexity theory, chaos, general systems theory and even in biology and the social sciences claim that they accept that laws of microphysics govern all the properties of the systems they study, yet the 'higher-level' properties and laws that they study are importantly novel.

Each area — indeed each author — may well differ on the criteria by which they mark out this novelty; but however it is made out these three offerings from metaphysics all appear unsuitable, for the general reasons just listed.

## 1.7 Applying the Lessons from Metaphysics

Let us review our position. In §1.2 we identified the central doctrines of New Emergentism as microphysicalism, together with the existence and novelty of certain systemic properties. In §1.3 we tried to fill out microphysicalism in the form of a clear metaphysical doctrine — microphysical supervenience. We tried to fill out the distinction of "novelty" in four different ways, first (§1.4) as an epistemological or methodological doctrine, and then (§1.5) with three different metaphysical

literature for a firm analysis of emergence: they seem focussed on the idea that 'emergent' properties are those that can be possessed by systems, but not by the parts (e.g., Bar-Yam, 2004).

approaches. Yet each of these senses have proved unsuitable.

Before we are forced to retreat, we should look at the lessons that have arisen from our metaphysical project, in order to disambiguate the New Emergentist claims. We claimed in §1.6.2 that the three metaphysical approaches to novelty came apart due to three different ways of understanding 'properties of parts'. In §1.7.1, we shall apply this lesson by making clear which of these senses the New Emergentists are concerned with.

But the tripartite distinction in the sense of 'properties of parts' has repercussions beyond making clear how different senses of emergence can hold under different circumstances. It can also be applied to distinguish three senses of 'microphysical properties'; a term that has been central to our discussion of the New Emergentists. In §1.7.2, we shall look at how these distinctions might further clarify the New Emergentist position. And in §1.7.3 we shall see how neglecting them has led to confusion amongst commentators on the New Emergentists' claims.

### 1.7.1 Parts and Wholes for the New Emergentists

The New Emergentists are concerned with a relationship between the properties of parts and wholes. So the question naturally occurs as to which, if any, of the interpretations separated in §1.6.2 is closest to their understanding. On this point at least, I believe that the claims of the New Emergentists can be best understood as an updated variation of British Emergence.

Here is Anderson on the relationship with which he is concerned:

> The behaviour of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles. Instead, at each level of complexity, entirely new properties appear, and the understanding of the new behaviours requires research which I think is as fundamental in its nature as any other. (Anderson, 1972, 393)

Lest we think that Anderson's focus on "understanding" and "research" means that he is here making only a subjective point about the methodology of science, consider this stronger statement from Mayr, who could almost be quoting Broad:

> ... characteristics of the whole cannot (not even in theory) be deduced from the most complete knowledge of the components taken separately or in other partial combinations. This appearance of new characteristics in wholes has been designated as emergence. (Mayr, 1982, 61-3)

It is difficult to read these claims as concerned with either the relationship between different descriptions of the same composite systems (inter-theoretic reduction) or with the properties that can be assigned to mereological parts of a composite system (mereological supervenience). Rather it seems clear that the New Emergentists understand the relation between parts and wholes in a similar way to Broad: properties of "wholes" are understood as systemic properties of composite systems, and properties of "parts" as properties exhibited by the components in isolation or as parts of other composites.

### 1.7.2 Physicalism and Microphysicalism Revisited

Recall from §1.2.3 that the central claims of New Emergentism both involved an appeal to microphysical properties. The first was that all properties of systems were determined by the microphysical; the second that there was a distinction of novelty between the microphysical properties and those of large systems. After the discussions of §1.6, the inherent ambiguity should leap out immediately: there are three possible senses of 'microphysical':

**Microphysical$_1$** – a sense associated with microphysical supervenience (and with intertheoretic reduction), which admits as "microphysical" all of those properties describable under a recognisably microphysical theory — say, the Standard Model. These may include properties of very large systems.

**Microphysical$_2$** – a sense associated with British Emergence, which admits as "microphysical" only those properties of systems below a certain size, which they display in isolation, or when they are parts of systems other than the one in question.

**Microphysical$_3$** – a sense associated with mereological supervenience, which admits as "microphysical" only those properties assignable to mereological parts below a certain size.

In regard to interpreting the New Emergentist's claims of microphysicalism, we can immediately dismiss microphysical$_3$ since it is violated by quantum entanglement. Microphysical$_1$ is the sense captured by microphysical supervenience as we have understood it, but we can see now that it is too weak a condition to exclude such positions as British Emergence. And we do wish to exclude them, for in contrast to Broad, Anderson feels that laws that 'we know pretty well' govern all the everyday properties, including those encountered in condensed matter physics. To make out their notion of microphysicalism, we must therefore add to the condition of microphysical$_1$ supervenience, a further hypothesis that there are no specific laws of composition.[33]

But for the New Emergentists' second claim, of novelty, it seems clear that we are speaking of microphysical$_2$ properties. An emergentist distinction of novelty should be between the properties of small systems and the properties of large systems, not between the descriptions offered by two different theories (as microphysicalism$_1$ would have it), nor the holistic sense of microphysicalism$_3$. Another way of putting the same points would be to say that whatever the New

[33]It might be thought that we could do this far more simply, by specifying that only microphysical laws operate (as we quoted Pettit: back in §1.2.1). My quarrel with this approach might have become clear in the discussion in §1.5.2 as to whether specific laws of composition are microphysical or not. Any ambiguity in "microphysical" applies just as much to microphysical laws as it does to microphysical properties, and even a large-scale specific law of configuration would be a microphysical$_1$ law.

Emergentist distinction of novelty may amount to, it is exhibited even amongst the set of microphysical$_1$ properties. In fact, a simple way to avoid further confusion would be to restrict the whole discussion to this set, and look for our sense of novelty within these.

### 1.7.3 Diagnosis

If these suggestions are correct, then we can begin to understand why there has been confusion when philosophers concerned with metaphysical discussions of emergence have tried to evaluate the New Emergentist claims. We saw back in §1.4 that the orthodox interpretation of the New Emergentists is that as a metaphysical thesis, their claims are simply inconsistent. And now we can see why this opinion might have arisen. They are explicit about accepting microphysicalism, but if this is interpreted as microphysicalism$_2$, then this would commit them to saying that all properties supervene on the properties of very small systems considered in isolation. And this claim would be inconsistent with their claim of novelty as we have just interpreted it. Fortunately, they are not committed to holding microphysicalism$_2$; they are committed to microphysicalism$_1$ and an absence of specific laws of composition.

Similar confusions between the different senses of parts and wholes can lead to bewilderment as to what the New Emergentists are claiming. For example, Don Howard recently attempted to interpret New Emergentism either as a failure of reduction or as a failure of mereological supervenience, and concluded that he could not see that it had any relation to either. He finished his paper with the frank admission that he was puzzled at the phenomenon of condensed matter physicists believing that their subject could be called "emergent" (Howard, 2005, 17). I suggest that his puzzlement can be explained by the New Emergentists' understanding of microphysicalism and the relation of parts and wholes in the

"British" sense. Since it is a distinct relationship to the one at issue in both reduction and mereological supervenience, it may hold under entirely different circumstances.

As well as missing the particular parts/wholes distinction that the New Emergentists mean to draw, it is possible to make errors by conflating the three relationships separated here. Confusing two (or even all three) of the senses, has — I believe — led to misunderstanding in discussions of emergence. Among those who I would accuse on these grounds are Kronz and Tiehen (2002), who discuss the failure of mereological supervenience as a form of British Emergence. They are commenting on an earlier article by Humphreys (1997), who I think falls into similar traps.

Humphreys draws a very strong metaphysical conclusion from the failure of mereological supervenience. His own notation is rather involved, but he introduces a notion of 'levels': $L_1, L_2, ...L_n$, each the domain of a particular special science, so in our terms, the "smaller" levels would concern microphysical$_1$ properties (Humphreys, 1997, 5). But he then claims that 'microphysical properties' cease to exist when we have non-separable states. Instead, microphysical properties become'absorbed' into the systemic properties of the whole. This is a conclusion he judges on criteria related to their causal powers, but still it is clear that the sense of 'microphysical' is now microphysical$_3$. When he moves from the failure of mereological supervenience to argue that what he calls a 'fused' property cannot be represented at a lower level, the force of his case appears to turn on this equivocation (Humphreys, 1997, 10-11). (The properties of mereological parts might be said to be absorbed when they are brought together in a composite, but not the properties assigned by a theory of microphysics).

When Kronz and Tiehen comment on Humphreys' article, they repeat his equivocations in their discussion, and also add a discussion of British Emergence

(331), taking on the old terminology of 'emergent' and 'resultant' properties, but still treating all three positions as if they are concerned with the same relationship between parts and wholes.[34]

## 1.8 Concluding Remarks

We have seen that a variety of claims for the existence of emergent properties and laws that have been made by physicists can be characterised by a shared acceptance of physicalism, and a recognition of a distinction of some systemic properties as importantly "novel" and thus qualifying as emergent.

Filling out these twin doctrines in metaphysical terms has led us to a rather disappointing conclusion: although there is no shortage of precise distinctions of novelty, they are not appropriate for the cases that the New Emergentists focus on. What is more, the orthodox response is to accuse them of a simple confusion; one which leads them to draw a metaphysical lesson from an epistemological distinction. Yet in discussing the metaphysical approaches, we have uncovered a series of ambiguities in the comparison between microphysical and macrophysical properties. And it seems possible that it is at least partly down to these ambiguities that the New Emergentist position has been misinterpreted by metaphysicians.

Now that the ambiguities have been recognised, we are left with the New Emergentist position holding a very similar position to the one occupied by Broad in the 1920s. They are concerned with the same species of micro-macrophysical

[34]Kronz and Tiehen introduce further distinctions of their own within the non-separability of quantum properties (specifically, they argue that the situation where both the unitary evolution and the states of a system are non-separable is more important than a situation with only non-separable states), but this further technical discussion does not bring them any closer to the sense of emergence that Broad had in mind. For while he is concerned with microphysical$_2$, they are examining distinctions as between the various kinds of non-separable properties (microphysical$_3$), and relating this back to Humphreys' levels of description by different special sciences (microphysical$_1$).

distinction: between the systemic properties of large systems, and the properties of their components considered in isolation. And they each believe that certain larger systems have systemic properties that are importantly novel, compared to those of these smaller systems. But the New Emergentists also reject specific laws of composition, and this prevents them from taking Broad's more detailed position.

Broad himself did try to give a logically exhaustive set of alternatives for an emergentist position, judging that his own was the only one viable.[35] So with their rejection of configuration laws, it looks as though the New Emergentists might have run out of metaphysical ground on which to stake their claim. In the final section of this chapter I would like to look ahead to the next, and consider whether these various disambiguations leave them any space for manoeuvre at all.

### 1.8.1 An Inductive Argument

With the New Emergentists' doctrines of microphysicalism and the novelty of emergent properties now clear in metaphysical terms, we can see that they have still managed to paint themselves into a fairly tight corner. For consider the following argument.

Take two uncontentiously microphysical entities, consider each in isolation, and list their physical properties. So long as these entities are indeed small enough, it is uncontentious that each member of this set will be amongst the microphysical$_2$ properties. Now consider $S_2$, the set of microphysical$_1$ properties of the system formed by the *composition* of the two entities. According to the New Emergentists, all the properties in $S_2$ must be derivable from the microphysical$_2$ properties, using

[35] I have not followed the structure of Broad's own detailed discussion (Broad, 1925, Ch.2) because it is not entirely clear that his points translate well to modern physics. In particular, his 'Ideal of Pure Mechanism' — supposed to contrast with the Emergentist thesis — is very much based on turn-of-the-century physics, and is not relevant to modern discussions. In Appendix A I outline how microphysical supervenience (microphysical$_1$, that is) precludes the existence of one variety of Broad's emergent properties (the non-behaviourably analysable), and in the main text, I have concentrated on his alternative: specific laws of composition.

only general laws of composition. Thus, there seems no room for novelty to enter into the properties that are admitted into this new set. Next, consider $S_3$, the properties of the composition of this joint system with some further microphysical entity. Again, only general laws of composition can be used, and again, there seems no room for novelty to enter. Now iterate this process, adding one uncontentiously microphysical entity at a time, and applying only general laws of composition at each stage. Since the New Emergentists grant entity microphysicalism, they must also grant that any physical system of $N$ microphysical parts could be constructed in $N$ iterations of composition.

But then it is difficult to resist the obvious inductive conclusion that one needs a merely algorithmic process to obtain the set $S_N$ containing all the microphysical$_1$ properties of this system. No novelty in the properties seems to have entered at any stage. Yet in the last section we concluded that the New Emergentists' distinction of novelty should appear amongst this microphysical$_1$ set.[36]

There are certainly a number of ways to resist this sort of inductive argument. For example, it could be pointed out that we will encounter steps where there are no physically possible systems of the requisite type. Most obviously, if we started considering the properties of the very smallest entities, we might find ourselves attempting to list the properties of free quarks — arguably a non-sensical project. But the New Emergentist case does not seem to turn on these technicalities — they would make their same claims if our microphysical theory came without such restrictions.

Another response would be to point out that the form of the argument is very reminiscent of Sorites reasoning. One could then argue that from accepting that no novelty enters at any single stage, we cannot infer the conclusion that no novelty

[36]If we are unhappy with this last step, an alternative would be to appeal directly to microphysical supervenience, and conclude that all properties are fixed by the algorithmic procedure above, since they supervene on the microphysical$_1$.

enters overall, and argue on analogy with one's favourite response to that paradox.

But even if we are inclined to resist the argument in these ways, it still looks as though in principle, one needs no more resources to derive the whole set of systemic properties of any large concrete system than the resources we use to deal with small ones. In both cases, we just need microphysical$_2$ properties, and the same small number of general laws of composition. All that is needed to deal with larger systems, is an enormous amount of calculating power.

Assuming we accept the force of the inductive argument, our way forward seems very restricted — but it is also clear. We are searching for a metaphysically significant distinction of novelty amongst the systemic properties, but we have established that the membership of the appropriate set of systemic properties for any entity is fixed by a process that admits of no novelty.

So if there is such novelty, and with metaphysical significance (rather than being "merely" epistemic) we must look for it in distinctions *within* the fixed set of systemic properties. Fortunately, there are many distinctions that metaphysicians recognise in the status of sets of properties. For example, while the above argument seems to fix the set of first-order properties, it does not touch certain varieties of second-order properties, i.e.: properties *of* properties. It is not clear that it fixes the division of intrinsic/extrinsic, nor any distinctions of naturalness, nor any matching with universals. If such distinctions could be shown to change in an unexpected manner as we consider the systemic properties of larger systems, then it is possible that some novelty in these characters could mark emergent properties. The next chapter will look at one approach along these lines which — I shall argue — is very promising in implementing the themes of the New Emergentists.

### 1.8.2 Reasons for Optimism

Given the almost relentlessly negative conclusions from this chapter, one might wonder whether physicists and metaphysicians are concerned with the same issues when they speak of emergence, physicalism and microphysicalism. Why should we reject the orthodox conclusion that the two sides are simply talking past one another, one addressing practical concerns of physicists, the other more abstract concerns of philosophers? We close this discussion with some reasons for hope in this regard.

Metaphysicians have put forward a variety of explanations for the popularity of physicalism, and specifically of microphysicalism, but by far the most popular motivation is a desire to adopt a metaphysics that is informed by the success of physics. Crane and Mellor (1990) sum up this attitude in their discussion (and criticism) of physicalism. They attribute the 'now almost orthodox' belief in physicalism and microphysicalism to the progress achieved by physical sciences, this 'grants physical science a unique ontological authority: the authority to tell us what there is.' (394).

I believe that this is a laudable desire, and that it carries across directly to a need to capture a metaphysical sense of emergence faithful to any that may have been highlighted by discoveries in physics. The caveat is that we must be careful to separate the success of physics, and a particular attitude to that success. In accepting a supervenience on microphysical$_1$ properties and the rejection of specific laws of composition, we are following the authority of physics, which it has acquired mainly from to its empirical success. But in according a privileged role to microphysical$_2$ properties, and to the microphysical laws that explicitly mention them, we are inspired not directly by the practice and success of physics, but rather by a particular attitude to that practice and success. This emphasises the microphysical$_2$ theories such as the Standard Model, and the properties mentioned

in them as particularly fundamental. But this attitude by no means shared by all physicists, it is not supported by the empirical success of physics, and in particular it is precisely the one which the New Emergentists strongly oppose.

# Chapter 2

# Interpreting New Emergentism: A New Approach

In the last chapter, we looked at the difficulties inherent in relating metaphysical conceptions of emergence to those claimed by physicists. The claims of Philip Anderson were taken as representative of a movement dubbed 'New Emergentism', which we tried to characterise in existing metaphysical terms, though with limited success. In this chapter, we shall try to build on those largely negative conclusions and construct a metaphysical position that could be occupied by Anderson and his followers; one which supports a clear conception of emergence founded on well-understood physical principles.

The main negative conclusion of the last chapter was that existing metaphysical interpretations of emergence were unsuitable to make out the position held by the New Emergentists. I argued that emergence was best thought of as a distinction of novelty between microphysical and systemic properties, but that any such distinction was wracked with ambiguities. Once these were cleared away, I argued that the most promising place to look for novelty was in distinctions of metaphysical status within the set of systemic properties. In this chapter, I shall

propose an interpretation of the New Emergentist claims, along those lines.

In §2.1 & §2.2, I shall look at examples of distinctions of metaphysical status that might be suitable to make out the New Emergentist thesis. In §2.3, I discuss how the New and British Emergentist theses can be seen as parallels, the critical advance being that the New Emergentists look to well-understood physical theory to produce novel properties, rather than having to appeal to speculative metaphysics, as did Broad. In §2.4 I present Anderson's principal example of a "novelty-producing" process, that of symmetry-breaking. Finally, in §2.5, I address the very different interpretations of Anderson's views that have appeared in the philosophical literature.

## 2.1 Intension

In metaphysical terms, not all properties are created equal. There are a variety of distinctions that can be recognised: intrinsic/extrinsic, determinables/determinates, fixed degree/multigrade, structured/unstructured and categorical/dispositional are just some of the best-known divisions. Most relevant to our concerns is a distinction between the sparse "metaphysically respectable" properties that mark objective resemblance, and the plentiful gerrymandered properties that may be definable in a certain language, yet play no significant role in the world.

For example, David Armstrong holds that just because predicate "P" and predicate "Q" each pick out physically significant properties, it does not follow that the disjunction does so. Some, like Armstrong (1978), make out such distinctions in terms of universals, others prefer a trope theory or the "naturalness" of properties. Following the useful, but slightly non-standard terminology of Wilson (1993, 54), I shall refer to all such distinctions as between the *intensional* characters amongst the systemic properties, using the term to cover any consideration that makes a

distinction of privilege amongst properties, distinguishing those that mark objective resemblance from "mere" gerrymandered ones; or the physically significant from the insignificant. The importance of making an intensional distinction was most famously summed up by David Lewis:

> Any class of things, be it ever so gerrymandered and miscellaneous and indescribable in thought and language and be it ever so superfluous in characterising the world, is nevertheless a property ... Because properties are so abundant, they are undiscriminating. Any two things share infinitely many properties, and fail to share infinitely many others ... Properties carve reality at the joints — and everywhere else as well. If it's distinctions we want, too much structure is no better than none. It would be otherwise if we had not only the countless throng of all properties, but also an élite minority of special properties. (Lewis, 1983, 13)

Lewis counts as 'adequate' only those theories of properties that recognise such a distinction, naming theories due to Armstrong, Quinton, Quine and Bealer as adequate in this sense (1983, 13n6). But it was Lewis himself who went on to propose perhaps the most influential scheme of all.

### 2.1.1 David Lewis' Natural Properties

Lewis' famous aim was to vindicate 'Humean Supervenience': to show how an adequate metaphysics can be founded on a particularly sparse supervenience base: the intrinsic properties of point-like objects, and the spatio-temporal relations between them.

Lewis begins by dictating that the properties in his supervenience base provide a standard of *perfectly natural* properties.[1] Other than the aforementioned conditions, he leaves it up to a future physics to determine membership of this set,

[1] Other influential metaphysicians have constructed similar schemes, and appeal to a hierarchy of natural properties, though few start from such a sparse supervenience base as does Lewis. As well as those mentioned by Lewis himself, we can count proposals from Shoemaker (1980), Putnam (1975b, 305-322) and Armstrong (1978).

noting only that present theories suggest that charge, spin and quark colour might be among them.[2]

He then imagines a restricted language, which contains only predicates that refer to perfectly natural properties. All other properties can then be arranged on a descending scale of *naturalness*, depending on the length and complexity of the definition needed to specify them in such a language. (Lewis (1983, 14) argues that a distinction of degree is preferable to a binary one.)

This second-order property of 'naturalness' then finds application throughout his metaphysics. Without going into details (which would be unnecessary for our discussion) we can summarise by saying that he employs natural properties to determine duplicates, causal roles and questions of reference.[3]

When his scheme is worked through, it soon becomes clear that Lewis' spectrum of naturalness is often at odds with the physical sciences to which it is supposed to be faithful.[4] As an example, consider an elementary physical system: a single free classical particle. Since Lewis is taking lessons from physics, its natural properties should include the particle's mass, velocity and spatio-temporal position — those found significant in physical descriptions of the particle, and appearing in the laws which govern it. Unfortunately, when we move to systems made up from more than one such particle, agreement between naturalness and physical importance breaks down.

For consider a system made up from two such particles. Firstly, Lewis' scheme appears to pick out the wrong quantities as its natural ones, at least if we are to be guided by their importance to physical theory. Relativity tells us that the

[2]Lewis considers three schemes for setting up these perfectly natural properties: in terms of universals, tropes, or simply taken as a brute fact about certain properties. These schemes are set out in most detail in Lewis (1983) and updated in Lewis (1986b, 1994).

[3]Lewis (1983) is the best source for the central roles Lewis assigns to natural properties. It is also updated in Lewis (1986b, 1994).

[4]Lewis (1994, 226) makes carefully nuanced claims that his metaphysics is inspired by physics and aims to accommodate its discoveries.

physically important total mass is not a simple sum of the masses of the parts, but must also take into account the energy of any interaction between the two. But this mass-energy supervenes on Lewis' completely natural properties in a far more complex way than the simple sum, since it also involves the interaction between the two particles.

Secondly, Lewis' scheme does not privilege those quantities that *are* physically significant. For example, in our two-particle system, the centre of mass $(\mathbf{r}_1 m_1 + \mathbf{r}_1 m_2)/M$ expresses an extremely important quantity of the system, as does the angular momentum around this abstract centre. But these are just two among the bewildering variety of properties that supervene on the microphysical, and there is nothing to distinguish them from countless others that can be derived with equal or greater simplicity.

Add more elements, and move beyond simple particles and potential wells, and the variety of physical properties increases rapidly to bewildering levels. Yet we know from our very ability to do macroscopic physics that there will often be a privileged minority of systemic properties: they will be causally potent, important in physical theory, determining facts such as how a large system vibrates, breaks, bends or heats up. But these properties will not in general be picked out by the ease with which they can be defined in terms of the microphysical theory; as such Lewis does not count them as natural. It is puzzling that while metaphysicians such as Lewis are often happy to bow to the authority of physics to make intensional distinctions at small scales, this authority somehow dissolves when we move to larger, composite systems.[5]

[5] Although some authors follow Lewis in tying a naturalness scale to the mereological composition of large systems, it is worth noting that there are other approaches available, which together cover many different ways in which we might compose properties together to yield properties of composite systems. One such possibility is to recognise *structural universals*, as do writers such as Forrest (1984), and Bigelow and Pargetter (1989). Lewis (1986a) criticises such schemes, arguing that mereological analysis is the only metaphysically respectable approach.

Not nearly enough has been said here to establish that Lewis' scale of naturalness is unsatisfactory, but there exists a fast-growing literature raising very serious problems (insurmountable, in my view). Some of these arise due to the sparseness of Lewis' choice of supervenience basis (e.g., Robinson (1989), Zimmerman (1999)), so they might be accommodated by expanding this basis in a suitable way, but there are also more general objections, which call into question the tenability of the whole approach (e.g., Black (2000), Mainwood (2003)). Not least amongst the more general criticisms are those of Mark Wilson, which we shall look at in §2.2. But before this, we need to address a danger, brought into sight by schemes of naturalness such as that of Lewis.

### 2.1.2 Microphysical Supervenience — Yet Again

In the last chapter, we allowed that the New Emergentists accepted microphysicalism, and analysed it using the apparatus of microphysical supervenience. Our supervenience basis was not quite so rigidly circumscribed as Lewis', but it consists wholly of microphysical properties, and shares its supervenience physicalism.[6] So now there seems to be a problem associated with accepting microphysical supervenience as an analysis of microphysicalism: namely that we might implicitly assume that microphysical properties are a privileged minority, and the rest less so, by analogy to the scale accepted explicitly by Lewis.

Let us put this reasoning explicitly: any property that we count as part of the microphysical supervenience basis qualifies as physical in an elementary, fundamental sense. For the thesis of microphysicalism counts them directly as physical. In contrast, if asked to decide whether a large-scale systemic property is physical, we decide it by a process that ultimately returns to this same supervenience base.

[6] As mentioned in §1.3, there are issues of the modal range over which the supervenience thesis is formulated. Lewis (1983, 25-33) discusses the range of possible worlds that he takes for supervenience physicalism.

Thus, the large-scale systemic properties might also be physical, but only qualify in a derivative way. This makes it look as though microphysical supervenience has a built-in bias to count microphysical properties as having a higher intensional status — a clear conflict with the New Emergentist theses.

But this impression is mistaken: not only is there no pressure on the New Emergentists to adopt such a scheme, they could not do so even if they wanted to. To see this, take the properties mentioned by the Standard Model as the microphysical set, as Anderson suggests. It is natural to think first of examples such as the "colour", "charge" and "spin". But we now need to recognise the distinctions in the meaning of the term 'microphysical' that we drew in Chapter 1. These most natural examples are all microphysical$_2$: those which can be assigned to lone particles, or to very small systems in isolation. But these are not the only properties available in the supervenience base allowed by the New Emergentists. Theirs is the microphysical$_1$ set, i.e., all the properties that can be accommodated in the Standard Model, and this includes enormously complicated properties of large systems, which would qualify as physical just as directly as do the microphysical$_2$.

So one microphysical prejudice — the belief that supervenience physicalism grants a privileged place to properties of the very small — arises as a result of confusing microphysical$_1$ and microphysical$_2$ properties. Once we look at the scheme carefully, we see that systemic properties of large systems qualify as physical just as directly as do those of point-like particles. So even if we do imbue the supervenience basis with some higher intensional status, the systemic properties of larger systems lose nothing compared to the very smallest. Lewis' own scheme only imposes such a promotion because of his commitment to a microphysical$_2$ supervenience basis which forms part of his 'Humean project'. We could of course follow him, but we have already seen that this is not the position that the New Emergentists hold.

Alternatively, microphysical$_2$ properties might be thought to qualify as having intensional status on slightly different grounds, namely that they appear directly in the laws most central to the Standard Model (which is the theory that qualified them as microphysical in the first place). Of course this is precisely the reverse of Lewis' own scheme — for him it is perfect naturalness that qualifies properties to appear in fundamental laws, not the other way around (Lewis, 1986b, 1994). But we are free to take such a position, and this would also introduce a bias towards microphysical properties, and this time towards microphysical$_2$.

But this position is also mistaken, though the confusion is a different one. If the laws central to the Standard Model distinguish some basic set of properties, so — equally — does any other physical theory.[7] And it is central to the New Emergentist position that the laws of microphysical theories such as the Standard Model are *not* privileged *a priori* in any such respect. That is, any distinction of intensionality cannot be decided by their appearance in laws central to microphysical theories, as against their appearance in laws central to higher-level ones, unless we have already rejected the New Emergentists' position. For handing over a judgment of a physical property's intensional status to one particular microphysical theory would be to beg the question against their position before it has even be made out.

To help dissolve this second form of microphysical prejudice, note first that the same problem does not arise once we assign properties some other role outside physics (as biological or chemical properties, say). We may be quite happy to call a property such as 'valence' a fundamental *chemical* property, because of its fundamental status within the laws of chemistry, yet also recognise that it has no fundamental physical status. Indeed, it is a strong advantage which supervenience

[7]It might be suggested that it is the *formulation* of a particular physical theory that privileges a particular set, and that different formulations of the same theory might differ in their choice. I am happy to accept this, and it does not affect my point.

formulations hold over rival notions of physicalism: a property can be said to have a certain status *qua* physical, and another with respect to other special sciences.[8]

The science of physics itself appears to be the one exception to this rule. But this appearance is simply an illusion. If microphysical theories such as the Standard Model are to confer any higher intensional status, it must be restricted to a status *qua micro*physics, not *qua* physics.[9] That is, it might be useful if we want to know how "microphysically fundamental" is a particular property, and all properties would (presumably) have such a status; but it must be recognised that other physical theories might award it differently. Physics is not just microphysics (in any of our senses). For again, conflation of physics and microphysics is exactly the sort of picture that the New Emergentists are concerned to *deny*.

So, while property microphysicalism may be accepted by the New Emergentists, metaphysicians abuse their acceptance if they think that this implies some pressure to hand over all authority to that microphysical theories to decide questions of intensionality.

## 2.2 Emancipationists and their Enemies

Conflicts with New Emergentism are far from the most serious accusation that philosophers have made against the traditional ways of drawing intensional distinctions. In a pair of papers, Mark Wilson (1985, 1993) presents a wide-ranging critique of the accounts of property physicalism presented by contemporary metaphysicians, and on the basis of this critique he then puts forward some minimum standards for any such distinctions between genuine and gerrymandered properties.

[8]Barry Taylor sets out an explicit proposal of this type with his 'vegetarian' alternative to Lewis' red-blooded metaphysical distinction of naturalness, judging intensional status as relative to a theory (Taylor, 1993). In these terms 'valence' would be a natural property relative to chemistry, but not relative to physics.

[9]I have doubts as to whether such a scale is even useful in these contexts, but it should certainly have no significance beyond them.

His standards are relatively generous: any distinctions of intensionality must be at least minimally informed by physics, in the sense that those considered central to physical laws must be judged as genuine — i.e., to have a respectable intensional status, and not judged as just a gerrymandered "quasi-property".

In §2.2.1 I shall look at Wilson's own criticisms of the metaphysical schemes for drawing distinctions of intensionality, and his own minimum standards for any such distinction. I shall then look at the status these minimum standards should have, in particular, in regard to how much say physics should have in determining the metaphysical status of properties (§2.2.2). Finally, in §2.2.3 I shall argue that if we adopt such a standard, then discovering intensionally important properties of larger systems from the properties of their parts is a non-trivial procedure. This looks ahead to §2.3, where I shall argue for an interpretation of the New Emergentist claims based on this fact.

### 2.2.1 Wilson against the Emancipationists

Wilson (1993) labels those metaphysicians whose approach he objects to as 'Property Emancipationists', counting amongst them Lewis, Armstrong, Fodor, Shoemaker and Putnam. These Emancipationists hold that properties cannot derive any metaphysical status they may have from the roles they play in our language, but must be "freed" to allow their status to be judged on purely metaphysical grounds. He agrees broadly with the aims of this movement, but not with the details of their approaches. In particular, he objects to any appeal to 'basic laws' of physics, where these are taken to be some unknown set of universally quantified statements — assumed to pick out predicates referring to 'natural kinds', 'perfectly natural properties' or some other set of fundamental physical properties.

Wilson's papers make a persuasive (and entertaining) case that these modern metaphysicians appeal to the authority of "science" or "physics" in a piecemeal

manner, cherry-picking those aspects that that suit the requirements of their metaphysical systems, and ignoring less welcome lessons. Wilson complains in one of his memorable lines, that 'so often the "Science" cited is simply Philosophy in dark glasses with a phony passport'.

Lewis' scheme, outlined in the last section, provides a good example of the sort of practice that Wilson objects to — Lewis appeals to physics to provide a list of a basic set of properties, expressible in terms of natural kind predicates, on which his scale of naturalness can be built. And this aspect is shared, at least to some extent, by all the Emancipationists. Wilson attacks such schemes on the grounds that they are inadequate to account for the actual practice of physics. For physicists will happily go ahead and formulate laws that mention a far wider variety of properties than can be derived in any reasonably restricted sense, from any set of such basic predicates. What is more, they will consider some amongst this wider set as amongst the most physically significant for many systems.

To make this case, we will follow Wilson in looking at general frameworks used in physics to define properties of a wide variety of physical systems. In this chapter we will concentrate on a fairly simple set of examples, the state spaces of classical mechanics. (The discussion of Chapter 3 covers other examples of state-spaces from a rather different point of view.)

### The State Space Framework

A state space allows an abstract representation of all possible physical states of a system: each represented by a single point in a multi-dimensional space. The number of dimensions of the space is usually determined by the number of degrees of freedom of the system in question (though it may sometimes be useful to consider restrictions to certain lower-dimensional subspaces of the full state space, or to embed it in a higher-dimensional state space, if the problem requires it.)

As a simple example, let us consider a classical, Newtonian system $S$ made up of $N$ component particles, each of constant mass $m_1, m_2, ..., m_N$, perhaps interacting in some way. There are $3N$ configurational degrees of freedom: $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3..., \mathbf{q}_N$ — where each $\mathbf{q}$ is the the Cartesian position 3-vector for each particle. The spatial configuration of the whole system can thus be represented by a single point in $3N$ dimensional *configuration space*. Each of these particles also has its associated velocity — $\dot{\mathbf{q}}_1, \dot{\mathbf{q}}_2, \dot{\mathbf{q}}_3..., \dot{\mathbf{q}}_N$ — the derivative of each $\mathbf{q}$ with respect to time. In Newtonian physics, the full physical state of a system is determined by this data, so the full physical state can be represented by a single point in a $6N$ dimensional *state space* (also called the velocity phase space) coordinatised by the $\mathbf{q}$'s and $\dot{\mathbf{q}}$'s. And a temporal evolution of the system can be represented by a path in the space, linking the instantaneous states to one another.

But the very specific situation we have considered here (a particulate system of constant masses described in 3D Cartesian coordinates) is a necessary condition neither on the applicability of the state space approach, nor of classical Lagrangian mechanics. One of the great advantages of the approach is that the $\mathbf{q}$'s can be *generalised coordinates*, rather than Cartesian position vectors above and may represent a variety of degrees of freedom, so long as they and their associated *generalised velocities* $\dot{\mathbf{q}}$, completely specify the physical state. Masses can vary, and we can consider extended bodies.

Beyond the classical Lagrangian scheme, closely-related state space approaches include the Hamiltonian and Hamilton-Jacobi frameworks. These differ both in the ways in which the degrees of freedom are represented in a state space, and in the functions which are defined and extremised to obtain the physical evolution of the system.[10] And it is not just finite-dimensional classical mechanics that can

[10] An introduction to classical state space techniques requires far more discussion than we can engage with here. Classic works on the foundations and techniques of analytic mechanics include Arnold (1989) and Lanczos (1986). Careful discussions from a philosophical standpoint

be accommodated in the state space framework (at least not in the rather broad sense that I wish to use the term). Field theories use infinite-dimensional state spaces to accommodate their continuous degrees of freedom, with a Lagrangian of the field defined on such a space. And the states of quantum systems are represented in a space of a very different sort (a complex Hilbert space), with other important differences, such as observables of the system being represented by Hermitian operators, rather than real-valued functions on the state-space. I do not wish to go into the relations between these many powerful frameworks, but merely to point out that the overall approach of state space methods can be generalised far beyond our simple example. For the moment, we follow Wilson in focussing on finite-dimensional classical, Lagrangian mechanics, and focus on making general philosophical points about the framework.

We are free to define any functions from points of this space to the reals, to $n$-tuples of reals, to vectors, tensors or to anything else we desire. Each of these functions has the potential to represent properties, or qualities of the system.[11] Some, such as the function mapping each state to the $n$-tuple representing the acceleration of each particle, should be amongst the intensionally significant — they are important when we seek to characterise the system's behaviour. At the other end of the intensional scale, a function that maps all states to a single, arbitrary imaginary number, seems to have little, if any link to the physics of the system.

---

are provided by Butterfield (2003, 2006a).

[11] Wilson uses such terms quite freely, but we can be a little more careful while trying to accommodate their usual usage. Recall the distinction of classes of properties as determinate values of determinables as introduced in §1.2.2. Roughly, a function can be taken to correspond to a determinable, and a particular value of that function to a determinate value of that determinable. In what follows, I shall refer to the determinables as quantities; and to a particular determinate value of them, as a property. For example, a determinable such as the velocity of particle $i$ corresponds to a function from the state space to the vectors: $v(i) = (\dot{q}_{i,x}, \dot{q}_{i,y}, \dot{q}_{i,z})$. But a particular velocity — say $10m/s$ in the $x$-direction — would be a particular value of that function. So velocity would be a quantity, and a particular value would be a property: more or less in keeping with everyday use.

Familiar examples suggest that we can recognise the intensionality of large-scale systemic properties through their roles in the physics of the system. A good example is provided by rigid bodies, where a very short list of properties: its total mass, velocity, angular momentum and moment of inertia determine the evolution of a system that may be made up of an extremely large number of interacting particles. The rigidity of the system provides a *constraint*, which allows a very few, very specific, quantities to determine the physical state of the entire system. Thinking of this in state-space terms, it constrains the system to evolve within a much lower-dimensional hyper-surface than the full state space, allowing us to characterise its state with very few coordinates. Or if we are presented with a fluid, its viscosity and density will be especially significant, and its incompressibility will act as a similar constraint (though this of course allows far more freedom of the state space, than would a full rigidity).

Wilson gives a particularly striking set of examples: the Chlandi modes, which are patterns of vibration exhibited by metal plates when they are held by an edge and struck. (The craftsman E. F. F. Chlandi first discovered these patterns in the eighteenth century, by clamping plates in particular positions, and sprinkling fine sand on them. The areas with the largest motion shook off their sand, the stationary areas retained it, revealing the *modes* of vibration.) These modes are amongst the most physically important properties that the plates possess. They dictate the sounds that a plate will make when struck gently, and if struck harder, the places at which it would break.

Wilson emphasises that the set of physically important properties and quantities vary widely from system to system. Different shapes and constitution of plate will give rise to different modes, and so it is a great advantage of state space techniques that we can *re-coordinatise* the state space in terms of those quantities that express these modes most simply. Expressed in their terms, the evolution

of the system through different modes looks very simple, so they are often called 'good coordinates' for a particular system.

Wilson argues that any proposal for recognising intensional distinctions must recognise good coordinates such as the Chlandi modes as amongst the physically significant. Yet he argues that proposals such as Lewis' will inevitably omit them. For they base intensional distinctions on the ease by which properties can be defined in a language that refers directly to some perfectly natural set. But for any perfectly natural set, we can come up with systems with good coordinates that are only definable with extreme difficulty — by infinite-length connectives for example.

### Wilson's Criterion ∃

To demonstrate the poverty of the Emancipationist's proposals, Wilson presents an example of a bead sliding without friction on a wire, twisted into an irregular shape. The position of the bead $a$ along the wire from its starting position is given by a triple of quantities which we could express as a function $Q(t)$. For example, this could be expressed fairly straightforwardly in Cartesian coordinates $Q(t) = f(x(t), y(t), z(t))$. Wilson objects:

> But given any finite language $L_m$ we can choose our shape $S$ so that the needed $f$, and consequently $Q$, will not be definable in $L_m$, by [...] cardinality considerations [...]. But in several ways, the undefined $Q$ will be an *important quantity* for our system. For example ... $Q$ proves a natural quantity for predicting $a$'s future states; $a$ will move around $S$ with a constant $Q$-"velocity" The composite quantity $P = m\dot{Q}$ is likewise a conserved quantity ... But although $Q$ and $P$ represent some of $S + a$'s most *important physical properties*, they won't be definable in $L_m$. (Wilson, 1993, 231-2)

Wilson's more general point can be put as follows: It turns out that the specific features of the problem can make the most ostensibly arcane functions on state space express important quantities, even those that could not be simply defined

from any "basic" set of properties. Accordingly physicists have placed some properties centrally in physical laws that would be judged overwhelmingly "unnatural" from any Emancipationist standpoint.

Anything similar to Lewis' scale of 'naturalness' is especially vulnerable to such objections, since he draws distinctions based on the length of the logical connectives needed to define them, and $Q$ and $P$'s definitions will be infinitely long. Nevertheless, the quantities $Q$ and $P$ are central to the physics of the system.[12]

Certainly, there are responses available for Emancipationists in general, and for Lewis in particular. We could devise a more sophisticated approach for judging 'complexity' than a simple counting of logical connectives in a privileged language. And for the specific example of the bead, Lewis could reply that he means for all spatio-temporal properties to be perfectly natural, even those difficult to define from some original Cartesian system. However, we shall not pursue a detailed analysis of the validity of Wilson's criticisms, nor whether each of the Emancipationist authors would be able to make adjustments to accommodate them. Instead, we shall look at the minimal standards that Wilson sets for property physicalism as informed by physics; for these are of interest whether or not we believe that the Emancipationists can satisfy them.

Wilson argues that to take account of a realistic physics, *any* real-valued function on the state space must be considered a physical property. Accordingly, he suggests a minimum standard for the properties, or traits of a system, which he calls 'criterion $\exists$'. (A similar criterion is discussed in Wilson (1993)).

> **Criterion** $\exists$: For any system $S$ with Lagrangian $L$ and any real-valued mathematical relation[13] $r$, there is a physical trait $\phi$ such that $\phi(S, t)$ if

[12]Note that the microphysicalism accepted by the New Emergentists does not fall foul of this argument. Supervenience formulations of both physicalism and microphysicalism may include values of $Q$ and $P$ as physical properties since it is permissible for a physical property to be definable from a (finite) basis set only in terms of an infinitely long disjuction in $L_m$ (as we saw in §1.5.1).

[13]Wilson uses the terms 'relation' and 'function' quite freely (and as far as I can see, inter-

> and only if $r(\mathbf{q}(S,t), \dot{\mathbf{q}}(S,t), t)$ where the $\mathbf{q}$ represent Cartesian location coordinates for the members of $S$ and the $\dot{\mathbf{q}}$ are the corresponding velocities. (Wilson, 1985, 232)

But Wilson goes further: each of the quantities counted by $\exists$ must also have potential to be judged amongst the sparse, intensionally significant set.

> For virtually any of ($\exists$)'s postulated traits, no matter how convoluted, one can concoct a system for which it proves a good coordinate, a fact which helps explain ($\exists$)'s great ontological generosity. (Wilson, 1985, 237)

Wilson is thus attracted to $\exists$ since it provides a suitably wide-ranging set of physical quantities and properties, suitable to circumvent his objections to the Emancipationists. The physical traits (= quantities for Wilson) are as various as real-valued functions, which is to say all the mappings from points in the state space to the real numbers. He does not even impose minimal conditions of continuity or smoothness (though he admits that in this he differs from some physicists Wilson (1993, 75-80)).[14] Wilson holds that metaphysical theories should be held to a minimum standard: to admit the range of properties recognised by physics, and claims that the Emancipationists' approaches fail to meet this standard.

### 2.2.2 The Hempelian Dilemma

Before we go on to apply our new conceptions of physicalism to the New Emergentist thesis, there is an important objection to be addressed. Nelson (1985) charges Wilson with letting concerns from physics encroach illegitimately upon metaphysics. Nelson claims that there are certain issues of physicalism that can

changeably) in his discussion.

[14] Once we move beyond classical, Lagrangian, finitistic mechanics, even Wilson's conditions must be relaxed. Certainly, we may wish to generalise beyond the reals to consider vectors or tensors as physical. When we quantize a system, we replace functions by Hermitian operators, and it is a question central to the interpretation of quantum mechanics if these operators can be understood as representing physical properties.

be recognised as distinctly 'physicist's inssues' and others 'philosophers' issues', and that there is no reason to think that approaches to one set should have any bearing on approaches to the other. He explicitly includes amongst the latter the question of which are the fundamental properties of the world, as well as how to re-identify properties over time (Nelson, 1985, 279). According to Nelson, the state-space approach is only of help in addressing questions of 'physicists' physicalism', such as how to express their equations of motion in tractable form, and Wilson is mistaken to think that they should form any kind of standard for answers to the philosophical questions.

Ross (1999) replies to Nelson on Wilson's behalf, arguing that Nelson's attack is question-begging. The issues he identifies with the 'philosophers' physicalism' are just those that can be addressed by a logical analysis of a basic set of predicates in a privileged language. Yet whether this Emancipationist approach is valid, is precisely the question at issue. Ross argues that Nelson has not realised the significance of Wilson's proposals. According to her, Wilson's project is to restructure the whole issue of physicalism, rejecting traditional metaphysical approaches to defining physical properties, and replacing them with methods faithful to physical practice.

I believe that Ross is largely correct in her criticism of Nelson — his criteria for judging an issue a 'philosophical' one simply precludes any approach other than the Emancipationist one. But in defending Wilson, she overstates the scope of his project. Wilson himself does *not* claim that his criterion $\exists$ is meant to be a general criterion for property physicalism. Rather, his attitude is better expressed as the claim that $\exists$ expresses the range of a general framework now used to discuss physical laws. He does not claim that the framework is maximally general, even for present physics, much less than it will continue to be in the future. And he does not claim approaches from outside physics cannot accommodate the set delimited

by $\exists$. His criticism is that the present Emancipationist approaches cannot even meet the present standards.

Indeed, if we follow Ross' example, ignore Wilson's own caution, and put forward $\exists$ itself as a naturalistic criterion for physicalism, and the 'good coordinates' as naturalistic criteria for an intensional distinction, we run straight into a version of the 'Hempelian Dilemma' that we mentioned in §1.3.1. There we saw that any naturalistic approach of physicalism faces the choice between restricting any future physics, and making a trivial definition of physical properties. And this applys equally to any intensional distinction: either we restrict how future physics judges such a distinction, or our criterion is trivial.

Yet this "Ross-Wilson approach" navigates the Hempelian Dilemma with remarkable success.[15] It is true that most physical theories were not originally formulated in a state-space approach, but I do not know of a modern successful theory that *cannot* be expressed in its terms or in the terms of some close relative. The principles on which it is based: The Principle of Least Action and its variants, appear remarkably general elements of physical law.[16] Yet it is also difficult to see how non-physicalist theories can be placed in such a framework — for example, it is very hard to see how primitive mental properties can be governed by a least-action principle.

Despite this unexpected success, it is perhaps wiser to take Wilson's modest path. We require that any definition of physicalism, and any intensional distinc-

[15] Ross (1999, 100) suggests that the particular state space approach is not essential to Wilson's proposals, but any future physics could be expected to yield a modified version of $\exists$. I am not sure that Wilson would agree — it seems to trap the approach on the "trivial" horn of the Hempelian dilemma.

[16] Claims are sometimes made that the roots of classical physics in quantum mechanics explains why the least action principle holds so widely across physics. Taylor (2001, 179-182) is typical in holding up Feynman's 'sum-over-histories' approach as showing how the quantum mechanical paths that extremise action will be the classical ones. As far as I can see, such explanations make no suggestion of why similar principles hold in QM itself, nor in areas well away from these considerations, such as general relativity.

tions must accommodate the quantities picked out by good co-ordinates, for we know that they are important in today's physics. It also appears inevitable that state-space approaches will continue to be useful in future physics, but to rely on it seems a risk without reward. If we take a modest approach we can impose a necessary condition that good coordinates must be recognised in any set of privileged natural properties. We can safely leave aside the questions over whether they could also constitute a sufficient criterion.

### 2.2.3 The Elusive "Good Coordinates"

To give an intuitive idea of how state space methods can help discover good coordinates, Wilson considers the path traced out in state space by the evolution of a system over a certain period of time. Suppose that we write this path in terms of Cartesian coordinates: $q(x, y, z)$ and $\dot{q}(x, y, z)$ which track the position and velocity of each of the particles. If we were given a list of how these (microphysical$_1$) quantities change with time, their significance would be opaque, and in general the evolution would not look at all regular.

> But now introduce a quantity **u**, supplying the *path the system is on*, and a quantity **v** marking the *system's current place around that path*, measured as an angle. That is, the trajectory travels on a level surface for **u**. Viewed from **u** and **v**'s perspective, the motion looks very regular. (Wilson, 1993, 77)

In this example, **u** is called a *constant of motion* of the system, since its value does not change as the system evolves through time. Wilson emphasises that if there exists a quantity which "slices" the state space into such level surfaces, along which a system evolves (a *foliation* of the state space) we may be able to specify a system's state in terms of a smaller set of parameters. One subset specifies which amongst a certain class of 'orbits' the system follows, another subset serves to specify its place along that orbit.

This leads to several obvious questions: given a physical system, are there always good coordinates? And if there are, how are they to be discovered? These questions turn out to have no straightforward answers. One approach to the first is summed up in the local existence and uniqueness theorem for systems of ordinary differential equations[17] which, under a wide range of conditions, guarantee the existence of a *local* set of good coordinates on state space; i.e., a neighbourhood in which the solution to the equations of motion (expressed as differential equations) will be trivial. But this startling result has two important caveats: first, such a coordinatisation cannot always be extended outside the local state space neighbourhood; second, any guarantee that a coordinatisation exists is no guarantee that it can be easily discovered.[18]

Certain methods do exist for uncovering these coordinatisations. One is to exploit any *constraints* on a system, such as the rigidity of a body, or incompressibility of a fluid, as mentioned above. It is often possible to incorporate these constraints into a judicious selection of good coordinates which can lead to a great simplification of the problem with a significant reduction of the number of degrees of freedom.

Another approach which may be applied more generally, is to discover symmetries of the dynamics of a system. By exploiting certain dynamical symmetries we can discover associated good coordinates by a variety of methods. Amongst the most powerful results is Noether's theorem: which relates a variational symmetry of the Lagrangian $L$ to a conserved quantity in the motion of the system. This theorem becomes even more powerful when translated into the Hamiltonian approach, where the conserved quantity gives an immediate reduction in the number of linked equations that need to be solved. Also within the Hamiltonian approach,

[17]For a detailed treatment of this theorem see e.g., Olver (2000, Prop. 1.29).

[18]There will also exist local constants of the motion, though sadly, for most problems they will *only* exist locally. Butterfield (2006a) discusses the limited sense in which this theorem guarantees a 'solution' to all problems in classical mechanics.

the powerful modern techniques of symplectic reduction can be used to generalise the techniques associated with Noether's theorem even further.[19]

We now find an unexpected consequence of adopting a physicalism sufficiently broad to accommodate the properties recognised by physical science. With the enormous range of physical properties provided by $\exists$, we will have no problem in algorthimically discovering many physical properties of a large system, indeed we are overwhelmed with them. But somewhere amongst this multitude lurk the good coordinates which represent the few, simple characteristics that dictate the important features of the system's behaviour. But the good coordinates are not at all linked to the simplicity with which they are definable from microphysical properties. In particular, there is no general algorithm to discover them from the good coordinates of smaller parts.

## 2.3 New Emergentism and British Emergentism

In the last chapter, we concluded that the contrast between the properties of parts and of wholes that concerned the New Emergentists, was identical — or at least very close — to the contrast which concerned Broad and the British Emergentists. But we were puzzled as to where the novelty could come from for the New Emergentists, since Broad's proposal of specific laws of composition had been, in effect, empirically refuted.

In adopting an intensional distinction, we have come to demand more than Broad. He was concerned with being able to derive the set of physical properties of a composite system from the physical properties of its parts. We further demand that we be able to distinguish natural properties from amongst the vast set of

[19]Sussman and Wisdom (2001, 208-220) gives a full discussion of the advantages of the Hamiltonian and Lagrangian approaches in exploiting conserved quantities. A detailed discussion of these techniques from a philosophical point of view, without any sacrifice of rigour, is given in (Butterfield, 2006b).

physical properties of a composite system. And we have just seen that it is *not* easy at all to discover these from the properties of their parts.

Except in the very simplest cases, there are no general algorithms into which we can feed the "good coordinates" for systems made up of parts in isolation, and obtain the good coordinates for the whole system. In this sense, a clear analogue of Broadian emergence could be said to be endemic in physics. Uncovering the good coordinates is a task specific to each system (or small classes of systems) requiring creativity, insight — or, putting it in tendentious terms —'research as fundamental as any other.'

With these elements in place, we can see our way to a surprisingly straightforward interpretation of the New Emergentist claims, which I shall present in this section and the next. First, in §2.3.1 I shall argue that both Broad and Anderson follow a strategy of first exhibiting evidence for emergence: the novel and unexpected character of certain systemic properties, and only then presenting a physical process — a "mechanism" — that explains how such novelty can arise. Then, in §2.3.2 I shall bring out the ways in which methodological differences might provide evidence for a metaphysical distinction behind them, in the writings of both Broad and the New Emergentists. In §2.3.3 I shall bring out and answer a serious objection to the interpretation, before moving to look at the metaphysical mechanism that the New Emergentists suggest, in §2.4.

### 2.3.1 Emergence: Evidence vs. Mechanisms

Let us return to the New Emergentist claim of 'anti-constructivism' which we first met in §1.7. This is the claim that although we know the microphysical laws that govern the interactions of the parts, it is impossible *in practice* to use these to derive the significant properties of large systems. Nonetheless, they continue, fields such as condensed matter physics have discovered laws which govern the

properties of these larger systems, though these laws are often qualitatively very different than those that govern the interactions of the parts.

This situation can be compared directly with the earlier distinction we made between Broad's evidence for emergence, and the mechanism for it (§1.5.2). In chemical phenomena, Broad recognised a qualitative difference between the dynamical laws that successfully describe the properties of composite systems, and the dynamical laws that successfully describe the properties of their components. He also noted that it seemed impossible to derive the systemic properties from the components: his observations could be said to be the anti-constructivism of his time.

But if this analogy is a faithful one, then the modern observation of anti-constructivism is only *evidence* for the existence of emergence,[20] but it cannot constitute it. This is in opposition to the orthodox line discussed in §1.4: that it is the practical difficulty of derivation that constitutes and defines emergent properties.

Broad's own next move was to put forward a bold metaphysical hypothesis to explain his methodological evidence: specific laws of composition. But here the New Emergentists part company with Broad. One of the central themes of *More is Different* and its follow-up articles is to show how an appeal to modern physical theory can make such speculative moves redundant. The qualitative differences between the laws governing the systemic properties and the microphysical laws are shown to be explicable by a well-understood physical process, without any appeal to Broad-style configurational laws.

Recognising this distinction between evidence for emergence and a constitutive mechanism, is central to the reading of New Emergentism that I shall present.

[20]There is a separate class of evidence as well as anti-constructivism: the apparent irrelevance of many microphysical details to certain large-scale physical properties. Laughlin and Pines concentrate more on this class, and call such properties *protected*. Their claims, and the arguments that they build upon them, I shall consider in Chapter 5.

Unfortunately, the distinction is not made explicit in Anderson's original paper, nor is it emphasised particularly strongly in later writings. Accordingly, there exist several alternative readings of the New Emergentist claims, which I shall postpone to a later section (§2.5).

### 2.3.2 Parallels — Useful and Misleading

I feel that the parallels between British and New Emergentist approaches extend beyond the evidence/mechanism distinction I have just drawn. Indeed, the characterisation of the main claims of the New Emergentists, back in §1.2.3, could equally be a characterisation of those of the British Emergentists.

Both the British and the New Emergentists start from the observation that the natural sciences appear to be split into distinct "layers" or "levels" of description, roughly corresponding to the domains of the particular special sciences. These are purely methodological observations: the systemic properties of larger, more complex systems are studied using very different techniques from those that are used to study the smaller ones. And in general, the systemic properties of larger systems appear qualitatively different from those of smaller ones. Further, the science at each level appears to involve a great deal of research which is largely uninformed by the research going on at any other level — in particular, it is uninformed by the study of the properties of smaller systems. For Broad, the principal example of a distinct higher level was chemistry; for Anderson it is condensed matter physics.

This methodological separation is contrasted with their shared belief that all of the entities studied in any of the levels are all composed from microphysical parts, and interact according to general laws, which completely determine all their systemic properties. Both the British and New Emergentists want to discover how this striking separation comes about. How do the systemic properties of larger systems appear so different from the properties of the smaller ones, when they are

composed from them, and obey the same laws?

In the separation of chemistry and physics, Broad was presented with such a methodological division, and was led to postulate a metaphysical mechanism which explained it: his specific laws of composition. The laws also formed the barrier for mechanically deriving the properties of wholes from the properties of the parts. The New Emergentists do not think there is such a barrier to *Broad's* procedure: all of the properties of the systems which we considered in §2.4 could have been derived from the properties of the parts that made it up, so long as we had the resources of Broad's Mathematical Archangel. But they do think that there is a barrier to something else: to mechanically derive the intensionally significant properties of large systems, from those of their parts.

Following Wilson's suggestions, we have assumed that this set must include those represented by the 'good coordinates' of the state space of each system in question. So, the suggestion is now obvious: if the set of good coordinates can be specific to a system (or a class of systems) this has the potential to explain the qualitative difference in the most important properties and laws used in particle physics, and those used in condensed matter physics. Accordingly, the New Emergentists must follow Broad in proposing a *mechanism*, by which the good coordinates of large systems in certain configurations could be radically different to those of their component parts. This mechanism would then possess both metaphysical and physical significance, but there is no suggestion either that it need contravene microphysical supervenience, nor need it be as speculative as Broad's specific laws of composition. But if this interpretation is correct, the mechanism must be one that switches the good coordinates of large systems away from those of other systems made up from the very same constituent parts.

Though I have been pushing these parallels hard, at this point they abruptly come apart: British Emergence seems not to be realised empirically, and the New

Emergentists claim that their mechanism does not share this drawback. But before looking in detail at the mechanism that the New Emergentists propose (in §2.4) we should address a very direct objection. The microphysicalist can plausibly claim that *whatever* mechanism the New Emergentists present, so long as they grant microphysicalism and deny specific laws of composition, there *will be* a general method of deriving the good coordinates of a large system from those of their components.

### 2.3.3 General Laws After All?

The New Emergentists grant that is possible to use the general laws of composition to construct *all* the physical systemic properties. But then, given this (enormous) set, and contrary to what I have suggested, Reductionist opponents can point to a totally general procedure for discovering the good coordinates. The general laws of composition allow them (or at least Broad's Archangel) to trace the physical evolution of the system through state space, even if it is coordinatised with the very complex "microphysical" coordinates. But then they can simply *define* a set of good coordinates from this coordinatisation. Wilson gives an explanation of how this might be done very simply (which, strictly, ignores various issues of continuity and smoothness). If we take the path representing the actual evolution of the system in state space, and define a vector $\mathbf{u}$ as the tangent to this path at every point, this will let us define a set of good coordinates, a further parameter giving the position along the path, constructed in a similar way to Wilson's 'action-angle' coordinates of §2.2.3. True, the definitions may be extremely complicated (indeed they may be infinitely long), but it is still a *general* algorithm to obtain good coordinates. So if we put this together with the procedure for obtaining all the systemic properties, we have a direct analogue of Broad's general laws of compositon: a procedure which starts from the properties of the parts of a system

and produces the good coordinates of any system composed from them, using only laws that apply generally.

There are two clear answers we could make to such a challenge. The first is that by defining them in such a way, we risk trivialising the notion of good coordinates and of associated constants of motion. Nontrivial definitions of the constants of motion of a system should include certain restrictions (e.g., of differentiability) since the trajectory of *any* $n$-dimensional system will have co-dimension one and thus there will always exist $n-1$ "trivialised" constants of motion. I am somewhat suspicious of this argument, since although there is no *guarantee* that the constants of motion and good coordinates discovered by this method will be differentiable, they may well be. That is, if there are suitable constants of motion, according to whatever criteria we choose to avoid triviality, there is nothing to stop us discovering them by this method. (Of course, trivial ones may be discovered as well, but we can discard them.)

Fortunately, we need not head off our reductionist at that point. We can admit the objection, for it says nothing against the central New Emergentist case. For she has shown, first that she is able to define the good coordinates in terms of microphysical properties; and second, that she has a very complex, but general procedure for doing so. Broad himself may have been disappointed by this second point, but the New Emergentists can accept it. There is still no practical general method, and there is a metaphysically important mechanism behind this practical impossibility — a switch from one set of good coordinates to a completely different set. This is not the sense of novelty in the specific sense of underivability that we were surveyed earlier §1.4, but it remains a metaphysically important distinction of novelty.

It is here that a spurious parallel between New Emergentism and British Emergence might have led us astray. For Broad believed that the practical failure of

deriving chemical properties from the general laws of composition and atomic physics, was evidence for a mechanism that played two roles. First, it explained how the failure came about, though the science of chemistry was still possible. Simply, chemistry takes into account the effects of specific laws of composition, while atomic physics does not. Second, the same mechanism also showed that the practical failure would also be a failure in principle. The New Emergentists think that their corresponding practical failure shows up a mechanism that plays only the first of Broad's roles. It explains why there is a practical barrier to derivation, yet condensed matter physics is still practically possible, since we can take advantage of the simplifications provided by the new set of good coordinates. It does *not* show that our failure of microphysical practice is also a failure in principle, but we do not need this to establish that the mechanism in question is both physically and metaphysically significant.

## 2.4 A Mechanism — Symmetry-Breaking

The New Emergentists stress that there is no single mechanism lying behind all instances of anti-constructivism nor behind all emergent properties and laws. Those mentioned by Laughlin, Pines and Anderson include: (1) *Renormalisation*, especially in association with the properties of 'soft modes' that occur near quantum and classical phase transitions; (2) both fractional and integer *Quantum Hall Effects*; and (3) Anderson's own principle of *Localisation* which appears in certain types of superconductivity and superfluidity. (The mechanism lying behind high-temperature superconductivity, they admit, remains elusive.) However, there is one class of mechanisms which they judge to be particularly basic, important, and which may provide the basis for a widely generalisable framework — *symmetry-breaking*.

In the introduction to the collection of his most important papers, Anderson identifies symmetry-breaking as 'the clearest instance of the process of emergence which lies behind *More is different*' (Anderson, 2005, vii) a judgment which already appears in that paper itself:

> In my own field of many-body physics, we are, perhaps, closer to our fundamental, intensive underpinnings than any other science in which non-trivial complexities occur, and as a result we have begun to formulate a general theory of just how this shift from quantitative to qualitative differentiation takes place. This formulation, called the theory of "broken symmetry", may be of help in making more generally clear the breakdown of the constructivist converse of reductionism. (Anderson, 1972, 393)

So, to understand the claims of the New Emergentists, it seems worth examining the theories associated with symmetry-breaking. In §2.4.1, we will narrow our focus to a particular kind of symmetry-breaking, in which it is the size and stability of the system that is responsible for the breakdown of the symmetries. This gives a clear connection to the mechanism that the New Emergentists are searching for, which is responsible for a "switch" in good coordinates when the system takes on certain configurations. Then in §2.4.2, we move to consider some concepts associated with symmetry-breaking states, using the example of a classical ferromagnet passing through its Curie point.

### 2.4.1 Varieties of Symmetry-Breaking

A word of caution before we engage with more general examples of broken symmetry. The New Emergentists are looking for a distinction between the intensionally significant properties of individual parts and those of the whole. From our discussion of §2.2, we might suggest that this will happen when the good coordinates of larger systems are very different from those of their parts. And we saw in §2.2.3 the strong links between symmetries of the dynamics and the good coordinates. Thus,

changes in these symmetries sound as they could bring about such a distinction. But to find a suitable mechanism for emergence, we must pay careful attention both to the particular symmetries that are at issue, and how they are broken. We need the symmetries broken to be those that determine the good coordinates of each system, and we need them to be broken by some aspect of the size or arrangement of the system in question. But these are not the only distinctions covered by the textbook definitions of 'symmetry-breaking'. Rather, these definitions refer to a situation in which there exist states of a system with lower symmetry than the dynamical laws whose solutions they are. And these are indeed the situations that are discussed in both the most elementary examples, such as a bead spinning on a vertical loop of wire (Greenberger, 1978); as well as the most advanced, such as vacuum states in quantum field theories. But this distinction between laws and their solutions is *not* directly related to the emergentist distinction between large and small systems: Anderson called his 1972 paper 'More is different', not 'Solutions are different'.

We can move to the variety of symmetry-breaking relevant to emergence, in two stages. First, we narrow down our concerns to *spontaneous* symmetry-breaking, which Anderson defines as follows:

> *Spontaneously* broken symmetry. Definition: Although the equations describing the state of a natural system are symmetric, the state itself is *not*, because the symmetric state can become unstable towards the formation of special relationships among the atoms, molecules, or electrons it consists of. (Anderson, 1984, 265)

This is meant to contrast with *explicitly* broken symmetry, where some symmetry-breaking term is inserted into the equations describing the state (for example, by introducing an external field term with a particular spatial direction).[21] So

[21]The application of terminology varies in different areas. Particle theorists recognise further distinctions such as *dynamical* symmetry breaking. A discussion is given in Castellani (2005).

Anderson is concerned with a variety of symmetry-breaking where it is the stability of the state of the system that leads to the breaking of symmetry.

This situation, where the energetic stability of the states leads to symmetry-breaking, is particularly associated with large systems at low energies. Perhaps the most familiar example are the regular periodic crystals. For a wide variety of forms of interaction amongst a large set of atoms or molecules, the most energetically favourable arrangement is some regular lattice-like configuration. And so, at the relatively low temperatures of our everyday experience, regular, rigid, lattice configurations are extremely familiar. These low-energy states are symmetry-breaking, in that they are not symmetric under continuous translations or rotations, though the Hamiltonian of the interaction between the parts do have such symmetries. At higher temperatures, the stable states (liquids, gases) tend to take on the full symmetry of the Hamiltonian, with no regular spatial arrangements. This exemplifies a general rule that large and cold systems, with low-energy states tend to break symmetries, though small and hot systems, with high-energy states do not.

Another familiar example is the phenomenon of ferromagnetism, a simple classical model of which we shall look at in the next section. Here, the symmetry broken is one of spatial orientation. At low temperatures, there are stable states with spins all aligned along a particular direction, but above the Curie temperature the states are unmagnetised, sharing the symmetry of the Lagrangian.

But we still need a second step to complete a link between this sort of symmetry-breaking and emergence: we need to relate the symmetries of the states with the symmetries that govern the properties of large systems. Here we must distinguish between low and high-energy processes, for the former are governed by the broken symmetries, the latter by the unbroken symmetries of the original Hamiltonian.

The properties of large and cold systems are dominated by low-energy dynam-

ics, which are by definition small deviations from their ground, or stable state (i.e., a state that may be symmetry-breaking). The good coordinates for these low-energy dynamical properties are not determined from the symmetries of the original Hamiltonian, if these have been broken. So we obtain exactly the contrast we need for symmetry-breaking to be a mechanism for emergence — in our sense of producing novel good coordinates for the system. Small systems have dynamics determined by the original set of symmetries of the interaction Hamiltonian, larger systems have a very different set specific to their state, which may not be shared by others made up from the same components.

Summing up, the symmetries of a large system are often not the same as the symmetries of the interaction Hamiltonian. And for the lower-energy processes that dominate the properties of these systems, the good coordinates will be determined by the broken symmetries, which can be utterly different to those exhibited by some different arrangement of the same microphysical components. Thus the most significant properties of a large system can be qualitatively different from those of a different system made up from the same parts — exactly what we are looking for, to understand the New Emergentist claims.

There is plenty of textual support from the New Emergentists for this being an appropriate way to understand the importance of symmetry-breaking:

> The Hamiltonian describing the internal interactions in our macroscopic system still obeys the initial symmetry group, whatever it may have been; the *state* of the matter involved does not, in some sense, at low enough temperatures; do any consequences flow from that? The answer is that some of the most important and interesting properties of matter flow from precisely this fact. (Anderson, 1984, 30)

And though Wilson does not identify himself as one of the New Emergentists, he sometimes speaks in very similar terms. Here he is outlining the surprising ways "good coordinates" may be found anywhere amongst the physical properties of a large system. This passage is important to what follows and is worth quoting in

full:

> The fact that "good coordinates" may potentially lurk anywhere within the bounds [allowed by criterion ∃] may be understood as follows: as one forms a system out of a collection of component molecules, their interactions may lead various sorts of "ordering" to lock in among them. Analogy: if one assembles complex shapes by hinging together triangles along their edges, certain configurations (e.g., Buckminster Fuller's famous "geodesic domes") will "lock in" a rigidity absent in most of the otherwise floppy complexes that one builds. Similarly, the intermolecular ties within Chlandi's plates manage to "freeze in" patterns of cooperation extending across all the molecules of the plate. This cooperation allows us to meaningfully ask how much energy is stored in each mode, giving rise to a set of good coordinates for the system. The "locking in" of a good coordinate can be so startling that physicists are sometimes tempted to speak of "the emergence of traits". Wilson (1993, 79)

Not only are physicists tempted; but I believe that they do so, at least in the case of the New Emergentists. Wilson's "locking in" of good coordinates can take place under many circumstances, but one of the most striking is when symmetry breaking states of matter are formed. The metallic crystalline solid which make up Chlandi's plates is just one example of such a state.

Anderson separates the most striking 'novel' properties of symmetry-breaking states into four classes: the discreteness of phase transitions and physical phases, the development of collective excitations (of which Chlandi's modes are an example), his own 'generalised rigidity', and the structure of defects in large systems (Anderson, 1984, 18-19). He goes on to examine each in some detail, but in this chapter we shall look only at the development of collective excitations, with a brief discussion in the next section. The discreteness of phase transitions is a tricky subject that forms part of the topic of Chapter 4, and we will say little further about the 'static' properties, such as the defect structure of solids. But it is important to realize that there are theoretical treatments available from physics for a wide range of both dynamic and static systemic properties, and in each case we can

trace their 'novel' nature back to the symmetry-breaking of the state concerned.

### 2.4.2 Order Parameters, Collective Excitations and Good Coordinates

Unfortunately, Anderson's 1972 claim that condensed matter physics was approaching a general theoretical understanding of the mechanism behind emergent properties, was rather premature. In particular, Anderson's own choice of examples in *More is Different* turned out to be slightly unfortunate. They concerned molecules, such as ammonia, which we observe in stable dipole-possessing states (Anderson (1972, 394-5), see also (1984, 27-30). As Joos (1996, 89-96) makes clear, the issue of why we observe dipole-possessing states of large molecules, rather than their non-symmetry-breaking superpositions is rather more complicated than Anderson suggests. Environmentally induced decoherence may be relevant, and this involves an appeal to an external system which distracts from the main points that Anderson wanted to illustrate.

It seems fair to say that the general framework which Anderson anticipated in the quotes above, does not yet exist, despite theoretical progress on many phenomena associated with symmetry breaking. That said, some steps have been made in such a direction, and here we can mention a few examples. First, in his own discussions Anderson focusses on extending Landau's concept of an *order parameter* characterising a phase of condensed matter system. Second, one of his original contributions is an associated concept of *generalised rigidity*, which aids general discussion of elementary excitations in condensed matter systems.[22] We could also mention the very different approach of the algebraic framework for quantum statistical mechanics, in which clear distinctions can be made between global, local

[22]For more details on his suggestions, see especially Anderson (1981, 1984, 1989).

and quasi-local observables, each relevant to different levels of description.[23]

Anderson's own discussions give some impression that symmetry-breaking can be discussed within a single theoretical framework. However, he moves very quickly between cases of quantum and classical symmetry-breaking, which are handled in rather different ways; indeed most of his case studies, as well as his suggestions for general frameworks are specific to the quantum context.

I feel it may not be helpful to complicate general points about emergent properties, with the inevitable questions that arise when we ask how to understand *any* properties of a quantum system. Therefore, I shall not follow Anderson's own case studies, nor his discussions of more general frameworks for symmetry-breaking since, despite their promise, they raise controversy that does not affect the philosophical points we will need. I shall rest satisfied with a brief discussion of a classical case. This should serve to give a qualitative idea of how a large (classical) system may acquire good coordinates which would not be suspected from examination of the interactions of its parts.[24] And I hope that it will provide a clear example of the mechanism for emergence which Anderson has in mind, without bringing in too many extraneous complications.

### Order Parameters and Order Parameter Fields

One of the simplest models of a ferromagnet is known as the classical Heisenberg model. Let us consider the two dimensional version: an infinite lattice of sites, on

[23] Algebraic QSM is perhaps not what Anderson had in mind when he suggested an emerging general approach, since it developed independently of his ideas. However, it is of much interpretative interest of its own, and we shall discuss some issues in §4.5.

[24] It could be argued that in taking this path, we will miss the trees while in search of the wood. It is characteristic of the New Emergentists that they concentrate on technical examples such as many-body quantum systems, which require second-quantization to make precise such concepts as quasi-particles or localization. In ignoring the specifics of these examples, it may well be that we miss important lessons. Certainly, we shall miss much of interest. However, given that they repeatedly stress that their general metaphysical points extend beyond their specific examples to properties in classical physics, chemistry, and biology, I think our course is a reasonable one.

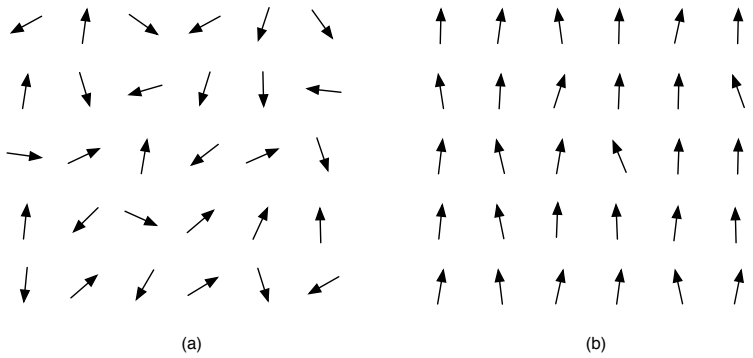



Figure 2.1: Heisenberg spins in a disordered phase (a) and ordered phase (b). In the disordered phase (a) the mean magnetisation is zero, in the ordered phase (b) it is a non-zero-vector pointing upwards, and as such is not invariant under a rotational transformation. There are are continuum of distinct ordered phases with the mean magnetisation pointing in each of the directions in the $x-y$ plane.

each of which is defined a "spin" vector $\mathbf{s}_i$, of unit length, but free to point in any direction in the $x-y$ plane. Without an external magnetic field, the interaction energy of the model can be written:

$$V = -J \sum_{<ij>} \mathbf{s}_i.\mathbf{s}_j \tag{2.1}$$

where the sum is taken over nearest-neighbour pairs of sites. $J$ is a parameter expressing the strength of interaction, positive values giving an energetic advantage to aligned spins. Note that this potential is symmetric under an overall rotation. Also, since we are neglecting the mass of any bearers of spins, the potential term is the only one in the Lagrangian and in the Hamiltonian of the system.

The Hamiltonian is thus invariant under rotations in the $x-y$ plane, the orthogonal group $O_2$. And in the higher-energy phase, the disordered states have spins pointing in all directions, as illustrated in Figure 2.1 (a), so they share the rotational symmetry of the Hamiltonian. This is the general case for condensed matter systems. But the lower energy, ordered states, as illustrated in 2.1 (b) are

not invariant under rotations.[25]

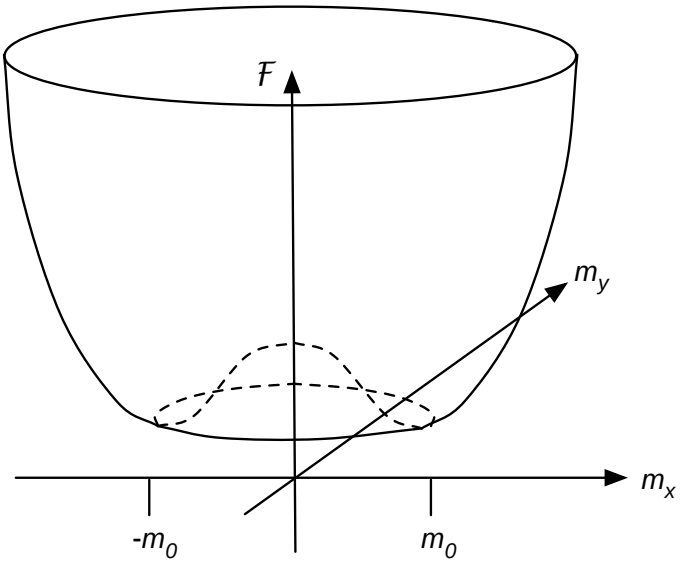



Figure 2.2: The free energy plotted against the $x$ and $y$ components of the mean magnetisation $\mathbf{m}$. The 'rut' of lowest energy is that of the stable degenerate ground states, which display broken symmetry.

To illustrate the low-energy ordered states, Figure 2.2 shows the free energy of the system, a form often referred to as a 'wine bottle' or 'mexican hat' potential for obvious reasons. The minima of the free energy, which characterise the stable states, describe a circle, the radius of which gives the magnitude of the mean magnetisation of such states. The *mean magnetisation* $\mathbf{m} = -J \sum_i \mathbf{s}_i$ is the *order parameter* for a low-temperature state of the Heisenberg model. Two useful ways of expressing it are as a 2-dimensional vector with a magnitude and direction: $\mathbf{m} = m(\cos\theta, \sin\theta)$; and as a complex number, with a magnitude and phase: $me^{i\theta}$. Its suitability as an order parameter is suggested by the fact that it is not

[25]N.B.: A difference in symmetry is a sufficient, but not a necessary condition for phase difference. For example, liquids and gases are different phases, but share symmetries. The order parameter concept can be extended to these cases too, where it is connected to the density of the matter. However, we shall not discuss examples of phase transitions in the absence of symmetry-breaking.

invariant under the symmetries of the full Hamiltonian, its value distinguishes the different ordered phases, and its magnitude quantifies the degree of ordering of each state.[26]

Another way of looking at broken symmetry is as the operation that maps one ordered state to another one in a different ordered phase but with the same free energy. That is, the rotations of the order parameter represent transformations between different equilibrium phases. In the two-dimensional ferromagnetic case, the symmetry broken is the $O(2)$ rotation group of the spins, isomorphic to the $U(1)$ group of transformations of the phase of a complex number. We can also introduce the space in which the order parameter can be represented: the *order parameter space*, which in the case of the 2D ferromagnet is the unit circle $S_1$. For 3D ferromagnets, the space is the surface of a unit sphere: $S_2$.

A few other examples may be useful to illustrate these ideas. Nematic liquid crystals are made up from long crystals which can align along particular axes. The symmetry broken is again a rotational one, but because parallel and anti-parallel alignment give the same energy expression, we identify antipodal points in order parameter space. This yields a space of a hemisphere rather than a sphere ($\mathrm{R}P^2$). The order parameter concept also extends to quantum systems, where the symmetries broken can be local, as well as global (i.e., the Hamiltonian is invariant under local applications of the operation in question). The most notable example is the local gauge symmetry of the superconductor.

This simple view of an order parameter as spatially constant over the system is already useful to characterise some of the large-scale static properties of the system, such as its heat capacity, electrical conductivity, and many of the defects that may form in it. But we are more concerned with dynamics, for which it

[26]For some suggested requirements for quantities to qualify as order parameters, see Anderson (1981) and Chaikin and Lubensky (1995, §3.6). In general, the complete description of ordered states require a full description of how the order parameter transforms under the action of the symmetry group of the original Hamiltonian.

is helpful to extend the order parameter concept, and allow $\mathbf{m}$ to vary spatially, giving us an *order parameter field*. So the order parameter field is a mapping from each point of physical space of the material, to its order parameter space. So for magnets, the order parameter field is given by the local magnetisation at each point: $\mathbf{m}(\mathbf{x})$; for nematic crystals it is given by the local alignment to the axes.

The order parameters of crystals are slightly harder to think about. As a state crystallises from the higher temperature liquid or gaseous states, it breaks a translational symmetry. As an example, take the lowest-energy states to be a regular square lattice arrangement. Now consider any more complicated state of a crystal as a possible deformation from this lattice to the actual one. We can then define an order parameter "vector" field, with $\mathbf{u}(\mathbf{x})$, defined at each point in the crystal, as the displacement from their equilibrium position. Of course, this $\mathbf{u}(\mathbf{x})$ is not a true vector, since once the real atoms move a distance comparable to the lattice spacing from their ideal placings, there is an ambiguity as to which ideal atom they should be associated with. When we choose a different reference atom, $\mathbf{u}(\mathbf{x})$ changes by a multiple of the lattice spacing $a$. Therefore, the order parameter space is a square with periodic boundary conditions, $\mathrm{T}^2$, most easily thought of as the surface of a torus.

### Collective Excitations

Let us consider the physical properties of a large system after a perturbation, say the physical properties of a two-dimensional crystal after it has been struck. Often, the most fruitful method is often to consider the stable states, discover the lowest energy variations from that state and analyse the effects of any perturbation in terms of a decomposition into these *elementary excitations*.

Stable states tend to be those with a uniform order-parameter field,[27] and

[27]The interesting cases of *frustrated equilibrium*, such as those found in the spin glasses are

elementary excitations can be characterised as variations in this field. We can see why they are so important for the dynamical properties of a symmetry-breaking state if we consider the energy cost associated with changing the order parameter field. Return to the Heisenberg magnet and consider a local variation of spins, i.e., a local change in $\mathbf{m}(\mathbf{x})$. Because of the underlying rotational symmetry of the interaction, the energy of such a variation can depend only on spatial derivatives of the order parameter field. That is, while a rotation of all the spins by the same amount would cost no energy, a local variation may. In ferromagnets local variations of this sort are called 'spin waves': excitations due to local changes in spin direction. Thus as we consider a less and less local variation, (i.e., as the wavelength of the spin-wave goes to infinity) the energy cost must go to zero. Thus the *elementary* excitations, costing the least energy, have very long wavelengths, and there will be a spectrum of higher-energy excitations with shorter wavelengths. Similarly, long-variations of the order parameter of the nematic crystal are known as 'rotational waves', and the square lattice system has sound waves (or phonons) which are local spatial variations from the equilibrium positions. Each of these are transmitted through the system with a dynamics which is highly dependent on the nature of the order parameter of the state.

One can gain insight into the physics by considering again the free energy of the system. It is easy to see that a spatially uniform change in the order parameter does not lead to an energy change (it is clear from Figure 2.2, for example). But it is also clear that spatially *non*-uniform changes will result in an energy change, since neighbouring spins must take up configurations other than fully aligned. Let us assume that the free energy of the system with an elementary excitation: $F_{el} = F[\mathbf{m}(\mathbf{x})] - F[\mathbf{m}(\mathbf{x} = const)]$, is analytic in $\nabla\mathbf{m}$. We require that a uniform state is minimal with respect to all possible variations in the order parameter, so

ignored here.

there are no linear terms in the expansion of $F_{el}$ around this uniform state in terms of $\nabla\mathbf{m}$. The simplest form which is consistent with such requirements is:

$$F_{el} = \frac{1}{2}\int d^2\rho_s[\nabla\mathbf{m}(\mathbf{x})]^2 \tag{2.2}$$

where $\rho_s$ is called the *spin-wave stiffness*, and is an instance of what Anderson calls *generalised rigidity*. This value is dependent on the strength of the interaction, the dimensionality of the lattice, and our chosen temperature. In our model $\rho_s = m^2Jz/8$ at $T = 0$, where $z$ is the *co-ordination number*; the number of others that each spin interacts with, so $z = 4$ for a square-lattice two-dimensional Heisenberg model. Observe that the energy cost vanishes with the magnitude of this gradient. This feature suggests that small disturbances away from the stable state will typically manifest themselves as long-wavelength fluctuations in the order parameter. And this is indeed the case; this rough sketch of an argument can be made precise and generalised as Goldstone's Theorem. Roughly speaking, this states that any state that breaks a continuous symmetry (i.e., a group of symmetry transformations which has an uncountable continuum of elements) will exhibit long-wavelength excitations. (In the quantum field context, a version of Goldstone's result was very important, since the absence of the predicted massless spin-zero particles led to the prediction of the Higgs boson — yet another case of parallel developments in high-energy and condensed matter physics).

### Good Coordinates

These elementary excitations, or 'soft' long-wavelength modes, dominate the low-energy dynamics of symmetry-broken states. At higher energies, excitations of shorter wavelength become important, and we view them as superpositions of higher modes of vibration. Thus these modes provide a means of discovering good

coordinates for the physical evolution of the system; that is, to recoordinatise the state space of a crystal in such a way that its physical evolution looks simple.[28]

In fact, Wilson's own examples of good coordinates — the Chlandi modes of crystalline plates — are examples of elementary excitations (though to be more precise, the Chlandi modes are standing waves, and the boundary conditions provided by the shape of the plate must also be taken into account). So here we have an example of how good coordinates can be derived from the nature of the order parameter, which sums up how the state has broken the original symmetries. The claim of the New Emergentists is that in the phenomenon of symmetry-breaking we have a mechanism by which the set of "good coordinates" of the whole can be entirely different from the sets of good coordinates which apply to the constituent parts when in isolation or in other wholes. When we find a system such as a Chlandi plate in a symmetry-breaking state, we can guarantee that many of the most significant properties of the system, such as the Chlandi modes, will be entirely different from the significant properties of smaller systems, even if they are made up from the same components. The phenomenon we called 'anti-constructivism' comes about (at least in these cases) because it is not possible to discover the good coordinates needed to describe the physics of large symmetry-breaking states from the good coordinates of other systems made up from similar components, for they may have a very different order parameter.

Naturally, symmetry-breaking is only one example of such a mechanism by

[28]It is important to notice that there need be nothing approximate or non-exact about a decomposition into elementary modes, or excitations. It is true in many cases, that we are particularly interested in small variations, and so we are satisfied with a Taylor expansion around the stable state, cutting off terms above a certain order. But this is not intrinsic to the approach: we can continue any such series to higher orders, or we can consider exact expressions. For if we consider the variation of the order parameter as a field on space, decomposing any variation in terms of the elementary excitations is analogous to the decomposability of any field variation into its Fourier components. In practice, continuing to higher orders turns out to become mathematically intractable, but then so does consideration of the interactions of all the smallest parts, and this approach usually allows us a great deal more accuracy while remaining tractable.

which we can explain a striking separation between the good coordinates associated with large systems and those associated with smaller ones. Even the general theory of symmetry-breaking that Anderson hoped for would not be a general theory of emergence.

Even in condensed matter physics, symmetry-breaking is not the only mechanism that the New Emergentists appeal to. Perhaps the most significant change since Anderson's 1972 article is the development of understanding of renormalisation techniques.[29] I shall postpone a discussion of renormalisation techniques to Chapter 3, where I shall make some comments about its relevance to the New Emergentist's themes. But for the moment, suffice it to repeat that broken symmetry is only a single example of various mechanisms that are understood to give rise to emergent properties: the New Emergentists never claim that it is the whole story.

## 2.5 Alternative Readings

As we saw in §1.4, it is orthodox for philosophers to dismiss the New Emergentists as simply confused between practical, "epistemological" issues and "ontological" issues. It has been one of my aims to call into question these quick dismissals, and to try to suggest some possible sources of the misunderstanding. But there are also some more careful metaphysical analyses of their writings, and none of them (as far as I am aware) has suggested any connection between New Emergentist claims, and the difficulty of searching for good coordinates. So, it would be remiss not to look at some of these interpretations, and contrast them with the one presented here.

[29] A development that Anderson foresaw. He justified his inclusion of a discussion in his 1984 book by suggesting that 'the renormalization group has the potential of providing an even more complete conceptual unification of the science of complex systems than has yet been realized.' (Anderson, 1984, 5).

### 2.5.1 Liu — Infinite Systems mark Emergence

Here we follow an observation of Liu (1999), that many theories of condensed matter physics require the infinite limit of a large system to be taken. Liu concludes that when such phenomena appear in *finite* systems, they should be called emergent, since even a calculator with the abilities of Broad's Archangel would miss them.[30] For such a being would have no reason to consider an infinite system, nor any reason to apply an analysis of an infinite system to the finite ones under consideration. Thus Liu defends the 'emergence' in question as ontological rather than epistemological, since these symmetry-broken properties are not derivable from microphysical theories, even in principle.

Liu's interpretation has points in its favour. It explains the New Emergentists' fascination with symmetry-breaking, and it is supported to some extent by Anderson's own comments. We have already quoted Anderson as identifying 'our own macroscopic scale' with the '$N \to \infty$ limit of large systems' (§1.4.2). And when discussing the practical difficulties in deriving the behaviour of large symmetry-broken states, he emphasises that it would be impossible to use the microphysical laws, since:

> Starting with the fundamental laws and a computer we would have to do two impossible things — solve a problem with infinitely many bodies, and then apply the result to a finite system — before we synthesized this behaviour. (Anderson, 1972, 395)

He returns to the topic in his later book, where he points out some major confusions that can result if we do not take account of the discontinuities and 'attempt to calculate the properties of one state of matter by methods suitable to another.' Anderson (1984, 29).

[30] In fact, Liu worked with a conception of emergence as a failure of inter-theoretic reduction, but his suggestion is not dependent on this, and can be easily adapted to circumvent the ambiguities discussed in §1.7.

But all of these are strictly methodological points, concerned with the proper theoretical treatment of different phases. For any systemic properties of a state, even those associated with phase transitions and symmetry-breaking, *are* derivable in a finite model, at least if we substitute Broad's Archangel for Anderson's computer. We have no reason to think that such a being would need to take a diversion through the $N \to \infty$ limit, even if we do. That is, it would not have to run through the two impossible stages that Anderson mentions to derive the behaviour of such a system in any particular configuration. Instead, it would "just" have to run through a boring (and practically impossible) mathematical task of combining all the microphysical interactions.

It is true that Broad's Archangel would not discover those theoretical features that *we* use to model phenomena such as phase transitions and symmetry breaking. For rather than considering a realistic, finite system, theoreticians look to its infinite analogue and define phase transitions in terms of features that only appear there. They do this for reasons of mathematical power and convenience, as well as to isolate the large scale 'bulk effects' from spurious finite size effects.

In Chapter 4 I will expand on these themes, and on the larger debate on the nature of the idealisations involved in an appeal to the infinite limit. There, I shall argue that phenomena such as phase transitions, though requiring the infinite limit for their theoretical treatment, can nonetheless be recognised and defined as properties exhibited by finite systems. I suggest there that although some care needs to be taken due to the startling nature of the infinite idealisation, this is a matter of defining the relationship between physical phase transitions and their theoretical treatment. This is entirely separate from the metaphysical question of whether the associated properties are emergent.

### 2.5.2 Schweber — Emergent Properties are Properties of Solutions

Following a distinguished career as a quantum field theorist, Silvan Schweber has carried out extensive research into the history of field theory and the development of the Standard Model. In the course of his work, he wrote an article expressing the view that the New Reductionism/Emergentism debate was a symptom of a crisis amongst physicists. They had yet to confront conceptual developments that 'revealed a hierarchical structure of the physical world. Each layer of the hierarchy is successfully represented while remaining largely decoupled from other layers' (Schweber, 1993, 35). He feels that Anderson's main message is the following:

> ... it is not enough to know the "fundamental" laws at a given level. It is the *solutions* to equations, not the equations themselves, that provide a mathematical description of the physical phenomena. "Emergence" refers to properties of the solutions — in particular, the properties that are not readily apparent from the equations. (Schweber, 1993, 36)

This thought is certainly present in some of the early New Emergentist work, though it is defended more explicitly by other authors. Michael Polanyi, for example, emphasises that boundary conditions are needed to provide solutions, as well as the 'fundamental' laws, and that this provides a powerful barrier to the scope of 'unfettered reductionism'.

> A Laplacean knowledge which merely predicts what will happen under *any given* conditions cannot tell us what conditions should be given; these conditions are determined by the technical skill and peculiar interests of chemists and hence cannot be worked out on paper. Therefore, while quantum mechanics can explain in principle all chemical reactions, it cannot replace, even in principle, our knowledge of chemistry. (Polanyi, 1964)

However, it is a general feature of dynamical theories that boundary conditions are needed as well as the dynamical laws if we are to produce a full description

of the physical phenomena. Some solutions exhibit properties which are easily derived from the equations, others are not so easily derived. The whole point in coming up with a definition of emergence is to propose some way of distinguishing between trivial and non-trivial examples. Schweber and Polyanyi's conception of emergence provides no way of distinguishing 'trivial' solutions, whose properties are apparent from the laws, and 'non-trivial' ones.

In fact, Schweber seems to concede this point. Late in his article he interprets Anderson's title *More is Different* as the claim that there is novelty at each level of composition, which comes about in such a way that effective laws of properties at each level may be considered 'fundamental'. He suggests that the challenge is to conceptualise this novelty. I agree that this is indeed the challenge, but that Anderson has at least an example of how it may be achieved: namely: in his example of symmetry-breaking.

### 2.5.3 Cat — Intertheoretic Reduction

The issues addressed by Anderson in 1972 are deep, and it is startling to discover how differently careful interpreters can read his text. While Jordi Cat (1998) admits that Schweber's interpretation is in 'contrast' to his own, their discussions could almost be of two different articles.

In a long historical study of the circumstances surrounding the debate between Anderson and Weinberg, Cat offers an insightful analysis of the changes in attitude in different branches of physics that occurred at the end of the twentieth century. He reads Anderson as holding two separate positions: one a thesis about the methodology of physics, which he judges to be fairly clear; and one a set of metaphysical and ontological claims, which he argues is less so.

As to the former, he reads Anderson as recommending that we take very seriously the various viewpoints provided by the special sciences, on the grounds that

'the proliferation of new descriptions and their empirical adequacy would lead to potentially more fruitful lines of research' (Cat, 1998, 264). Particle physics is doubtless a creative and interesting subject, but new ideas crop up in the higher-level sciences: condensed matter physics, chemistry, biology and further up in scale. Cat claims that Anderson's *More is Different* is in part a critique of Kuhn's thesis of incommensurability, holding that cross-fertilisation between areas of science is a source not of misinterpretations, but of creativity and progress. In support, he points out that Anderson explicitly addresses Kuhnian themes in his article attacking large particle accelerator projects (Anderson, 1971, 515).

This reading of Anderson as putting forward heuristic recommendations to generate creativity in research, looks a good distance away from the main themes we have drawn from his discussion. I feel that Cat's interpretation focusses on a single point raised by Anderson as an argument for spreading government resources amongst areas in physics, rather than concentrating them to build a single expensive particle accelerator. But these are issues specifically to do with funding and status in the physics community, and are separate from his position on emergence.

Cat's discussion of Anderson's more explicitly metaphysical claims leads him to share with other commentators, the conclusion that 'Anderson's ontological reductionism is not clear' though he does recognise that his argument 'rests on one explicit ontological assumption, namely that composite systems present emergent properties.' But in developing his reading, Cat makes an interesting claim about where Anderson locates 'truly fundamental' laws and properties of physics:

> Anderson is ultimately concerned with the sort of epistemological relevance involved in "law-enforcement", a representation relation between levels of description and reality, between the most elementary laws and their concepts, on the one hand, and more complex systems and their properties, on the other. (Cat, 1998, 265)

Cat goes on to argue that these principles of "law-enforcement" have an element

of what he calls 'necessity', or 'non-derivativeness' in them, in that they are at least partially independent of the microphysical details of the system in question. He argues that this independence is sufficient to imbue the principles with a privileged status. 'Being "necessary", they cannot be derivative. It is precisely this unavoidable character that imbues them with the sort of fundamentality that Anderson stresses' (Cat, 1998, 265). He then goes on to mention symmetry-breaking as an example of such a set of 'fundamental principles'.[31] But regardless of his specific examples, Cat's suggestion is clear: it is not any set of microphysical properties that are fundamental or have a privileged intensional status. Rather, principles that constitute emergent mechanisms and bring about the appearance of emergent properties that are "fundamental". Symmetry-breaking is presumably amongst these.

This theme which Cat draws from Anderson is an extremely interesting one, but I cannot anywhere find Anderson explicitly defending such principles of 'law enforcement', as replacing microphysics in a privileged position. Instead, he repeats over and again that he judges the laws, principles and properties of each new emerging level as equally important. It is the role of principles and theories such as symmetry-breaking to show how novelty can come about at each level, not to constitute some alternative fundamental level themselves.

However, it is at least in the spirit of the New Emergentist movement to claim that 'higher-level principles' are as good candidates for some privileged status as are any other. Principles such as symmetry-breaking crop up across very different areas of physics: in the Standard Model of particle physics, in many-body quantum mechanics and in large-scale classical and celestial mechanics. Such principles that apply across physics of very different scales, give a sense of unity to physics, which

[31] Cat may be referring to 'protected' properties, despite the fact that Anderson does not mention them in *More is Different*, and they come to the fore only in later New Emergentist writings. I postpone discussion of this to Chapter 5.

implies no prejudice in favour of microphysics.[32]

Indeed, in a recent popular book Laughlin (2005) explicitly proposed that *all* presently known physical laws will turn out to arise due to 'large-scale collective principles' (amongst which he includes symmetry-breaking). He holds as a consequence that those entities currently treated as elementary, such as electrons, the other "elementary" particles, space and time will all turn out to have further structure.

But this position goes well beyond the core New Emergentist position. For as reported in §1.1.2, the New Emergentists have been willing to set aside this issue of entity reductionism, and grant the possibility that an ultimate micro-level may exist, governed by laws which could not arise due to the interaction of smaller parts. Cat's thesis may find some support from Laughlin's hypothesis: but both are separate, and far stronger doctrines than the central arguments of the New Emergentists.

## 2.6 Concluding Remarks

In the first chapter, I argued for two main conclusions. First: the New Emergentists have a consistent, but oft-misinterpreted position. I compared it to three well-established philosophical approaches to emergence, and argued that they shared

[32]Weinberg (1994) contains an interesting hint that he might be moving to embrace this alternative picture of the unity of science. For he distinguishes between 'Grand' and 'Petty' reductionism, where the former is the only one worth defending, it being: 'the view that all of nature is the way it is (with certain qualifications about initial conditions and historical accidents) because of simple universal laws, to which all other scientific laws may in some sense be reduced'. There is no mention in this statement that the laws must be those of the very small. Indeed, this is identified more easily with 'Petty reductionism ... the much less interesting doctrine that things behave the way they do because of the properties of their constituents'. And later in the same article (28), he again speaks of his arrows of explanation, repeating his earlier assertion that they end up at a common source, but with no suggestion that this source is the laws of the very small. Comparing these statements with, e.g., Weinberg (1987) suggests a substantial shift in his position.

many claims with the British Emergentists. But the philosophically orthodox reasons for rejecting their conclusions are valid only if we interpret them as claiming one of the other positions, and some commentators are also guilty of confusing the three. But we saw that in both accepting microphysical supervenience, and rejecting the existence of specific laws of composition, the New Emergentists do not take the exact position of Broad and his followers.[33]

In this chapter, we have explored the possibility that the New Emergentists are concerned with metaphysically significant distinctions amongst systemic properties. And while Broad's specific laws of composition are not recognised in modern physics, there is a close relative which could be said to be endemic. Thus we have arrived at an interpretation of the New Emergentist position which I believe metaphysically consistent, physically informed and to be faithful to their claims for the strong methodological differences between the various "levels" of physical science.

We have left various loose ends. One is the possibility of other mechanisms for emergence apart from symmetry-breaking. The only one that I shall discuss in this thesis is the phenomenon called "protection" by Laughlin and Pines. The rough idea is that while the systemic properties of larger systems are sometimes qualitatively "different" to those of smaller ones, they also exhibit eerie similarities across these larger systems, irrespective of their respective microphysical constitutions. I have already made a few remarks about these issues, which come to the fore in the Anderson-Weinberg debate and later writings of the other New Emergentists, but a full discussion requires the technical apparatus of the Renormalisation Group. I shall return to these and related issues after the discussion of renormalisation in Chapter 3.

[33]In fact, the reasons for their rejection decompose neatly into the two possibilities considered by Broad. As outlined in Appendix A, emergence$_1$ is excluded by accepting microphysical supervenience, and emergence$_2$ is excluded by rejecting specific laws of composition

Second, we have endeavoured to find a position for the New Emergentists while committing them as little as possible to specific metaphysical doctrines. We have found no outright contradiction while viewing matters from our rather high level of generality; but it is possible that if we attempted to construct a detailed metaphysical position, we might run into problems. In particular, we have remained almost silent on issues relating to laws of nature. The strictly Humean position of Lewis is the only one that has been mentioned, and that amounts a direct rejection of the New Emergentist position. Naturally, there are many others available, but we have not examined how a necessitarian view — say — would cohere with the rest of the New Emergentist doctrines.

Finally, the whole project has been to try to interpret and defend a rather narrowly-defined position, based on some very specific physical examples. Yet Anderson has been cited as inspiring a small intellectual revolution which has given rise to a renaissance in the physics of larger, more complex systems. (Indeed the very term 'complex systems' has become the label for some of the most fashionable research.) Do any of the interpretative issues we have been concerned with relate to these exciting areas? I shall take up these issues in Chapter 5. But whether there are uniting themes to this 'New Kind of Science', in particular, whether the mechanisms of symmetry-breaking have the claimed implications for biology and chemistry is, as always, down to the ultimate success of the physicists, not to the dictates of metaphysicians.

# Chapter 3

# Universality, Multiple Realisability and Renormalisation

Over the last decades, renormalisation techniques have become central to many areas of moden physics, a development not yet matched by attention from philosophers. Following the historico-philosophical study by Cao and Schweber (1993) and Teller's ground-breaking book (1995) the commentary that does exist, has been centred on the significance of the Renormalisation Group (henceforth, RG) to quantum field theories.[1]

Perhaps surprisingly, there has been less attention paid to the area in which the modern understanding of the RG was originally developed, viz. in condensed matter physics. Yet it may be that the philosophical significance of the technique is as striking in this context as when it is applied to quantum field theories. An exception to this lack of philosophical attention has been the work of Robert Batterman, who has made several bold claims about the significance of renormalisation as a strategy for explanation. In this chapter, I shall discuss and criticise Batterman's

[1] A valuable cross-section of the current philosophical debate is available in the collection of Kuhlmann et al. (2002), and other notable commentaries have come from Huggett and Weingard (1995), Hartmann (2001) and Castellani (2002).

claims, using them as a foil to bring out what I feel are the most philosophically interesting aspects of renormalisation techniques.

The discussion is structured as follows: In the first section I will introduce Batterman's theses about renormalisation; in particular, his claims that it provides a strikingly original explanation of the phenomenon that physicists call *universality*, and that similar strategies might explain the metaphysicians' problem of *multiple realisability*. Batterman's discussion is based upon a commitment to firm criteria by which an account could provide a satisfactory explanation of universality. The explanatory strategy of RG is supposed to be original in being able to satisfy these criteria.

To examine these claims in more detail, I introduce the work-horse example of condensed matter physics in §3.2: the Ising Model, and discuss how the phenomenon of universality is exhibited there. In §3.3, I look at two older approaches to explaining universality, and show how they fail to satisfy Batterman's explanatory criteria — as indeed he claims they should. In §3.4 and §3.5, I distinguish two varieties of renormalisation: the first one used by most authors (including Batterman) to explain the concept in qualitative terms; but the second a more significant advance in explaining universality. I argue that neither of them can satisfy Batterman's criteria for a satisfactory explanation of universality. In §3.6, I identify several different attitudes to the significance of the RG, and suggest that they are focussing on the significance of its applications and aims, not on the technique itself. Finally, in §3.7, I show that clarity with regard to the points highlighted in the preceding sections is crucial to evaluate the significance of renormalisation.

## 3.1 Introduction — Batterman's Thesis

Batterman holds that renormalisation exemplifies a form of scientific explanation unappreciated by philosophers of science, and that its differences from standard accounts are important for a variety of philosophical issues, including inter-theoretic reduction, emergence and the nature of scientific explanation. He puts his case in several articles (Batterman, 1992, 2000, 2002b, 2004) and in his book: *The Devil in the Details* (2002a).

Each of Batterman's theses are stimulating and controversial, inspiring commentary and criticism from several authors.[2] Perhaps most comprehensive is the survey given by Hooker (2004), who dissects a wide variety of his claims, especially focussing on those which impact on approaches to inter-theoretic reduction. My range will be a great deal narrower; I want to focus on a very specific set of his claims about the significance of the RG, which — if valid — would have profound implications for debates central to metaphysics and the philosophy of mind.

Batterman argues that the phenomenon much-discussed in metaphysical circles as *multiple realisability*, is best understood as an instance of a phenomenon closely studied in physics, where it is known as *universality*.[3] Batterman goes on to claim that the RG explains the physicist's universality with a profoundly original strategy that has not been considered by metaphysicians or philosophers of science, and that a similar strategy can be reasonably expected to explain all instances of multiple realisability.

[2]Belot (2005) and Redhead (2004) both concentrate their criticism on Batterman's conclusions about inter-theoretic reduction, questioning the significance that Batterman sees in singular limits. Batterman (2005) gives a response to these and related issues.

[3]Batterman writes as though it is clear that universality is the broader notion, encompassing multiple realisability, so that 'the phenomenon of multiple realisability is itself an instance of universality' (Batterman, 2000, 116-117). It is not clear to me that the former encompasses the latter, rather than vice versa. But in the present discussion, we can ignore the issue, for we will concentrate only on the domain of critical phenomena, where the two are claimed to be coextensive.

The existence of a profoundly new account of multiple realisability might have vast implications. For over thirty years, the "Multiple Realisability Thesis" — that homogeneous macroscopic properties can be realised by heterogeneous microscopic properties — has played a prominent role in debates in the philosophy of mind and science. It is often identified as the principal argument in favour of a functionalist attitude to mental properties (e.g., Putnam, 1975a) and invoked for debates on the autonomy of the special sciences (e.g., Fodor, 1974). Claims with such striking implications should invite close scrutiny.

Batterman's case that the strategy of renormalisation can help explain multiple realisability can be split into two claims, here paraphrased from Batterman (2000, 116-7):

(i) The phenomenon of multiple realisability is itself an instance of universality.

(ii) The explanatory strategy of the Renormalisation Group (though likely not the very same explanatory apparatus) can be extended to explain the metaphysical mysteries surrounding multiple realisability.

Hooker's commentary dissects each of these claims, arguing with respect to (i): that on the contrary, universality only captures part of the phenomenon of multiple realisability; and therefore to (ii): that any explanatory strategy capable of accounting for universality cannot provide a full account of multiple realisability (though he grants that it may go some of the way).

Though I am sympathetic to many of Hooker's claims, I shall not challenge Batterman's claims (i) & (ii) directly. Instead, I shall challenge two presuppositions that lie behind them:

(1) The Renormalisation Group explains the phenomenon of universality.

(2) The explanatory strategy of the Renormalisation Group is an original one,

> profoundly different to the old and more familiar attempts to explain either multiple realisability or universality.

Batterman defends both of these claims, going into some detail as to the explanatory success of renormalisation, and making explicit proposals as to those aspects in which the strategy is original. But before looking in detail at the explanation, we must first take a look at the phenomena that are to be explained: multiple realisability and universality.

### 3.1.1 Introduction to Multiple Realisability

The principal reason that multiple realisability is of interest in the philosophy of mind, is the claim that any satisfactory account of mental properties must allow that they can be multiply realised by physical properties. This claim is often put at the basis of an argument for a functionalist theory of mental properties.[4] But it is helpful to start by considering simpler cases that do not immediately involve us in controversies as to the nature of the mental. Such a "clean" introduction is provided by the famous example of the mineral jade, which can be realised by two compounds: jadeite and nephrite, each a distinct chemical kind.[5]

But we shall not be concerned with arguments that are built upon the multiple realisability thesis, but with the thesis itself. For, as Jerry Fodor puts it, whatever conclusions we draw from the existence of multiple realisability, the phenomenon itself presents 'a metaphysical mystery':

> Damn near everything we know about the world suggests that unimaginably complicated to-ings and fro-ings of bits and pieces at the extreme

[4]The most prominent of these arguments are those of Putnam in a series of papers through the 1960s, reprinted in his 1975a collection. Other influential examples of arguments based upon multiple realisability include those of Kim (1992), who argues that multiply realised properties cannot be true natural kinds.

[5]In order to continue to speak of properties rather than entities, we shall treat the example as that of the higher-level mineralogical property: "being jade", multiply realised by more than one lower-level chemical property: "being jadeite" and "being nephrite".

> *micro*level manage somehow to converge on stable *macrolevel* properties. (Fodor, 1997, 161)

Fodor confesses that the phenomenon of a truly heterogenous set of microlevel properties giving rise to a stable homogeneous set of macrolevel properties is one he does not even 'know how to *think* about'. Batterman suggests that physicists *do* know how to think about it, for multiple realisability is an instance of the physicist's phenomenon of universality, and Fodor's mystery can be explained by the Renormalisation Group.

Before moving to see how multiple realisability might be related to universality, we must mention that many commentators do not think it as mysterious as do Kim, Fodor and Batterman. Sceptics argue that most, or even all, proposed instances of multiple realisability fail to concern a truly homogeneous set of higher-level properties realised by a truly heterogeneous set of lower-level ones. As such, they can be explained by strategies that are in no way exotic or original. I shall mention just two such criticisms here: they are particularly relevant to universality, but by no means exhaust the concerns that have been raised about multiple realisability.

First, it has been pointed out that certain forms of multiple realisability are ubiquitous, so common that we hardly notice them. Crane and Mellor (1990) see the phenomenon everywhere, noting that properties such as 'masses, volumes, and temperatures' can be realised by almost any system, implying a vast range of microphysical realisers. Such widespread examples have led some to write of a 'trivial' sense of multiple realisability, when lower-level realiser properties are distinct only due to differences manifestly irrelevant to the higher-level properties they realise. Some, like Bechtel and Mundale (1999) argue that all instances of multiple realisability are trivial in this sense, and arise from confounding of 'broad-grained'[6] and 'fine-grained' criteria for distinguishing properties. Others, such as

[6]'Coarse-grained' is a more usual term than 'broad-grained', but I shall follow Bechtel and Mundale's terminology.

Shapiro (2000) grant that there may be both trivial and non-trivial cases and search for criteria to separate them.

Shapiro puts forward a causal criterion, suggesting that multiple realisability is non-trivial only when realisers differ in an aspect 'causally relevant' to the aspects that define the higher-level properties. So a case in which pain were realised in both ordinary neurons and in neurons that were stained purple, Shapiro would judge as trivial, since colour is not causally relevant to the higher-level property of pain. Shapiro doubts that many famous cases pass his test, arguing that the realisation of pain both in neurons and in silicon is also trivial, since the causally relevant aspects of electrical activity are present in both cases.[7] And the jade example seems vulnerable also. For we can argue that when we pick out the mineralogical properties, we abstract from chemical differences that are causally irrelevant to that level. At the lower chemical level we do not, and so judge that there are two properties instantiated.

Somewhat related to this first distinction is the observation that there is more than one way in which a property might be said to be multiply realised. First, we could look at the model provided by the determinate-determinable relation, and identify a sense in which a determinable property can be multiply realised by its determinate instances. In this sense, the property of blueness can be said to be multiply realised by the properties of being various shades of blue: aquamarine, azure, turquoise or cornflower. But there is a second sense in which even a *maximally specified* shade of blue may be realised in different ways. A shade of cobalt blue could be realised by the element cobalt, by cobalt-coloured paint or by a pixel in a computer screen — all exactly that same shade of cobalt. Once we have made this distinction, it is clear that the former 'determinate-determinable' sense should

[7] Shapiro's specific case-study rests upon patterns of electrical activity being the causal ground of pain, and common to both silicon and neural realisation, and it thus looks vulnerable to strengthened examples; but his distinction stands nonetheless.

also be counted a 'trivial' sense of multiple realisability since there is no mystery of how to explain it. We simply point out that we employ different standards of accuracy to distinguish properties on the higher and lower level, so it is not surprising that we see heterogeneous properties on one level and homogeneity on another.

The explanations of these 'trivial' cases of multiple realisability could be called 'trivial' themselves, for they involve discovering a sense in which the lower-level realisers were not heterogeneous after all. Batterman is clear that we are concerned with a sense of multiple realisability, and an explanation of it, that is not 'trivial' in either of the senses mentioned here.

### 3.1.2 Introduction to Universal Properties: Critical Exponents

The most striking examples of universal properties are observed near a certain type of phase transition, known as second-order, continuous or critical transitions. These *critical phenomena* can be observed near the Curie point of a ferromagnet.

Consider a large sample of ferromagnetic material with average magnetisation $M$ per unit volume. In an experiment or simulation, we usually control two independent variables: a temperature $T$ and an applied external magnetic field $H$, both of which affect the magnetisation. Typical results are represented qualitatively in Figure 3.1.

It is clear from the diagram that the magnetisation $M$ of a ferromagnet may change with $H$ in very different ways, depending on its temperature $T$. If we keep it at a low temperature $T < T_c$ and vary $H$, from positive to negative values, we observe a discontinuous 'jump' in $M$ at $H = 0$, as the ordered state switches from a magnetisation in one direction to another. This is an example of a *first order*

phase transition;[8] other familiar examples include the abrupt liquid/vapour transition, boiling, where the density jumps suddenly with a small change of applied temperature or pressure. These transitions are usually associated with a radical reordering of the structure of the material and the transfer of some non-zero quantity of latent heat. But while they are both theoretically interesting and challenging, there are no universal properties associated with first-order phase transitions.

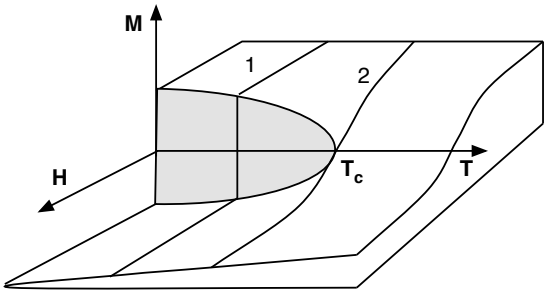



Figure 3.1: Phase diagram for a Ferromagnet. Path 1, with $T < T_c$ exhibits a first-order phase transition, Path 2, goes through a continuous phase transition at $H = 0$, $T = T_c$ (often called the *Curie* temperature).

At $T \geq T_c$ the first order phase transition disappears, so at the Curie temperature $T = T_c$ and zero external field, we say that a *continuous* phase transition occurs, and the system is at its *critical point*.[9] The behaviour of the thermodynamic variables near the critical point is strikingly different from elsewhere. If we define a *reduced temperature* $t = \frac{T-T_c}{T_c}$ (the dimensionless temperature difference from $T_c$) we find that the magnetization $M$ goes as a power law with $t$, with a certain *critical exponent* $\beta$. $M$ also goes as a power law when $t = 0$ and $H$ is

[8]The classification of phase transitions is troublesome. The 'Ehrenfest' classification and its descendants are outlined in §4.1.1, but briefly: we call a phase transition 'first order' if the free energy is discontinuous in the first derivative of the free energy with respect to any thermodynamic variable, and 'continuous' if the non-analyticity is of any other form. Further classification is usually made on the basis of symmetry considerations.

[9]This example also exhibits symmetry-breaking at the critical point. Symmetry breaking, a central theme of §sec:breaking, is closely associated with critical phenomena but I shall not discuss it further here.

varied, with a different critical exponent $\delta$. Other thermodynamic properties such as the magnetic susceptibility at constant temperature: $\chi = \frac{\partial M}{\partial H}\big|_T$, and the heat capacity at constant field: $C_H = T\frac{\partial S}{\partial T}\big|_H$, also exhibit power-law behaviour in the region and have the exponents $\alpha$ and $\gamma$ respectively. These four critical exponents are defined as follows:

$$C_H \sim |t|^{\alpha} \tag{3.1}$$

$$M(T, H \to 0^+) \sim |t|^{\beta} \qquad \text{(for } T < T_c\text{)} \tag{3.2}$$

$$\chi_T(T, H \to 0^+) \sim |t|^{-\gamma} \tag{3.3}$$

$$M(T = T_c, H) \sim H^{\frac{1}{\delta}}. \tag{3.4}$$

The values of these exponents are found to stand in certain simple relations to one another, the Rushbrooke and Griffiths Equalities:[10]

$$\alpha + 2\beta + \gamma = 2 \qquad \text{(Rushbrooke)} \tag{3.5}$$

$$\delta - 1 = \frac{\gamma}{\beta} \qquad \text{(Griffiths)} \tag{3.6}$$

It is striking that similar behaviour appears in other systems that appear to share very little structure with ferromagnets. A clear example comes from experimental studies of liquids as they make the transition through their fluid phase. Here the independent variables are the pressure $P$ and temperature $T$, which determine the density $\rho$ of the fluid at equilibrium. Again, there is a first order phase transition (boiling) which disappears at a continuous phase transition, where $P = P_c$ and $T = T_c$, and the thermodynamic variables of the fluid again display power-law behaviour. Consider thermodynamic quantities: again $C_V = T\frac{\partial S}{\partial T}$, the heat capac-

[10]Rushbrooke (1963) and Griffiths (1965) themselves first postulated *in*equalities on purely thermodynamical grounds. The limiting equalities of Equations 3.5 & 3.6 come from experimental evidence, and later, the scaling hypotheses introduced in §3.2.2.

ity; $\rho_l$, the liquid density; $\rho_g$, the gas density; and $\kappa_T = \frac{1}{V}\frac{\partial V}{\partial P}$, the compressibility. Compare these critical exponents defined for a fluid system with Equations 3.1 - 3.4:

$$C_V \sim |t|^{\alpha} \tag{3.7}$$

$$(\rho_l - \rho_g)(T, P \to P_C) \sim |t|^{\beta} \qquad \text{(for } T < T_c\text{)} \tag{3.8}$$

$$\kappa(T, P \to P_C) \sim |t|^{-\gamma} \tag{3.9}$$

$$P - P_C(T = T_c, \rho) \sim |\rho_l - \rho_g|^{\frac{1}{\delta}} \tag{3.10}$$

The truly astonishing experimental discovery was that the values of the critical exponents $\alpha$, $\beta$, $\gamma$ and $\delta$ are numerically identical in the case of the fluid and the ferromagnet. What is more, these are not the only systems for which these values appear. In the formation of long-chain polymers, in percolating systems, and in the critical behaviour of beta-brass, the systems all exhibit a continuous phase transition, and near it, their analogous macroscopic quantities also obey power laws, with exactly the same numerical values of the critical exponents.

Because the nature of the transition is very different in each case, these large-scale properties can be thermodynamic, mechanical, magnetic, electrical, or chemical; but still, their quantitative behaviour near phase transitions shows this remarkable similarity — the values of the exponents are identical. Yet there is no suggestion that the microphysical structures of these various systems have much in common with one another; and their behaviour away from $T_c$ is as different as could be expected. We call these critical exponents *universal properties*.

Further investigation suggests that all systems that exhibit continuous phase transitions fall into one of a small number of *universality classes*, defined by sharing the same values of the critical exponents. Each universality class may contain systems as different from one another as fluids, magnets or percolating systems,

which appear to share none of their behaviour away from their critical points, nor anything much in their microscopic constitution (so far as we know). We find that all they appear to have in common is their spatial dimensionality, and some general symmetry properties associated with the phases either side of the phase transition (more exactly, the symmetry of the order parameter of Landau, introduced in §3.3.2).[11]

### 3.1.3 Multiple Realisability as Universality

To understand how Batterman connects the phenomena of multiple realisability and universality, we must be clear as to exactly which of the properties associated with phase transitions are to count as the higher-level, macrophysical properties, i.e. as candidates for properties to be multiply realised, and which as lower-level properties, which are to realise them. In these central cases, this seems fairly clear: those dependent on the experimental variables are the higher-level set: $M$, $T$ and $H$ for the ferromagnet; $P$, $T$, and $\rho$ for the fluid. The lower-level properties are the microscopic details: the very different microphysical structure of the fluid and ferromagnet.

But it is only a few of the higher-level properties that are candidates for multiple realisability. Specifically, the universal critical exponents appear to be multiply realised; for example, the higher-level property of having a critical exponent $\beta = 0.31$ is realised by the heterogeneous microstructures both of fluids and of ferromagnets, as well as the microstructures of the other substances in that universality class. (More exactly, we should say that the lower-level *properties* of these microstructures realise the higher-level properties). But away from the critical point, wide differences in the lower-level properties are reflected in wide differences in higher-

[11] Lavis and Bell (1999, 22) follow Kadanoff (1976) in hedging their bets, quoting three properties that may influence the universality class a system falls into: a) the physical dimension of the system, b) the symmetry group of the order parameter, c) perhaps other criteria.

level properties. And even near the critical point, there are a large number of higher-level properties that are not universal and so not multiply realised. For example, the critical temperature $T_c$ is different for each system within a universality class.

Let us test universality as a case of 'true' multiple realisability, against the two 'trivial' senses highlighted in §3.1.1. It is certainly difficult to see it as trivial in the first sense: that the different realisations of critical exponents within a universality class are distinguished only by differences manifestly irrelevant to the higher level. In particular, Shapiro's criterion of causality makes this look untenable, for the only properties in common between the various microstructures are dimensionality and symmetry, which surely cannot be enough to provide any meaningful causal basis for the critical exponents. Just as challenging would be to follow Bechtel and Mundale, for we would have to explain why (almost) every microphysical difference between fluids and magnets would be irrelevant to the higher-level properties, and thus that our apparent instance of multiple realisability is nothing more than a mismatch between grades of discrimination between realised and realiser.

To claim that the cases are trivial in the second sense would be to claim that are dealing with a simple case of a determinate-determinable relation. And this seems just as implausible as the first sense. For the value of the critical exponent appears to be maximally determinate — a specification of its exact value is limited only by experimental technique (or lack of computational power when it comes to numerical simulations).

One further comment should be made, before we move to consider Batterman's thesis. The appeal to 'microphysical' or 'lower-level' properties might lead us to recall the distinctions drawn back in the first chapter, where we distinguished several different senses of these terms. The ambiguity most relevant to the present situation is between what we called there microphysical$_1$ and microphysical$_2$ properties,

for these give us two different versions of the Multiple Realisability thesis. In the first sense — microphysical$_1$ — the thesis would be that two properties, described as heterogeneous by a microphysical theory, each would provide a supervenience base for a single macrophysical property, as described by a macrophysical theory. But in its second sense — microphysical$_2$ — the thesis would be the following, very different claim: that there are large systems which have properties identical (or at least very similar) to one another, but whose respective parts have very different properties when alone, or when formed into different systems than those two in question.

In the discussions of Putnam, Kim and Fodor, we are clearly meant to think of multiple realisability in the first sense. For they are concerned with issues of inter-theoretic reduction, and especially with the autonomy of special science theories from one another. And in §1.5.1, we saw that it was microphysical$_1$ properties that was associated with these questions of reduction. Our examples bear this out, for it is not just the microphysical$_2$ properties of a fluid and a ferromagnet that will be heterogeneous, but the microphysical$_1$ properties as well. Fluids and magnets have a diverse microstructure, and different interaction potentials between their component parts. It is not just that their components are heterogeneous while separated, they interact according to entirely different laws and have different properties when within the larger systems as well. So, we need not be concerned with the ambiguous sense of 'microphysical'; in this chapter I shall use the terms 'lower-level' and 'microphysical' to mean microphysical$_1$, without any further specification.

So let us grant that judged on these considerations, the universal properties exhibited by values of critical exponents appear as non-trivial examples of multiple realisability, and thus as among the most interesting cases of Fodor's 'metaphysical mystery'. Let us now move to look at Batterman's alleged 'new explanatory

strategy' that physicists have developed to explain the universality of critical exponents.

### 3.1.4 Desiderata for a New Explanatory Strategy

Batterman holds that the explanatory strategy of the Renormalisation Group (henceforth RG) is a new one to both philosophers of science and to metaphysicians. In fact, he believes that the mainstream models of scientific explanation are powerless to explain universality. Therefore, it is not surprising that the RG provides a mode of explanation new to the philosophy of science, for any satisfactory explanation of universality *must* be new in this sense.

However, any claims of limits to the potential of scientific explanation are sure to be controversial, and indeed Batterman's argument for the inadequacy of the existing models appears to be far from conclusive. I shall argue that his views on the original nature of the explanatory strategy of renormalisation can be separated from his (controversial) views and criticisms of existing strategies.

Nonetheless, it is worth reporting Batterman's objections to two classes of orthodox explanatory strategies, since they help illuminate his criteria for a good explanation of universality.[12] He follows a fairly standard practice in discussions of scientific explanation by dividing contemporary approaches into *Causal-Mechanical* and *Unificatory* accounts.

The former class include the explicitly causal approaches of Salmon (1984), as well as the probabilistic models such as Railton's 'D-N-P' account (Railton, 1981). They share the aim of giving what Batterman calls a "local" account: 'they explain particular phenomena in terms of collections of particular causal processes and interactions — or, perhaps, in terms of noncausal mechanisms, if there are such

[12]The brief remarks below are taken from (Batterman, 2002a, Ch. 3). Further discussion of Hempelian deductive-nomological and inductive-statistical accounts is given in Batterman (1992).

things.' (Salmon, 1989, 182). And Batterman's problem with Causal-Mechanical accounts is simply this "local" aspect, for he argues there will be a mismatch between what a causal-mechanical approach can explain — a particular instance, or event; and the explanandum of universality, which is by its nature, a uniformity across many instances. The approach of the Causal-Mechanical theorist must be to examine every system within the universality class and identify some microphysical mechanism by which the critical indices take the values that they do — i.e., they will give a separate account for each microphysical realiser. Since these realisers are, by hypothesis, heterogeneous, we cannot explain why they are the same across the many different systems by giving a Causal-Mechanical account — this must be left a mystery.

This leads Batterman to one of his criteria for a satisfactory explanation of universality: it cannot merely provide a list of lower-level properties that would be sufficient to realise the higher-level properties (2000, 135-6).[13] Instead, a satisfactory account such as RG, must 'depend upon a demonstration that from the point of the lower level theory many of the microstructurally relevant details are irrelevant for the upper level phenomenology of interest' (ibid, 136).

This criterion also precludes the situation in which the list of lower level properties can be shown to have something in common, as in the first of the 'trivial' senses of Multiple Realisability, in §3.1.1. An example of an explanation that Batterman dismisses on these grounds is the account of multiple realisations of temperatures by their shared mean kinetic energy — a lower-level property they have in common (ibid, 134-135). But this must be qualified: Batterman accepts Ned Block's 'Disney Principle': the lower-level theory must impose *some* constraints on what can

[13]In the context of his discussion, Batterman is concerned with opposing a very specific view, expressed in (Papineau, 1993, Ch.2), which demands a list of lower-level properties sufficient *and* necessary to realise a higher level one. His defense is the failure of sufficiency, so that is the criterion I have focussed on here.

realise a higher-level property.[14] Indeed, he feels that this principle stands in need of 'explanation and justification' (119), but this explanation cannot be provided simply by listing the elements we find in all the realisers.

Batterman's objection to Unificatory accounts (Kitcher, 1989; Friedman, 1974) is a little harder to pin down. A unificatory explanation is one which shows how the phenomena to be explained fall under a unifying, or covering law, and so can be shown to be specific instances of a more general physical process. Batterman argues that any unifying law that could cover all the instances of universality would amount to a stipulation that the regularity must occur. This law would not be an explanation, but simply a restatement of the fact of universality. Batterman holds that any unificatory account leaves out the explanation for *why* such a statistical law holds; this can only be given in terms of a 'fundamental theory'. While it is not entirely clear as to what qualifies as a 'fundamental theory' in general, Batterman gives explicit criteria for in the specific case of critical phenomena. He demands an account of the higher-level regularity, made in terms of the lower-level microphysical realisers, and it is just this account that he argues the unificatory accounts cannot provide (Batterman, 2002a, 30-36). This leads Batterman to his second criterion: a satisfactory explanation for universality must be 'completely grounded in the lower level theory' (Batterman, 2000, 136).

Thus Batterman looks on the explanatory strategy provided by the RG as being original, and distinct from orthodox explanations of multiple realisability, in that it succeeds in fulfilling the following two desiderata, both taken from Batterman (2000, 136):

[14]The name is an odd one. Block explains it as follows: 'In Walt Disney movies, teacups think and talk, but in the real world, anything that can do those things needs more structure than a teacup. We might call this the Disney Principle: that laws of nature impose constraints on ways of making something that satisfies a certain description.' (Block, 1997, 123). Since the principle asserts that such constraints of microphysical detail *do* exist, the name 'Anti-Disney Principle' seems more appropriate.

(A) The explanation must be 'completely grounded in the lower-level theory'.

(B) The explanation is not dependent on identifying a list of microphysical properties sufficient to realise the higher-level ones.

It is these two conditions that we shall focus on in what follows: we leave aside any further differences from orthodox accounts of explanation. In the next section, we will look at a particular condensed matter system which exhibits universality, and then look at how both pre-RG and RG approaches aim to explain the phenomenon, concentrating on the ways in which they might contravene one or the other of Batterman's conditions.

## 3.2 The Ising Model

The Ising Model has been called the *Drosophila* of condensed matter physics: a simple and well understood case on which the phenomena associated with phase transitions can be studied. More properly, it should perhaps be named the Lenz-Ising model, since Lenz proposed the lattice of two-state spins as a model of ferromagnets in 1920. His student Ernst Ising solved the one-dimensional version in his doctoral dissertation, finding that it does not exhibit a phase transition, but remains in a paramagnetic phase (except at $T = 0$). He conjectured that this would also hold true in higher dimensions. With such a dismal outlook, the model then fell into obscurity, until Rudolf Peierls realised that Ising's conjecture might be mistaken. He gave an argument that seemed to show that transitions would occur in two dimensions and higher.[15] In 1944 Lars Onsager analysed the two-dimensional Ising model in the absence of an external field, and managed to solve it exactly (i.e., he provided an analytic expression for the partition function) and

[15]Later, some minor flaws in Peierls' argument were found and the argument was made rigorous by Griffiths (1964).

showed explicity that it had a finite temperature phase transition. To date, the three dimensional model has resisted all attempts to discover an exact solution.

Where the model cannot be solved exactly, there are a wide variety of techniques available to give approximate results. Series expansions about the high and low temperature limits are surprisingly powerful; and Monte Carlo and other numerical methods can be implemented on powerful computers, giving results that are believed accurate to several decimal places.

Despite its simplicity, the Ising model provides a rather versatile framework that can be applied to many physical systems, with little or no modification: as well as the original purpose of modelling (suitably anisotropic) ferromagnets, we can interpret the spins to represent occupied and unoccupied sites in lattice gas models and configurations of atoms in binary alloys, such as beta-brass.[16] Since experimental data on these systems are available, and they also easily modelled on computers, the theoretical results can be tested against a variety of benchmarks.

### 3.2.1 The Model

Consider a large lattice of sites, indexed by an integer variable $i$. On each of there is defined a variable we call a *spin*, which can take one of two values: $s_i = \pm 1$. Each interacts pairwise with its neighbours through an exchange interaction $J^2_{ij}$, triplewise with the interaction $J^3_{ijk}$, etc. There may also be an external field, $J^1_i$, that couples to all the spins individually. The Hamiltonian of such a system can be written as follows:

$$-\mathcal{H} = \sum_i J_i s_i + \sum_{ij} J^2_{ij} s_i s_j + \sum_{ijk} J^3_{ijk} s_i s_j s_k + ... \tag{3.11}$$

[16]It should be emphasised that an Ising model is not versatile in the sense that it could be reinterpreted as a model of all systems in its universality class, and thus provide an alternative explanation for universality.

which allows for interactions between an external field and each spin, interactions between pairs, between triples and so on. In what follows we shall concentrate on a well-known special case, and the one that Ising himself studied — the *nearest-neighbour* Ising Model. Here, the interactions are non-zero only between adjacent pairs of spins, so using the more common notation $J \equiv J_{ij}^2$ and $H \equiv J_i$, the Hamiltonian can be written:

$$-\mathcal{H} = J \sum_{<ij>} s_i s_j + H \sum_i s_i \tag{3.12}$$

where $< ij >$ means the sum is taken over nearest neighbours, and $J$ is positive so as to favour parallel alignment of spins. However, we will sometimes find it useful in what follows to retain Equation 3.11, since the RG technique often generates effective Hamiltonians of a more general form than Equation 3.12.

We assume the validity of orthodox classical Gibbsian statistical mechanics, so the thermodynamical variables and associated quantities can all be derived from the canonical partition function:

$$\mathcal{Z}(T, H) = \sum_r e^{-\beta \mathcal{H}} \tag{3.13}$$

where the sum is taken over all the possible spin configurations, each of which give a different value for the Hamiltonian $\mathcal{H}$. We also define the free energy:

$$\mathcal{F}(T, H) = -\frac{1}{\beta} \ln \mathcal{Z}(T, H) \tag{3.14}$$

which we extremise to find any thermodynamically stable states. It is possible to calculate macroscopic quantities such as magnetisation, specific heat capacity, and magnetic susceptibility from $\mathcal{Z}$ or $\mathcal{F}$ and their derivatives, and we shall introduce these quantities as needed.

As well as these thermodynamical variables, we can also consider further properties associated with the Ising system. These may be obtained from *correlation functions* on the spins. For example, the spin-spin correlation function measures the correlation between spins on sites $i$ and $j$ (with position vectors $\mathbf{r}_i$ and $\mathbf{r}_j$) away from their mean value:

$$\Gamma(\mathbf{r}_i, \mathbf{r}_j) = \langle (s_i - \langle s_i \rangle)(s_j - \langle s_j \rangle) \rangle \,. \tag{3.15}$$

(Here, the triangular brackets indicate a mean taken over the spin configurations.) For a translationally invariant system, $\langle s_i \rangle = \langle s_j \rangle \equiv \langle s \rangle$, and $\Gamma(\mathbf{r}_i, \mathbf{r}_j)$ depends only on the separation $r$ between the spins $i$ and $j$. That is, we can show that:

$$\Gamma(r) = \langle s_i s_j \rangle - \langle s \rangle \,. \tag{3.16}$$

For the Ising model with short-range interactions (a class which includes, but is not limited to, nearest-neighbour models) it can be shown that the spins become uncorrelated as $r \to \infty$, the function $\Gamma(r)$ dropping off exponentially with distance. From this property we can define one of the most important properties for the discussion that follows, the *correlation length*, $\xi$:

$$\Gamma(r) \sim r^{-\tau} e^{-\frac{r}{\xi}} \tag{3.17}$$

(where $\tau$ is a constant that becomes unimportant as $r$ becomes large). Near the critical point it can be shown that both $\xi$ and $\Gamma$ also go as power laws, and define two additional critical exponents $\nu$ and $\eta$. They also display universality, within

the same universality classes as did $\alpha$, $\beta$, $\gamma$ and $\delta$.

$$\xi \sim t^{-\nu} \text{ , as } t \to 0; \tag{3.18}$$

$$\Gamma(r) \sim \frac{1}{r^{d-2+\eta}}. \tag{3.19}$$

As usually defined (see note 8) a phase transition is a discontinuity in the free energy or any of its derivatives with respect to the thermodynamical variables. At first sight it is difficult to see how such a discontinuity could occur in the Ising model as defined here, since any finite sum of analytic functions is also analytic, and looking at the sums involved in constructing $\mathcal{Z}$ and $\mathcal{F}$, it appears there is nowhere a discontinuity could occur. We shall look at how this "paradox" is resolved in Chapter 4, but briefly: the usual method of studying phase transitions in statistical mechanics is to take the *thermodynamic limit*. This is an idealisation which carefully extends the number of spins to the infinite limit, while keeping relevant quantities such as their volume density at a constant value. The subtlety is that a discontinuity *can* appear in an infinite sum of analytic functions. The use of the thermodynamic limit brings up various philosophical worries, but I shall postpone consideration of these to the later discussion. Here it should just be noted that each of the approaches we shall meet, both RG and non-RG, make use of it.

So consider a large Ising system with average magnetisation $M$ per unit volume. ($M$ may be derived directly from the microphysical spins: $M = \frac{1}{V}\sum_{i=1}^{N} s_i$, but can also be defined from the free energy: $M = -\frac{1}{V}\left(\frac{\partial \mathcal{F}}{\partial H}\right)_T$.) We can then define the rest of the critical exponents, in the same way as we did in §3.1.2.

To make contact with the issue of multiple realisability, it is again important to separate the microphysical from the macrophysical properties associated with the Ising Model. The interaction Hamiltonian $\mathcal{H}$ specifies a detailed model in terms of

its microphysical properties and the function encodes all of the microphysical interactions involved in the system. This model and its associated coupling interactions corresponds to Batterman's talk of a 'lower-level theory' — a mathematical expression governing the microphysical properties directly. In contrast, the partition function and free energy are obtained from weighted sums over all possible spin configurations. The thermodynamic quantities (i.e.: all those derivable from the partition function $\mathcal{Z}$) may be taken as corresponding to Batterman's higher-level, macrophysical properties. Amongst these, the critical exponents form a further subset that are held to be *universal* and therefore multiply realised.[17]

### 3.2.2 The Widom Scaling Hypothesis

The variety of universal phenomena, as outlined in the last section, were originally discovered through experiment, along with the few isolated theoretical successes such as the Onsager solution. In response to the discoveries, a new framework of experimentally successful, but theoretically unmotivated conjectures began to grow up around the new results. One of the great triumphs of the RG methods was to derive these conjectures from physically plausible principles, and place them on a firm theoretical footing.

In the 1960s, several workers[18] independently suggested that the universality of critical phenomena might arise if, as they neared the critical point, the thermodynamic functions for disparate systems took on a particular form, that of a *generalized homogeneous function*. One way of putting the conjecture is to propose that the singular part of the free energy per spin, call it $g_s$, obeys the following

[17] It may be that all the macrophysical quantities are multiply realisable in one or more of the 'trivial' senses identified in §3.1.1. What we are concerned with are those that are genuinely universal, and so guaranteed to be multiply realisable in an interesting sense.

[18] These included Widom, Domb, Kadanoff and Fisher. An authoritative review is given by Kadanoff (1966).

equation as the critical point is neared:

$$g_s(t, H) = t^{\frac{1}{y}} \psi \left( \frac{H}{t^{\frac{x}{y}}} \right) \tag{3.20}$$

where $x$ and $y$ are parameters to be determined. This is called the *Widom Scaling Hypothesis*, and the function $\psi$ is called a *scaling function*. From it, we can calculate the critical exponents in terms of $x$ and $y$. For example, near the critical point the zero-field magnetisation becomes:

$$m_{H=0} = -\left( \frac{\partial g_s}{\partial H} \right)_t = -\left( \frac{t^{\frac{1}{y}}}{t^{\frac{x}{y}}} \right) \psi'(0) \text{ , so } \beta = \frac{1-x}{y}. \tag{3.21}$$

Similarly:

$$\alpha = 2 - \frac{1}{y} \text{ , } \gamma = \frac{2x-1}{y} \text{ and } \delta = \frac{x}{1-x}. \tag{3.22}$$

Eliminating the parameters $x$ and $y$ then leads immediately to the Rushbrooke and Griffiths' equalities.[19]

On its own, the Widom scaling hypothesis is not even close to providing a satisfactory explanation of universality. It is a hypothesis stated wholly in terms of the thermodynamic quantities, with no basis in microphysical details, no explanation of why the scaling form should appear near the critical point, and no indication of how to derive values for $x$ and $y$. In practical terms, this means that we neither have predictions for the values of the critical exponents, nor any indication of why they should be the same for all realisers in a universality class. So even if we accept Widom's scaling hypothesis, it does not give any account of

[19]There are two further equalities associated with critical exponents, the Fisher and Josephson laws, which involve the exponents $\eta$ and $\nu$. They can be derived from a compatible but distinct scaling hypothesis which imposes a form on the correlation function (Kadanoff, 1966). But in this chapter, I shall focus on the Widom hypothesis.

universality — it takes it as given, and uses it to make further conjectures about the macrophysical properties of near-critical systems.

That said, the Widom hypothesis might provide a step towards an explanation of universality, since it transforms our original problem to a different one. If we can explain why near-critical systems might have thermodynamic functions of the form of Equation 3.20, and why those belonging to the same universality class share values of the parameters $x$ and $y$, we shall have secured an explanation for universality.

## 3.3 Universality and Scaling — Two Classical Approaches

In this section, I shall present two theoretical approaches to critical phenomena, and discuss how they can be related to the Widom Scaling hypothesis. These are both *non*-RG, 'classical' approaches,[20] one based upon the idealisations of mean field theories, the other on a set of arguments due to Lev Landau.

Anachronisms abound, for both Landau's theory and the general ideas behind mean field theories had been around for several decades before Widom put forward his scaling hypothesis. Nevertheless, it is illuminating for our discussion to examine how classical approaches account for universality via scaling. For Batterman judges that no classical approach is able to provide a satisfactory explanation of universality. Thus, contrasting their deficiencies with renormalisation techniques, may give a clearer idea of what he holds to be original to the latter.

[20]In the context of theories of critical phenomena, 'classical' simply means an approach not based on renormalisation techniques; there is no implied contrast with quantum physics. For example: 'classical' mean field theories can be successfully applied to fully quantum-mechanical models. However, it must be noted that a surprising number of successful models are also 'classical' in the non-quantum sense, even when it is known that the underlying interactions must be quantum mechanical in nature.

We start with an example from the wide variety of mean field theories (henceforth MFTs). Far more sophisticated varieties are, even today, at the theoretical cutting edge, but a simple case suffices to exhibit the idealisations that are essential to the general mean field approach.

### 3.3.1 Mean Field Theory

It has been said that there are as many ways of approaching Mean Field Theories as there are authors writing on them, and this is probably a reasonable lower bound. The idea common to all is to model each spin as interacting with an external field chosen to summarise the effects of all the other spins that interact with it. So rather than attempting to calculate the partition function based on the original Ising Hamiltonian (Equation 3.12), we replace it with a *mean field* form, which couples each spin to a field term, which may be as simple or as intricate as the theoretician wishes. We shall consider a MFT for the Ising model with a fairly transparent physical motivation.

#### The Bogoliubov Inequality

We start from the following inequality, which follows from a result known as the *Bogoliubov theorem*:[21]

$$\mathcal{F} \leq \mathcal{F}_0 + \langle \mathcal{H} - \mathcal{H}_0 \rangle_0 \tag{3.23}$$

where $\mathcal{F}$ is the true free energy of the system with Hamiltonian $\mathcal{H}$, $\mathcal{H}_0$ is a trial Hamiltonian, $\mathcal{F}_0$ its free energy, and $\langle x \rangle_0 = tr\left[xe^{-\beta\mathcal{H}_0}\right]/tr\left[e^{-\beta\mathcal{H}_0}\right]$, the ground state expectation value of $x$. Now for any Hamiltonian of a given form, Equation 3.23 can be used to obtain the best possible estimate of the true free energy. For

[21] A proof of the inequality is given in most textbooks on condensed matter physics. For example, see McComb (2004, 163).

we simply write a trial Hamiltonian in the fixed form, with variable parameters, and then adjust them to minimise the right-hand side of Equation 3.23.

So consider again the Ising model, which has the Hamiltonian:[22]

$$\mathcal{H} = -J \sum_{<ij>} s_i s_j \tag{3.24}$$

and consider a trial Hamiltonian of the very simple form:

$$\mathcal{H}_0 = -H_0 \sum_i s_i \tag{3.25}$$

where the adjustable parameter $H_0$ gives the strength of the *mean field*. This gives us:

$$\mathcal{F}_0 = -NkT \ln(2 \cosh \beta H_0) \tag{3.26}$$

$$\langle s \rangle_0 = \tanh \beta H_0 \tag{3.27}$$

$$\langle \mathcal{H} - \mathcal{H}_0 \rangle_0 = -J \sum_{<ij>} \langle s_i \rangle_0 \langle s_j \rangle_0 + H_0 \sum_i \langle s_i \rangle_0 \tag{3.28}$$

For a translationally invariant system $\langle s_i \rangle_0 = \langle s_j \rangle_0 \equiv \langle s \rangle_0$. Substituting these into Equation 3.23 and minimising with respect to $H_0$ gives us:

$$H_0 = Jz \tanh \beta H_0 \tag{3.29}$$

$$\langle s \rangle_0 = \tanh \beta J z \langle s \rangle_0 \tag{3.30}$$

$$\mathcal{F}_{mf} = -NkT \ln(2 \cosh \beta J z \langle s \rangle_0) + \frac{NJz}{2} \langle s \rangle_0^2 \tag{3.31}$$

where $z$ is the number of directly interacting spins (so $z = 2d$ for a $d$-dimensional, nearest-neighbour Ising model on a square lattice). In some respects, these three

[22] For simplicity, we consider a model without an external magnetic field.

expressions appear good qualitative descriptions for the Ising model. For above a certain temperature the equations have only one solution $\langle s \rangle_0 = 0$, but below it, another pair of finite solutions for $\langle s \rangle_0$ appears. We identify this temperature with the critical temperature $T_c$ and the two new solutions as representing the two ordered, magnetised states. But the equations also lead to an expression for the critical temperature: $kT_c = Jz$, which is erroneous: and not only quantitatively, for it implies a finite temperature phase transition for the one-dimensional model, the non-existence of which was demonstrated by Ising himself.

### Mean Field Exponents

If we ignore these deficiencies then it is simple to use the framework of MFTs to calculate the critical exponents. For example, we can define the reduced temperature $t \equiv \frac{T - T_c}{T_c}$ and Equation 3.30 becomes:

$$\langle s \rangle_0 = \tanh \frac{\langle s \rangle_0}{1 + t}. \tag{3.32}$$

Expanding for small values of $\langle s \rangle_0$ and $t$ gives:

$$\langle s \rangle_0 = \langle s \rangle_0 (1 - t) - \frac{\langle s \rangle_0^3}{3} + O(\langle s \rangle_0 t^2, \langle s \rangle_0^3 t, \langle s \rangle_0^5) \tag{3.33}$$

Assuming that the system is sufficiently close to the critical point, we can ignore the higher terms and obtain:

$$\langle s \rangle_0 \sim (-t)^{\frac{1}{2}}. \tag{3.34}$$

This gives a value for one of the critical exponents, $\beta = \frac{1}{2}$. The others can be obtained similarly,[23] and we derive the following values for the whole set: $\alpha = 0$, $\beta = \frac{1}{2}$, $\gamma = 1$, $\delta = 3$, $\nu = \frac{1}{2}$, $\eta = 0$. I omit the full details of the derivation, but the important point to realise is that these values are foisted upon us as soon as we adopt a mean-field form for the Hamiltonian. That is, the exact details of the mean field $\mathcal{H}_0$ is irrelevant — our choice to use the idealisations inherent to mean field theory, fixes those values for the exponents.

### The Ginzburg Criterion

The most important idealisation involved in MFTs comes with imposing a particular form of Hamiltonian: that of Equation 3.25. This is a fairly radical simplification, replacing all the interactions between the different spins, by a coupling to a single field. Choosing a mean field value that saturates the Bogoliubov inequality will mean that we select the best possible value of $H_0$, but we must recognise that it may not be good enough. The form of Equation 3.25 can be too restrictive if the fluctuations away from the mean field contribute significantly to the free energy.

We can use MFT itself to give a self-consistent estimate as to the circumstances under which which this will occur — an estimate known as the *Ginzburg Criterion*. Again, I shall not give a full derivation here,[24] but since the energy of a typical fluctuation is of order $kT$, and its size is determined by the correlation length, we can estimate the size of the free energy due to the fluctuations to be:

$$\mathcal{F}_{fluc} \sim \frac{kT}{\xi^d} \tag{3.35}$$

[23] To derive some further exponents ($\eta$ and $\nu$, in particular), it is necessary to use a rather more sophisticated version of MFT, which allows magnetisation to vary smoothly with position. See for example (McComb, 2004, 171-176).

[24] Once again, proofs can be found in most textbooks, e.g., Chaikin and Lubensky (1995, 214-217).

which we need to be small compared to $\mathcal{F}_{mf}$. Near the critical point, as $t \to 0$ we can use the MFT values of the critical exponents to estimate when this will be so:

$$\mathcal{F}_{mf} \sim |t|^{2-\alpha} \text{ and } \xi \sim |t|^{-\nu} \text{ , so } d\nu > 2 - \alpha \tag{3.36}$$

With the MFT values $\nu = \frac{1}{2}$ and $\alpha = 0$, this implies that MFT will fail in less than four dimensions, since the correlation length $\xi$ will diverge near the critical point and the fluctuations away from mean field values can no longer be neglected. But the Ginzburg Criterion is just a test of self-consistency: it neither implies that MFT will be empirically successful while it remains self-consistent, nor that it will be unsuccessful even when it is inconsistent. Besides, as we see in the following section, there was an independent line of argument that seemed to suggest that — at least as regards the critical exponents — the MFT values were correct for systems of every dimensionality.

### 3.3.2 Landau Theory

The Landau theory of phase transitions was one of most successful approaches to critical phenomena until the advent of renormalisation methods.[25] It is based on very plausible, general principles, which makes it all the more surprising to find that it is in conflict with the exact results of Onsager, and with experiment.

Landau's starting point was to identify an *order parameter* for each system, a quantity chosen to characterise the collective order of the state.[26] For the Ising model, and in real ferromagnets, the appropriate choice is the magnetisation $M$: zero in the disordered high-temperature states, and rising smoothly to non-zero

[25]The canonical exposition of Landau's theory is in the textbook of Landau and Lifshitz (1959), where various refinements are also discussed.

[26]I discuss order parameters in a little more detail in §2.4. The original discussions are in Landau and Lifshitz (1959, 446-516), and illuminating ideas of how to formalise and extend the concept into areas where it is less clearly defined, are discussed in Anderson (1989, 16-19).

values as the system makes the transition to ordered states. Landau's idea was that if we restricted ourselves to spin configurations with a given large-scale order (a given *phase*), the free energy would be analytic in $M$. It is only if we average over a set of configurations belonging to more than one phase that we encounter the singularities associated with a phase transition. He assumed that close to $T = T_c$ it is possible to Taylor-expand the free energy $\mathcal{F}(T, M)$ in powers of $M$.

$$\mathcal{F}(T, M) = a_0(T) + a_1(T)M^2 + a_2(T)M^4 + ... \tag{3.37}$$

where the odd powers vanish, due to an assumed symmetry under the flip transformation $M \to -M$. We impose the condition that $a_2(T)$ must always be positive, since otherwise the free energy would fall with increasing $M$, giving an unstable system.[27] The critical point is then recognised by solving for the stationary point of $\mathcal{F}$, and by demanding non-trivial solutions, (i.e., that $a_2(T)$ and $a_1(T)$ have opposite signs):

$$\frac{\partial \mathcal{F}}{\partial M} = 0 = 2a_1(T)M + 4a_2M^3 \tag{3.38}$$

so that, neglecting the trivial solutions:

$$M = -\sqrt{-\frac{a_1}{2a_2}}. \tag{3.39}$$

[27] Recall that a minimum in the free energy gives thermodynamically stable states.

Assuming that $a_1$ and $a_2$ are also analytic in $T$, we obtain to a first approximation that $a_1 = At$, and $a_2 = B$ where $A$ and $B$ are constants, and $t$ is the reduced temperature. Substituting this back into Equation 3.37 gives us the non trivial minima:

$$M_{\pm} = \pm\sqrt{\frac{B}{2C}t} \text{ , for } T < T_C; \tag{3.40}$$

and by comparing this to Equation 3.2 we can see that the critical exponent $\beta$ is predicted to be $\frac{1}{2}$. A similar analysis lets us read off the *Landau values* of all the critical exponents as: $\alpha = 0$, $\beta = \frac{1}{2}$, $\gamma = 1$, $\delta = 3$, i.e., exactly the same as we obtained from Mean Field Theory.[28] So Landau theory also predicts universality. Indeed like MFT it goes too far: not only are values of the critical exponents independent of the microphysical details, they do not even depend on the dimensionality of the system.

But in the case of MFT, the Ginzburg Criterion has cautioned us that the technique was not to be trusted for systems with $d \leq 4$. Here there are no obvious caveats or alternatives to Landau's analysis that would give rise to different values for the exponents.[29] It is possible to extend Landau's theory by allowing the order parameter to be a smoothly varying function in space, which gives a far more realistic picture of the situation (recall the exponential drop-off of the correlation function we saw in §3.2.1), but this still does not change the values of the exponents, nor does it make them dependent on the dimensionality of the system.

It was therefore a great theoretical puzzle that the exponents observed in experiment, and those obtained from the Onsager exact solution, agreed in showing

[28] Again, we need to extend this simplest version of Landau's theory in order to obtain predictions for $\nu$ and $\eta$, but we shall not consider these elaborations here.

[29] Indeed Goldenfeld (1992, §7) provides a discussion of how dimensional analysis suggests that the Landau values for the critical exponents are the only ones possible.

| Values of selected Critical Exponents | | | | |
|---|---|---|---|---|
| Exponent | Variable | Landau Values | Ising $d = 2$ | Ising $d = 3$ |
| $\alpha$ | Heat Capacity | 0 | 0 | 0.12 |
| $\beta$ | Order Parameter | $\frac{1}{2}$ | $\frac{1}{8}$ | 0.31 |
| $\gamma$ | Response Function | 1 | $\frac{7}{4}$ | 1.25 |
| $\delta$ | Critical Isotherm | 3 | 15 | 5.20 |
| $\nu$ | Correlation Length | $\frac{1}{2}$ | 1 | 0.64 |
| $\eta$ | Pair correlation at $T_c$ | 0 | $\frac{1}{4}$ | 0.06 |

Table 3.1: A selection of critical exponents, comparing the Landau values, with values from the Ising Model in two and three dimensions. For $d = 2$, the values are taken from Onsager's exact solution. For $d = 3$ they are obtained by numerical simulation.

significant deviations from the Landau values. In fact, the Ginzburg Criterion of MFT turns out to be a remarkably good guide to their range of validity. Above the $d = 4$ *upper critical dimension*, numerical simulations of Ising models start giving critical exponents in good agreement with the Landau values. But below it, they diverge sharply from Landau's predictions.[30] In particular, experimental values show a strong dependence on symmetry and dimensionality which neither Landau theory nor MFT can account for (see Table 3.1). Thus, one could even say that these approaches fail by predicting *too much* universality, for the results suggest that the same values of the critical exponents will appear in an even wider range of systems than actually do realise them.

But it is worth focussing on the fact that both of the classical approaches *do* give correct predictions above four dimensions. There, the critical exponents cease to have a dependence on dimensionality, and the Landau values become completely universal. Thus, Landau and Mean Field theory are by no means lacking in empirical success, albeit under a limited set of circumstances. It appears

[30]For certain cases, such as tricritical points and other varieties of ferromagnet, the upper critical dimension can be three or lower, and Landau's values for the critical exponents can be observed in the real world.

rather bad luck that the dimensions we are most interested in (two and three) are not within this domain.

### 3.3.3 A Classical Explanation for Universality?

The partial empirical success of the classical theories in predicting universality might lead us to ask whether either MFT or Landau theory could count as a satisfactory explanatory strategy for multiple realisability on Batterman's criteria. I shall discuss this question further in §§3.4 & 3.5 when we can make a comparison with the RG techniques, but I hope it is already clear that they are both deficient, though they fall down on different points.

The universality predicted by Landau's theory is very wide-ranging, but it arises only because we imposed a form for the free energy on all near-critical systems. This fixed form of free energy also determines the critical exponents, so a prediction of universality is hardly surprising. It cannot fulfil Batterman's criteria, since is is not an explanation 'completely grounded in the lower-level theory', as he demands. Rather, Landau's theory ignores the details of the Ising Hamiltonian altogether.

When we turn to the MFT approaches, we do take account of the microphysical Hamiltonian, but only insofar as we replace it with one of Mean Field form. In certain circumstances this replacement might be justified, in others it might not. But by imposing it in all cases, we are imposing properties in common between all the ostensibly heterogeneous lower-level sets of properties: each Hamiltonian is judged to be idealisable to a common mean-field form. And so we would contravene Batterman's restriction that we cannot explain universality by simply identifing a set of lower-level properties sufficient to realise the higher-level properties. Imposing a Hamiltonian of Mean Field form would amount to such a set of properties, since it is sufficient to give us critical indices with a single set of values.

To conclude: Landau's approach violates Batterman's criterion (A), since it looks to the higher-level properties (i.e., those derived from the free energy $\mathcal{F}$) and imposes a common form upon these, giving an explanation which is not wholly (nor even slightly) grounded in a lower-level theory. In contrast, Mean Field theory violates his criterion (B), since it replaces each of the microphysical Hamiltonians by one with a common form — a form sufficient to determine universal values for the critical exponents.

Of course, this failure is to be expected once we read Batterman's main claims, since he holds that classical approaches fail to explain universality in a satisfactory manner. It is the new strategies of renormalisation that he thinks do better, and we turn now to those.

## 3.4 Kadanoff's Approach: Real-Space Renormalisation

Soon after Widom postulated his scaling relations, Leo Kadanoff put forward a very original approach to explain how they might arise, and to calculate values of the critical exponents for the Ising model. His first paper on the subject (Kadanoff, 1966) is often seen as the first application of renormalisation techniques to critical phenomena, though not yet as part of a comprehensive theory of critical phenomena (nor did he link it to the techniques in field theory). Kadanoff suggested that one could think of the effects of several spins, as combined into 'blocks', each of which could be thought of having a single magnetic moment, and a simple nearest-neighbour interaction. He then showed that this new block-model could also be modelled as an Ising system, with a different effective temperature and magnetic field. By applying these ideas, he was able to derive the Widom Scaling Laws (as well as the further scaling laws of Fisher and Wilson).

Kadanoff's ideas contained much insight, but they were not entirely correct in their assumptions. With no apparent malice, Wilson described his scheme as 'absurd, though leading to generalizations that are not absurd' (Wilson, 1971b). The scheme's modern descendants are now called *real-space* renormalisation schemes, which vary in their sophistication and empirical success, but all share the spirit of Kadanoff's original ideas. While they are relatively easy to understand, and usually provide the pedagogical introduction to renormalisation, they are *not* always the best approach to give accurate predictions. Our physical example is again the two-dimensional nearest-neighbour Ising model, with the Hamiltonian given in Equation 3.12. (The following discussion draws on those of Maris and Kadanoff (1978), Cardy (1996, §3), Goldenfeld (1992, §9) and Yeomans (1992, §8)).

### 3.4.1 A Real-Space Recursion Relation

The central idea of the Kadanoff approach is that if we can apply some transformation that maps an original problem to another at a different scale, we might be able to compensate for this reduction by altering (renormalising) the parameters appearing in the Hamiltonian or Lagrangian of the system, in a way that preserves all its large-scale properties. In a real-space approach to the Ising model this approach can be implemented by taking the original lattice of spins, and *coarse-graining* it according to some rule, to produce a new lattice that contains fewer spins. We then rescale so that the the new lattice can be compared to the old, and change the interaction parameters, to reproduce the large-scale properties. A transformation that changes the interactions in such a way, is known as a *recursion relation*.

It is useful to consider *reduced interaction couplings* rather than the general interaction parameters of Equation 3.11: $J^1_i$, $J^2_{ij}$, $J^3_{ijk}$, etc. We define $K^a_{ijk\ldots} \equiv J^a_{ijk\ldots}/kT$, and denote the whole set as $[K]$. Also, though a system is usually

considered in the thermodynamic limit, it is conceptually useful to consider a finite system. We can then think of the recursion relation as relating our original problem involving $N$ spins to one involving $N' \equiv N/b^d$, thus reducing the degrees of freedom by a factor $b^d$, (where $b$ is the spatial scaling factor, and $d$ the dimensionality of the lattice) described by new spin variables $\{s'_I\}, I = 1, ..., N'$, and a new set of coupling constants $[K']$. We can write a *recursion relation* for the coupling constants:[31]

$$[K'] = \mathcal{R}_b[K] \text{ where } b > 1. \tag{3.41}$$

We can divide the action of a general recursion relation into three stages:

1. Reduce the degrees of freedom — in this case, collect the spins into 'block spins'.

2. Rescale — in this case, shrink all the lengths to recover the original lattice spacing.

3. Renormalise — readjust the interaction parameters on the new lattice so that the higher-level properties are the same as the original.

One possible procedure of implementing stage 1 is illustrated in Figure 3.2. It is a particularly crude reduction method that collects spins into blocks, by simply throwing out half of them, a procedure known as *decimation*. In the notation used above, we can implement the rule with the specifications: $s'_i \equiv s_{2i}$ and $N' \equiv \frac{N}{2}$. Another well-known procedure for reducing the degrees of freedom is known as *blocking*, where we take the majority "vote" of several spins, and replace the block with a single spin in the direction voted for (see Figure 3.3). More sophisticated procedures can involve taking a transformation rule with several adjustable parameters, and tuning the values of these parameters to match experiment.

[31]Despite being called the Real-Space Renormalisation *Group*, the transformation need not

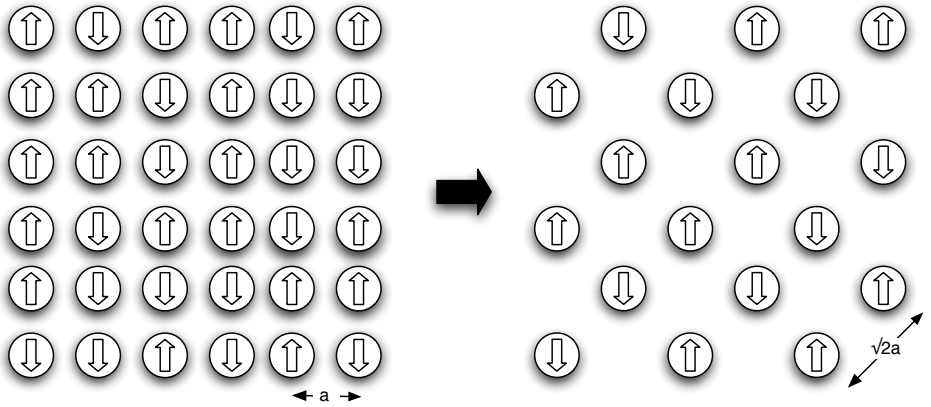



Figure 3.2: An arrangement of spins on a lattice, before and after a decimation transformation, which simply removes every other spin to produce a lattice which must be rescaled by a factor of $b = \sqrt{2}$ to be comparable to the original.

How we implement the rescaling of stage 2 then depends on the details of the first stage. For the decimation procedure illustrated in Figure 3.2 we just rescale all the lengths associated with the problem by a factor $b = \sqrt{2}$, since this is the factor by which all lengths have to be shrunk to reproduce the original lattice. For the blocking procedure of Figure 3.3, it is equally clear that $b = 3$. And the more sophisticated rules give rise to their own rescaling factors.

Stage 3 is usually the most difficult to implement. The aim is to transform the parameter values such that the large scale properties are preserved. As before, we take these to be the thermodynamic quantities obtainable from the partition function $Z$, and the free energy $\mathcal{F}$ (though it is easier to work with the free energy per spin $g$). So given the "block spins" $s'_i$ produced in the first step, we look for

have an inverse (since $b > 1$ always). However, it is transitive: $\mathcal{R}_{b_1 b_2}[K] = \mathcal{R}_{b_1} \cdot \mathcal{R}_{b_2}[K]$, and does have an identity $\mathcal{R}_1 = e$; so it is a *semi-group*, or *monoid*.

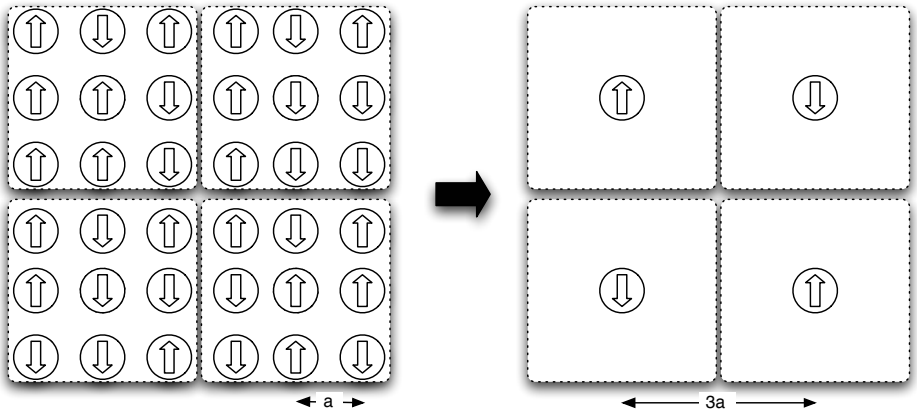



Figure 3.3: An arrangement of spins before and after a majority-rule transformation, which groups the spins into blocks of nine, prodoucing a lattice which must be scaled by a factor of $b = 3$ to be comparable to the original.

new parameter values $K'$ that preserve these functions, i.e. those such that:

$$Z_N[K] = Z_{N'}[K'] \tag{3.42}$$

$$g[K] = b^{-d} g[K'] \tag{3.43}$$

(the factor of $b^{-d}$ appears in the free energy per spin because of the reduction in the number of spins after the blocking). If we were to impose these conditions exactly, they would almost determine $K'$ in terms of $K$. For consider $P(\{s_i\})$, the probability that the original spins take the set of values $\{s_i\}$. This means that, considering the reduced Hamiltonian $\mathcal{H}$ as a function of these spins $\{s_i\}$:

$$P(\{s_i\}) = \frac{1}{Z} e^{-\mathcal{H}\{s_i\}} \tag{3.44}$$

But our chosen blocking rule has already told us how to obtain the set $\{s'_j\}$ from $\{s_i\}$, so the condition becomes:

$$P'(\{s_i\}) = P'(\{s'_j\}) = \frac{1}{Z'} e^{-\mathcal{H}'\{s'_j\}} \tag{3.45}$$

$$\text{where } Z' = \sum_{s'_j} e^{-\mathcal{H}'\{s'_j\}} \tag{3.46}$$

This constrains $\mathcal{H}$ to within an additive constant which can be fixed by some normalisation condition.

However, imposing the conditions of Equations 3.42 and 3.43 exactly must be seen as an ideal. Some very artificial models, such as the heirarchical Dyson model, allow it: given a suitable reduction rule, there is an effective Hamiltonian for the reduced system that preserves $Z$ exactly, and is of the same form as the original Hamiltonian. But this is not so for the Ising model, nor for any other realistic ones. If we are to faithfully reproduce the physics of the original lattice, our effective Hamiltonian would have to include not only nearest-neighbour interactions, but also the next-nearest neighbour, and introduce four-spin terms (since the nearest-neighbour interactions in the original lattice would have to be allowed for). This means that even if we start with the simple nearest-neighbour form of Equation 3.12, a single application of the recursion relation will immediately generate more complex terms, and repeated applications will quickly lead us to unmanageable Hamiltonians.

There are many different ways of making a compromise. We might cut off all interactions above a certain order, or design the recursion relation such that it does not generate higher-order terms. The design of real-space transformations can sometimes seem more than an art than a science. To quote Kadanoff: 'if we know the answer, we can always invent a renormalization transformation to find it'.

The question is whether these approximations that we are forced to make can be controlled and justified. The answer depends on both the model considered and the quantities we are attempting to calculate. And it is surprising how often reasonable results can be derived from fairly crude approximations. But whether the transformations are exact, or approximate, the crucial point is that the details of the procedure must be tailored for each model we consider. That is, to construct a real-space recursion relation we need to pay close attention to the microphysics of the system, and a recursion relation appropriate for one model will not in general be appropriate for another, and may give spurious results.

### 3.4.2 Parameter Space

Let us assume for the moment that an appropriate recursion relation has been constructed for our model. Now consider the high-dimensional space co-ordinatised by the set of parameters $[K]$. We call this a *parameter space* and can think of each point as directly representing a possible interaction Hamiltonian (though we can also think of each point as corresponding to the equilibrium state(s) of the system with that interaction Hamiltonian).

We can separate two mechanisms which would be represented by "motion" in parameter space. The more obvious one is an experimental operation by which we can physically alter some of the parameter values. The set of parameters $[K]$ represent both easily adjustable parameters such as temperature or external magnetic field, and less easily adjustable ones such as the the nearest-neighbour interaction strengths. When movement is along directions representing only changes the easily adjustable parameters, it is natural to associate the region covered by that movement with a single physical system. For example, a line could represent a single system at a range of temperatures; a plane, a system at different temperatures and magnetic fields. It is perhaps less natural to think of a single system in other

cases; for example, the most practical way of changing interaction strengths is to substitute a system composed of one material with another. However, we will consider an idealised experimental set-up and treat the two cases alike. As any parameter is varied, we can think of a single system as tracing out a path in the parameter space.

A second mechanism that could be represented as motion is the action of the recursion relation we considered in the last section. Each application of $\mathcal{R}_b$ maps a point $[K]$ to point $[K']$, thus moving from a point representing one Hamiltonian (and thus one equilibrium state) to another. Owing to the nature of the 'majority rule' and 'decimation' procedures, we have been considering discrete recursion relations; each application would be represented by a discrete 'hop' in the space. But it is usual to generalise the idea of recursion relations to include continuous versions, parameterised by the length rescaling parameter $b$. We call the paths traced by the action of the recursion relation, *flows* in the parameter space. Our conditions on designing an appropriate recursion relation should ensure that we shall never find ourselves leaving the original space (since the transformed Hamiltonian remains constrained to the original form), and that the large-scale physics is preserved along the whole flow (since the partition function and the free energy are suitably scaled and expressed in terms of new variables by equations 3.42 and 3.43).

If we think of each point in parameter space as representing an equilibrium state of the physical system with the corresponding interaction Hamiltonian, we expect that some will be tractable using the orthodox methods of statistical mechanics, and some will not. (In particular those systems which are near their critical points and have large values of $\xi$ are amongst the intractable). But tractable or not, we can be sure that if two states are related by a suitable recursion relation, their higher-level, thermodynamic properties will be the same. Kadanoff's original idea

was that if one could not analyse systems at, or near to their critical states, it might be possible to transform them to ones that were further away from criticality, and that could be analyzed by orthodox techniques.

But this is not quite how the modern procedure for calculating critical exponents works. Instead of looking directly at the systems related by the recursion relation, we look at the flows it induces. The *fixed points*: systems whose parameters are unchanged by $\mathcal{R}_b$, turn out to play a critical role (pun intended). At these points we have:

$$[K^*] = \mathcal{R}_b[K^*]. \tag{3.47}$$

Now consider the correlation length $\xi$ as a function of $[K]$. At a fixed point it will have a single definite value, since it is a function of $K^*$, but we also know that applying the recursion relation shrinks all lengths by a factor of $b$. So we can write:

$$\xi[K^*] = \frac{1}{b}\xi[K^*] \tag{3.48}$$

Therefore, $\xi[K^*] = 0$ or $\infty$. We call these *trivial* and *critical* fixed points, respectively. We assume that $\mathcal{R}$ is analytic[32] at $[K^*]$ and linearise parameter values close to this fixed point:

$$K_n = K_n^* + \delta K_n. \tag{3.49}$$

Applying the recursion relation $[K'] = \mathcal{R}_b[K]$ we obtain:

$$K_n' = K_n^* + \sum_m T_{nm}\delta K_m + O[(\delta K)^2], \text{ where } T_{nm} = \left.\frac{\partial K_n'}{\partial K_m}\right|_{K_m = K_m^*}. \tag{3.50}$$

[32]This is a strong assumption, since we know that the free energy itself is not analytic at the fixed point. See Wilson (1971a) for a discussion.

We now examine the eigenvectors and eigenvalues of this matrix $\mathbf{T}$. We can write each of the left eigenvalues[33] in terms of the length scaling factor $b$ to a certain power $y_i$. We call these left eigenvalues the *RG eigenvalues*:

$$\sum_n \phi^i_n T_{nm} = b^{y_i} \phi^i_m \tag{3.51}$$

where $\phi^i_m$ are the left eigenvectors. The sign of the exponents $y_i$ give us the stability properties of the fixed points. This can be seen by defining *scaling variables*: $u_i \equiv \sum_m \phi^i_m \delta K_m$, which are linear combinations of deviations away from the fixed point. Their significance comes from the fact that they transform linearly under $\mathcal{R}_b$ near the fixed point:

$$u'_i = \sum_m \phi^i_m \delta K'_m = \sum_{mn} \phi^i_m T_{mn} \delta K_n = \sum_n b^{y_i} \phi^i_n \delta K_n = b^{y_i} u_i. \tag{3.52}$$

It is useful to classify the eigenvectors according to the signs of their eigenvalues:

- $y_i > 0$: Iterating the RG increases $u_i$ away from the fixed point, and we have a *relevant* eigenvalue, direction and scaling variable.
- $y_i < 0$: Iterating the RG decreases $u_i$ towards the fixed point, and we have an *irrelevant* eigenvalue, direction and scaling variable.
- $y_i = 0$: We cannot tell from the linearised equations where $u_i$ will move, and we have a *marginal* eigenvalue, direction and scaling variable. Higher order terms may be needed to analyse these further.

The direction of the flows near the fixed point are determined by these distinctions. Let us call the set of initial parameters that flow to a particular fixed point, the *basin of attraction* of that point. (At present, we can only identify basins within

[33]The matrix $\mathbf{T}$ is not in general symmetric, so the left and right eigenvalues may not be the same. In the cases we are dealing with, the left eigenvalues will be real.

| Classification of Fixed Points (FP) | | | |
|---|---|---|---|
| Relevant | $\xi$ | Name | Physical Domain |
| 0 | 0 | Sink | Bulk Phase |
| 1 | 0 | Discontinuous FP | Plane of coexistence |
| 1 | 0 | Continuous FP | Bulk Phase |
| 2 | 0 | Triple Point | Triple Point |
| 2 | $\infty$ | Critical Fixed Point | Critical Manifold |
| $>2$ | $\infty$ | Multicritical point | Multicritical point |
| $>2$ | 0 | Multi-coexistence FP | Multi-coexistence FP |

Table 3.2: A selection of fixed points, and the phases associated with them (taken from Goldenfeld (1992))

the immediate region around each fixed point, but by linearity we expect the basins to extend outside these regions.)

There are two main properties by which we can categorise the fixed points (see Table 3.2). The first is their *codimension*, which is (here) defined as the number of relevant eigenvalues of the fixed point. This is important, because it determines the dimensionality of the space of nearby flows which will be "attracted" by that point. For example, if the fixed point has no relevant eigenvalues, it is said to have a *co-dimension* zero; these points are called *sinks*, since all nearby flows head towards them as the recursion relation is iterated. (This is what typically happens at high and low temperature limits.) The dimensionality of their basins of attraction is equal to the total dimensionality of the parameter space.

The significance of sinks can be understood by considering a nearby point falling within their basin of attraction. Since iteration of the recursion relation will bring any flow closer to the fixed point, changing its parameter values, but not the nature of its free energy or partition function, we might expect it to share the thermodynamic characteristics of that fixed point. And this is indeed the case (assuming that the fixed point is a simple attractor). Fixed points with codimension zero indicate *bulk phases* of a substance, and in a ferromagnetic system, the points

in these basins of attraction will be the directionally magnetised low-temperature phases, and the unmagnetised high-temperature phase. In a fluid system, they might include the gaseous phase and the liquid phase. Note that because of their codimensionality, these bulk phases are very easy to find, we do not need to 'tune' our parameter values finely in order to put the system in the basins of attraction of their fixed points — accordingly, these are the phases of systems that we will observe, unless we prepare our experiments very carefully indeed.

Consider a fixed point in parameter space with a non-zero number of relevant scaling directions. This number $n$ is the codimension of the fixed point, and our basin of attraction will be $(N-n)$-dimensional, where $N$ is the total number of parameters.[34] So we must 'tune' $n$ parameters to locate a state in the basin of attraction of a fixed point of codimension $n$. Accordingly, these states will be much harder to explore than those of a bulk phase. If these subspaces are not reachable by adjusting the experimentally controllable parameters, it is possible that the corresponding phases might not physically appear in certain systems.

Given that the high and low temperature phases of the ferromagnet both contain sinks, this suggests that there may be a 'watershed' in the parameter space between points that flow to the high and low temperature fixed points. And this is so; the "watershed" hypersurface contains a fixed point of codimension two. All flows within the surface are attracted to this fixed point, but those outside it end up in high or low temperature phases.

The second property useful for the categorisation of fixed points is the value of the correlation length $\xi$. As suggested in Equation 3.48, at a fixed point $\xi$ can only take the values 0 — a *trivial* fixed point, (such as high and low temperature) and $\infty$ — a *critical* fixed point. This is important, because as we saw in §3.3.1, classical methods tend to fail as $\xi$ becomes very large. But here, a word of warning: due

[34] We ignore the fact that in realistic cases, the parameter space may be infinite-dimensional.

to this failure of classical methods, the analysis of systems nearing critical points is particularly associated with renormalisation techniques. But this particularly fruitful application of RG techniques must be distinguished from the techniques themselves, and it is important to emphasise these techniques have no intrinsic restriction to critical phenomena. It is just that other methods tend to fail in those regions.[35]

At a critical fixed point, the correlation length $\xi$ is infinite. As suggested by Equation 3.48, $\xi$ must also be infinite at all points within the basin of attraction of such a fixed point, a sub-space which we call the *critical surface*. Our watershed hypersurface is such a critical surface, for its fixed point turns out to be a critical fixed point, and as such, all points within the surface have infinite correlation length, suggesting that they are associated with systems reaching their respective critical points.

### 3.4.3 Near-Critical Systems

Consider the three dimensional parameter space represented in Figure 3.4, which contains a fixed point $K^*$ with one relevant and two irrelevant eigenvalues. From the discussion of codimension in the last section, we see that we need to adjust only one parameter of the system to end up in this two dimensional critical surface. Let us assume that this is easily adjustable — temperature, for example — such that our system's states are represented by a path such as the one illustrated in the diagram, passing through the critical surface at some point $K_c$. It is here that the system reaches its critical point. (In contrast, the Ising model has two relevant eigenvalues, so we need to adjust two parameters, usually the external magnetic

[35] An example of applying RG techniques to non-critical phenomena is the Nienhuis-Nauenberg criterion, which can be used to determine the existence of first-order phase transitions (Nienhuis and Nauenberg, 1975). The low-temperature fixed point must display one eigenvalue of $b^d$ if a plane of coexistence is to appear (Goldenfeld, 1992, 268-9). The correlation length $\xi$ remains small throughout this analysis, but it is an RG analysis nonetheless.

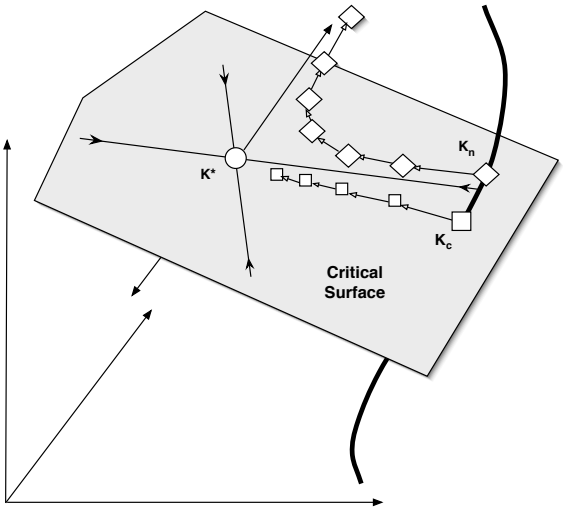



Figure 3.4: A critical surface in parameter space with a fixed point $K^*$, which has two irrelevant and one relevant eigenvector. The bold curved line represents a system passing through criticality at $K_c$, with the effect of four iterations of the recursion relation marked by the square boxes, remaining within the critical surface. $K_n$ is close to, but not on the critical surface, and the effect of six iterations of the RG are shown as the diamond-shaped boxes, at first approaching the critical point along the irrelevant eigenvector, and then veering away along the relevant one.

field and the temperature, to reach criticality).

First let us take the case of a system which comes to criticality in the immediate vicinity of $K^*$, where the linearised equation 3.50 is valid, and consider the effect of the recursion relation on the scaling variables. The distinction between the relevant and irrelevant eigenvalues becomes crucial, since as we iterate the recursion relation, the relevant components grow, while the irrelevant ones shrink. Any quantity that can be expressed in terms of the scaling variables will thereby become dominated by the small number of relevant ones as we approach the critical point.

Now consider a system at criticality, such as the one marked as $K_c$ in Figure 3.4, but far enough from the fixed point $K^*$ that the linearised equations are not

valid. From Equation 3.48, we can expect $\mathcal{R}$ to transform systems at criticality to systems at criticality, always keeping $\xi = \infty$, and thus remaining within the critical manifold. Both the system illustrated in Figure 3.4, and the simplest ferromagnetic systems have only a single fixed point within the critical surface, and so ignoring any strange limiting behaviour of the fixed point $K_c$ will flow towards $K^*$ by iteration of $\mathcal{R}$ and thus within the linearised regime.[36]

Most importantly, those systems that are *near*, but not at criticality also flow towards the fixed point, by continuity with those in the critical surface. If they are brought close enough to their critical points, the flow from states representing many realistic systems can be expected to pass within the linearised region near the fixed point. (Again, we ignore complications such as 'dangerous' irrelevant eigenvalues, cross-over behaviour, surface effects, etc.). Figure 3.4 shows this behaviour for our case of one relevant eigenvalue. $K_n$ is not within the linearised region of the fixed point nor is it actually within the critical surface, but the flow is still influenced by the critical fixed point. As shown in the diagram, under iterations of $\mathcal{R}$, the flow follows one of the irrelevant (attractive) eigenvectors to within the regime where the linearised equations become valid — then veers away along the relevant (repulsive) eigenvector.

### 3.4.4 Calculating the Critical Exponents

It is possible to use Kadanoff's insights to derive the Widom Scaling hypothesis for an Ising-like system with two scaling variables. We have to adjust two parameter values: the external magnetic field $h$ and the (reduced) temperature $t$, to arrive at the critical point, and so we expect there to be two relevant scaling variables, which

[36]The simplifying assumptions of there being only one fixed point within the critical surface, and the exclusion of odd limiting behaviour, disqualifies most of the more fascinating work on critical phenomena: cross-over effects and 'dangerous' irrelevant variables are just two important effects which are examples of exactly such behaviour.

we identify as relating to each of these parameters. From Equation 3.43, we have a rule for how the singular part of $g$ transforms homogeneously: $g[K] = b^{-d}g[K']$ under the recursion relation (or at least an ideal one).[37] Near the fixed point, we can write this as depending on the relevant scaling variables $u_t$, and $u_h$, ignoring the irrelevant ones for the moment:

$$g[u_t, u_h] = b^{-d}g[b^{y_t}u_t, b^{y_h}u_h] = b^{-nd}g[b^{ny_t}u_t, b^{ny_h}u_h] \tag{3.53}$$

where we have iterated the recursion relation $n$ times in the last step. This cannot be too large, or the linearity assumption will break down; so we must halt the iteration at some arbitrary point which is still within the region of linearity. If we make a suitable choice of scaling factors $t_0$ and $h_0$ we can solve for $n$ to obtain:

$$g[t, h] = \left|\frac{t}{t_0}\right|^{d/y_t} \Phi\left(\frac{h/h_0}{|t/t_0|^{y_h/y_t}}\right) \tag{3.54}$$

where $\Phi$ is a *scaling function*, depending on the scaling variables alone. This form of the free energy is of the form postulated by Widom (compare to equation 3.20), and so the Griffiths and Rushbrooke equalities follow directly. But now the parameters are not arbitrary, but determined by the eigenvalues associated with

[37] We ignore any terms of $g$ that remain analytic at the critical point, since they play no part in the critical behaviour.

the fixed point:

$$\text{Specific heat } C\text{:}\quad \left.\frac{\partial^2 g}{\partial t^2}\right|_{h=0} \sim |t|^{d/y_t-2} \text{ , so } \alpha = 2 - \frac{d}{y_t} \tag{3.55}$$

$$\text{Magnetisation } M\text{:}\quad \left.\frac{\partial g}{\partial h}\right|_{h=0} \sim (-t)^{\frac{d-y_h}{y_t}} \text{ , so } \beta = \frac{d-y_h}{y_t} \tag{3.56}$$

$$\text{Susceptibility } \chi\text{:}\quad \left.\frac{\partial^2 g}{\partial h^2}\right|_{h=0} \sim |t|^{d/y_t-2} \text{ , so } \gamma = \frac{2y_h - d}{y_t} \tag{3.57}$$

$$\text{Magnetisation (as } t \to 0)\text{:}\quad \left.\frac{\partial g}{\partial h}\right|_{t\to 0} \sim h^{d/y_h-1} \text{ , so } \delta = \frac{y_h}{d-y_h} \tag{3.58}$$

So if we have chosen an explicit form for the recursion relation, we can insert values for $y_t$ and $y_h$, and calculate the critical exponents.

Some further details must be mentioned. We ignored the irrelevant scaling variables in the analysis above, because the iterated action of $\mathcal{R}$ shrinks them down so that they become insignificant compared to the relevant ones. But it turns out they are not completely irrelevant: they provide small corrections to the values of the critical exponents. One can see that this is plausible by looking again at Figure 3.4. The irrelevant directions are those two on the critical surface that 'attract' the flow in towards the fixed point, and though their contributions are suppressed by the iteration of the recursion relation, the outward flow exits *near* the relevant direction, not quite on it. The amount of this deviation is controlled by the irrelevant scaling variables, so we require a skilled experimentalist to drive the system very close to the critical point, if we are to obtain accurate values for the critical exponents.

There are also situations in which *dangerous* irrelevant variables appear (i.e., those that are not expandable as a Taylor series around the critical point), and these can affect the critical free energy in more dramatic ways. Fixed points in the same parameter space can also interfere with one another, and "cross-over" behaviour can be predicted and analysed. But these can be accommodated only

in more sophisticated approaches to RG analysis, and quickly become technically challenging. We are still in search of the explanatory strategy itself, so will ignore all of these more interesting possibilities to concentrate on the simplest cases.

### 3.4.5 An RG Explanation for Universality?

With these results, it looks as though Kadanoff's method has allowed us to achieve the aims set out in §3.2.2, to derive the Widom scaling hypothesis, and to provide a method of calculating the critical exponents. At this point in introductory physics textbooks and philosophical discussions, it is usual to state that we have an explanation of universality, which proceeds along the lines of the discussion in the last few sections: since the critical exponents are determined by the features of the critical fixed point, and many different critical systems across the parameter space lie in the basin of attraction of the same fixed point under the recursion relation, these exponents will be universal. This certainly seems to be Batterman's own position:

> It turns out that *different physical* Hamiltonians can flow to the same fixed point. Thus, their critical behaviours are characterized by the same critical exponents. This is the essence of the explanation for the universality of critical behaviour. Hamiltonians describing different physical systems fall into the basin of attraction of the same renormalization group fixed point. (Batterman, 2000, 127)

Similar statements can be found throughout the literature on RG techniques and critical phenomena.[38]

But the Kadanoff real-space RG approach seems to me to *fail* to provide an explanation of universality — and also to fail in fulfilling Batterman's criteria.

[38] For the physics textbooks, see for example: Cardy (1996, Ch.3), Yeomans (1992) and Binney et al. (1992, Ch.5) (though the latter does offer a comparison with the explanation offered by other theories). For the philosophical discussions, see for example: Liu (2001), Yi (2000), Huggett (2002, 270n), Hughes (1999), Batterman (2002a, §4.1) and (1998, §5) as well as the work quoted above.

Further, I think that these faults goes deeper than a reliance on any over-simplified exposition of Kadanoff's methods.

Let us look at the Batterman's two criteria in turn. It is true that Kadanoff's approach does better than Landau's approach in satisfying desideratum (A), that the explanation must be 'completely grounded in the lower-level theory'. Recall that Landau imposed the requirement that all near-critical systems should have a particular form of the free energy, and this form determines the Landau values for the critical exponents. This was done without reference to the microphysical Hamiltonian of the systems in question, but by defining the order parameter $M$, and writing down a reasonably general form for $\mathcal{F}$, consistent with symmetry considerations. In contrast, the Kadanoff RG techniques start from the lower-level theory, dealing directly with a Hamiltonian which encodes all the microphysical interactions amongst the spins. The recursion relation was constructed in the light of the details of this Hamiltonian, and the critical exponents derived from the fixed points of that relation. In this sense, Kadanoff's approach is certainly an advance on Landau's; it explains on microphysical grounds (that is, in terms of the properties of the Hamiltonian) why the Ising model has the critical exponents it does.

But this attention to microphysical detail was also present in the classical Mean Field approaches. We dismissed these on the grounds that they violated Batterman's condition (B), that explanation not simply provide a list of microphysical properties sufficient to realise the higher-level ones. The Mean Field approach involves idealising each microphysical Hamiltonian to the best available Mean Field form, and this form is indeed sufficient to realise the classical critical exponents. But here we observe that Kadanoff RG has exactly the same fault, since we had to tailor the recursion relation to return the same form of Hamiltonian — in our example, the Ising form. Once this is decided upon, the explanation relies upon

imposing this single recursion relation on all the models it is applied to. Like the Mean Field approach, this amounts to an assumption that all the microphysical Hamiltonian share microphysical properties. Of course, we can make this assumption less restrictive by accepting a more sophisticated recursion relation, just as we could adopt a more sophisticated form of Mean Field theory. But in neither case does the additional sophistication change the fact that our choice determined the critical exponents.

Again, we face the question of whether this idealisation is a justified one. Certainly, it should be justifiable for Ising models, for the recursion relation was tailored to the microphysics of the Ising model. But consider applying a recursion relation tailored for the Ising model, to non-Ising models. Even if the action of this relation induces a flow which comes within the basin of attraction of the same fixed point, this does not demonstrate universality (unless we assume what we are setting out to prove) for we have no guarantee that it is appropriate to apply the relation to the original models.

Of course we can go ahead anyway, and draw wide-ranging conclusions from the nature of the fixed points belonging to our Ising-tailored recursion relation. But this is exactly analogous to what we did in the Mean Field approaches, where we did not worry about the Ginzburg criterion, nor any other indication that a Mean Field form of Hamiltonian was a justifiable idealisation, but went ahead to use it anyway.

So while the Kadanoff RG may give us a good derivation of critical exponents for each specific model, it fails to explain universality in a way superior to the classical approaches. What we argued in §3.4.2 is that all Hamiltonians whose flows approach a critical fixed point share the same critical exponents. So in order to explain universality we would need to show that all systems in a universality class have flows that come near to the same fixed point. But there seems little

hope of establishing such a conclusion by an argument along Kadanoff's lines.

Furthermore, the most obvious methods of plugging this explanatory gap negate the advantages of the Kadanoff approach. If we simply *stipulate* that all Hamiltonians sharing a universality class must share a form of recursion relation (or at least share its fixed points) then the universality "explained" by the Kadanoff approach seems no less an artefact of our stipulations than it is in the Mean Field approach. On the other hand, if we construct a recursion relation for each of the Hamiltonians in turn, and succeed in demonstrating that they share fixed points, this would be restating the *fact* of universality, not explaining it. We might as well model each system on a computer and demonstrate directly that the critical exponents matched. Finally, a simple appeal to empirical success cannot help us: recall that the classical theories *do* get the exponents right for $d > 4$, where it stipulates a particular form of Hamiltonian no less than it does for $d \leq 3$.

So it appears that in as much as a recursion relation is based on an analysis of a particular Hamiltonian, Kadanoff's approach can satisfy Batterman's criterion (A), but not (B), for much the same reasons as could MFT. Thus, if mean field theory is not a satisfactory microphysical explanation for universality, I cannot see that the Kadanoff real-space RG approach does any better.

None of this is to say that renormalisation techniques cannot succeed in providing an explanation of universality, but we need very different arguments. I shall argue that we need the insights of Kenneth Wilson, who provided far more powerful methods of renormalisation and put forward a very different understanding of the technique. However, as we shall see, the Wilsonian RG operates from a very different stand-point than does Kadanoff.

## 3.5 The Wilsonian Renormalisation Group

Our criticisms of the Kadanoff real-space approach suggests a path by which the RG approach could hope to explain universality. If we are to hold that the critical exponents are governed by the nature of flows near a fixed point, we must demonstrate why we expect all the systems in a universality class to flow within the vicinity of a single fixed point. That is, we must provide an account of how a transformation appropriate to each of the model systems can be shown to take us to the domain of attraction of that fixed point.

Just such an account was provided by Ken Wilson in his Nobel-prize winning work.[39] In contrast to Kadanoff's approach, Wilson's techniques do not tie themselves to any particular microphysical model, nor even to a lattice. Instead, we consider an *effective* Hamiltonian; that is, one that represents a continuous field on space, chosen to represent the collective properties of a large number of interacting spins. A particularly useful form for such a field is known as the Landau-Ginzburg-Wilson (LGW) Hamiltonian, which we will focus on in this exposition.

There are a variety of approaches to the LGW Hamiltonian: we can view it as an improvement to Landau Theory (Wilson, 1982/1993, 104); as a generalised approach to Ising-like models and their variants (Wilson and Kogut, 1974, 95); or even as a particularly sophisticated version of a Mean Field Theory (Binney et al., 1992, §7). In each of these cases, we end up with a Hamiltonian, partition function, and free energy for a field, with a form general enough to cover a large number of microphysical models. In what follows I shall follow Wilson in considering how we might develop the Landau theory of phase transitions.

[39] The comprehensive review article of Wilson and Kogut (1974) provides technical details, and Wilson's Nobel acceptance speech Wilson (1982/1993), includes a useful historical section which lays out the influence that Kadanoff's ideas had on his thinking.

### 3.5.1 The Landau-Ginzburg-Wilson Effective Hamiltonian

We mentioned in §3.3.2 that rather than considering the order parameter as a mean magnetization $M$, a constant scalar over all space, we can consider it as a function of position, $m(\mathbf{x})$. And the order parameter need not be a scalar: we can generalise further, and allow it to be a continuous $d$-component field $\phi(\mathbf{x})$. Any Hamiltonian, partition function or free energy that depends upon this field will not be a simple function, but a *functional*. We then want to consider a Hamiltonian *density*, rather than a Hamiltonian, which I will signify by $h[(x)]$, such that:

$$\mathcal{H} = \int d^d x h[\phi(\mathbf{x})] \tag{3.59}$$

The associated partition functional is then:

$$\mathcal{Z} = \int \mathcal{D}\phi e^{-\beta \int d^d x h[\phi(\mathbf{x})]} \tag{3.60}$$

with the free energy and other thermodynamic functions now functionals, but obtained from $\mathcal{Z}$ in the same way as before. The exact meaning of the functional integration of Equation 3.60 will be discussed later, but for the moment it can be understood as implying a sum over all possible field configurations.

Recall that the form of the free energy in the Landau approach was held to be:

$$\mathcal{F}(T, M) = a_0(T) + a_1(T)M^2 + a_2(T)M^4 + ... \tag{3.61}$$

In contrast to Landau's approach, we do not formulate LGW theory by imposing a form on the free energy. Instead we impose a form on the effective Hamiltonian density, $h$: one consistent with the symmetries of a ferromagnet, and with the existence of short-range interactions. In the absence of an external field, a plausible

form for a system with short-range interactions can be argued to be:[40]

$$h_{LGW} = \frac{1}{2}\alpha^2 \left|\nabla\phi(\mathbf{x})\right|^2 + \frac{1}{2}\mu^2\phi(\mathbf{x})^2 + \frac{\lambda}{4!}\phi(\mathbf{x})^4 \tag{3.62}$$

where $\alpha$, $\mu$ and $\lambda$ are phenomenological parameters, to be filled in to match the system in question. (This slightly unfortunate notation comes to us from quantum field theory. In particular, we should not confuse the parameter $\alpha$ in this equation, with the critical exponent.) Briefly: the first term is made to be dependent on the gradient of the field, to take account of a short-range interaction. Thus the parameter $\alpha$ controls the strength of a "nearish-neighbour" interaction. $\mu$ is a temperature dependent positive parameter which controls the phase transition. $\lambda$ is known as the *coupling* parameter, another terminological hang-over from quantum field theory.

It is important to emphasise that this is an effective, phenomenological Hamiltonian: one that makes no mention of the microphysical spins themselves. Instead, $\phi(\mathbf{x})$ is an order parameter field on space — in this case interpreted as a field of magnetisation-density — and it is meant to sum up the effective contributions of a large number of spins. The LGW form of Hamiltonian density in Equation 3.62 is intended to be general enough to cope with the effects of a large number of microphysical models (including the Ising model — Binney et al. (1992, 400-4)). Assuming once more the validity of Gibbsian statistical mechanics, the partition function for the LGW model is:

$$Z_{LGW} = \int \mathcal{D}\phi e^{-\beta \int d^d\mathbf{x} h_{LGW}(\phi)} \tag{3.63}$$

[40]There are many ways to motivate this Landau-Ginzburg-Wilson form. One accessible and intuitive approach is given in Binney et al. (1992, 178-183). The papers of Wilson and Fisher (1972) and Wilson and Kogut (1974, 90-123) discuss the reasons for believing that it will be a good approach for many models in dimensions close to four.

where the notation $\int \mathcal{D}\phi$ again signifies a functional integral: the integrand is a summed over all 'admissable' values of the order parameter field $\phi$. The meaning of 'admissable' needs to be considered carefully. We have good reason to believe that this Hamiltonian should not take account of variations in the fields below the scale of the lattice spacing $a$. For it is difficult to make sense of the order parameter field fluctuating smoothly with wavelengths smaller than $a$. This demand that $\phi(\mathbf{x})$, not vary on too small a scale can be implemented by considering the Fourier transform of $\phi(\mathbf{x})$ and introducing a *cut-off* parameter, $\Lambda$:

$$\phi(\mathbf{x}) = \int_{k=0}^{\Lambda} \frac{d^d\mathbf{k}}{(2\pi)^d} e^{i\mathbf{k}.\mathbf{x}} \tilde{\phi}(\mathbf{k}) \tag{3.64}$$

so that no functions with a wavenumber above $\Lambda$ will be considered. (To be more exact, any fluctuations on a scale smaller than this are assumed to have been averaged over, and absorbed into the parameters $\alpha$, $\mu$ and $\lambda$ of Equation 3.62.)

### 3.5.2 The Cut-off $\Lambda$ and a Recursion Relation

By these means, we introduce an effective Hamiltonian density, which depends on several parameters: $\alpha$, $\mu$ and $\lambda$ as well as the cut-off $\Lambda$. The full LGW functional integral of Equation 3.63 is usually tackled by setting $\lambda = 0$, reducing the LGW form reduces to the tractable *Gaussian* model, and then treating the full LGW form as some sort of perturbation to this Gaussian model. A variety of systematic approximation schemes have been invented for calculating $Z_{LGW}$, most of them adapted from previous work in quantum field theory.

Amongst the most powerful systems are those based on Feynman diagrams, where we associate each term in an expansion of the functional integral with a simple diagram, and then sort the behaviour of the terms by their topological properties. We can then evaluate the integral to various orders by ensuring that

we have included all expressions associated with the appropriate diagrams. The highly technical details of these techniques are not crucial to us here.[41] All we need to know is that all methods of evaluating the integrals give us diverging sums, unless we impose a cut-off $\Lambda$ as a further parameter — a phenomenon known as *ultra-violet divergence*.

In the context of condensed matter, this is not particularly surprising, once we think about the physics of the system. We know that the theory will not be realistic at small scales, since eventually we will be integrating field fluctuations with wavelengths comparable to the lattice spacing, and at this point our effective Hamiltonian can no longer be valid — the physics will become utterly different at these scales, and our approach will break down. Yet while a cut-off is necessary, we also require that the details of how we impose it should be irrelevant to the results. For example, let us say that we simply ignore all fluctuations below a certain wavelength: a particularly crude method. This imposition of a sharp cut-off is just as physically unrealistic a feature of a condensed matter system, as would be a series carried to arbitrarily short scales.

Thus, while we must always impose a cut-off, the details of how it is to be imposed should not be important. We ensure this by calculating the differences in the large-scale physics that occur when we change to a different value of cut-off $\Lambda \to \Lambda + \delta\Lambda$, and calculate the changes we must make to the other parameters of the effective Hamiltonian $h_{LGW}$ in order to compensate. This is equivalent to looking at how we can absorb the effects of all of the interaction details with wavenumbers between $\Lambda$ and $\Lambda+\delta\Lambda$, into *renormalised* values of the parameters $\alpha$, $\mu$ and $\lambda$. Under good conditions, we can come up with a transformation which can move us from an effective Hamiltonian valid for one cut-off, to a second effective

[41] Kaku (1993, Ch.7 & 14) gives a clear overview of the technicalities of renormalisation using Feynman graphs, with a focus on particle field theories. Binney et al. (1992, Ch.8 & 9) go into a great deal of detail, while remaining focussed on condensed matter.

Hamiltonian valid at another, and which gives the same large-scale physics.

This transformation is clearly analogous to the recursion relations we have already discussed. For if we think of the cut-off as corresponding to the lattice-spacing $a$, then the change to a new value of $\Lambda$ corresponds to considering a "blocking" transformation replacing the original lattice with a different spaced one and then rescaling. And the two procedures are usually discussed under the same title: both transformations are referred to as 'recursion relations', or 'renormalisation transformations'. (Naturally, the closest comparison is to the continuous, rather than the discrete version of the recursion relation).

### 3.5.3 The Critical Exponents from the LGW Hamiltonian

It is worth outlining the rather complex steps that are needed to extract critical exponents from the LGW Hamiltonian, and its recursion relations. The main reasons for the complexity arise from the fact that the LGW model is not exactly solvable (i.e., $Z_{LGW}$ cannot be calculated) due to the functional integrals, and in particular due to the $\lambda$ term. The methods that have been developed to get around this restriction are both interesting and ingenious. The stages can be summarised as follows.

First consider the Hamiltonian density as consisting of two parts:

$$h_{LGW} = h_G + h_{NG} \tag{3.65}$$

that is, as a sum of a tractable *Gaussian* part, and an intractable *non-Gaussian* part:

$$h_G = \frac{1}{2}\alpha^2 \left|\nabla\phi(\mathbf{x})\right|^2 + \frac{1}{2}\mu^2\phi(\mathbf{x})^2 \tag{3.66}$$

$$h_{NG} = \frac{\lambda}{4!}\phi(\mathbf{x})^4 \tag{3.67}$$

With the Hamiltonian thus divided, it is possible to solve the term $h_G$ (i.e.: to calculate $Z_G$, the partition function of the Gaussian model) and to apply a recursion relation to this Gaussian part. Finally, we appeal to varieties of perturbation theory to allow us to apply a suitable recursion relation to the full Hamiltonian, with the non-Gaussian term treated as a perturbation to the Gaussian one:

$$h'_{LGW} = \mathcal{R}_b h_{LGW} = \mathcal{R}_b(h_G + h_{NG}) = \mathcal{R}_b h_G + \langle \mathcal{R}_b h_{NG} \rangle_0 + ... \tag{3.68}$$

To achieve the last step calls for two heavy guns from the theoretical physicist's armoury. First, we must calculate the perturbation expansion and apply the recursion relation to it — a task which calls for the technology of Feynman diagrams. Second, as $T \to T_C$ (i.e., the cases we are interested in, close to the critical temperature) we lose the analytic base of our expansion, which is shown in a phenomenon known as *infra-red divergence*. However, this divergence only happens *below* four dimensions, so the trick is to set dimension $d = 4 - \epsilon$, and expand the results in a Taylor series around $\epsilon = 0$. Surprisingly, this gives good results even when we set $\epsilon = 1$, so we can use it to calculate results for three dimensional systems. This technique is known accordingly as the *epsilon expansion*, and gave rise to perhaps the most memorable title for a Nobel Prize-winning paper: *Critical Exponents in 3.99 Dimensions* (Wilson and Fisher, 1972).

Here, we must also introduce the concept of *anomalous dimensions*. The *scaling dimension* of a quantity sums up its behaviour under a scaling transformation. As an example, consider a rescaling of length by a factor $b$. Unsurprisingly, the lengths associated with the problem scale as lengths, and the wavenumbers as inverse lengths, i.e.:

$$x = bx' \ , \ k = \frac{k'}{b} \tag{3.69}$$

so the scaling dimension for these quantities is the same as the physical dimension — the one that we would expect from simple dimensional analysis. However, it turns our that the *fields* must be rescaled by a slightly different factor if they are to reproduce the large-scale results:

$$\phi(\mathbf{x}) = b^{-d_\phi}\phi'(\mathbf{x}') \tag{3.70}$$

therefore, in wavenumber space:

$$\phi'(\mathbf{k}') = \int d^d x' e^{-i\mathbf{k}'.\mathbf{x}'}\phi(\mathbf{x}') \tag{3.71}$$

$$= b^{-d_\phi}\int d^d x e^{-i\mathbf{k}.\mathbf{x}}\phi(\mathbf{x}) \qquad = b^{-d_\phi}\phi(\mathbf{x}) \tag{3.72}$$

The difference between the scaling and physical dimensions: $d_\phi - d$ is called the *anomalous dimension* of the LGW model. This anomalous dimension shows up in the values of the critical exponents, and all the classical approaches predict an anomalous dimension of zero — the same as simple dimensional analysis would predict. More than anything else, it was this non-zero anomaly that puzzled pre-RG theorists (see the historical remarks of Fisher, 1998, 657).

### 3.5.4 New Features of Wilson's Approach

The relations between the parameters of the LGW Hamiltonian as considered at different scales are thus determined by a recursion relation very similar to the sort considered by Kadanoff. True, the technical details are somewhat different: we typically work in wavenumber (momentum) space, and rescale our parameters according to the scales of the wavelengths we take into account, whereas Kadanoff's techniques usually operate in real space. But still, the effective Hamiltonian suitable for one scale, is related to a Hamiltonian suitable to a very different one, by a

recursion relation acting on its parameter values. But in an important difference to the Kadanoff approach for the Ising model, the LGW model can absorb the effects of interactions, without forced approximations, *and* absorb them into only a finite number of terms. For the LGW Hamiltonian has a very important property: it is *renormalisable*.

**Renormalisability**

To call a field theory (= a form of field Hamiltonian) *renormalisable* is to say that the action of a recursion relation does not generate an infinite series of new terms in the Hamiltonian, but merely adjusts a finite (and usually small) number. That is, changes in the cut-off value $\Lambda$ can be compensated for by changes to a finite number of parameters. The significance of this property can be seen if we consider the contrast in significance between the Ising and LGW Hamiltonians.

As we saw in §3.4.1, if we construct a fully accurate recursion relation for the Ising Hamiltonian, then each time it is applied it will generate new terms in addition to the originals. Our way of dealing with this is to neglect the higher terms at each stage, forcing the new Hamiltonian to again take on an Ising form, and just hope that none of the higher-order ones would become important. This is one reason for Wilson's description of Kadanoff's scheme as absurd — there is no guarantee that these higher order terms would not be far more important than the ones we kept.

The LGW Hamiltonian is different. We know that a LGW-form Hamiltonian with a certain cutoff, can be re-expressed as a second LGW-form Hamiltonian, with a different cut-off, and with different parameter values. But its renormalisability means that there will appear no extra parameters in addition to those we have already written down; that is, higher-order terms will not come to dominate. This is not to say that the calculations which we are faced with will be exact, or even

tractable, but we can at least be sure that if we start with a Hamiltonian of LGW form, and consider a new momentum scale, the rescaled version will not neglect the most significant aspects of the physics.

### Relevant and Irrelevant Operators

At the heart of the LGW approach to explaining universality is the idea of *relevance* and *irrelevance*. These terms have the same sense as in Kadanoff's procedure in §3.4.2, but since the Ising model has a different significance to the LGW model, the implications are different. It is traditional to refer to each term of the LGW Hamiltonian as an *operator* of order $n$: $O_n$, which contributes to the Hamiltonian a term of the form $\int d^d\mathbf{x}\phi^n$. (Once again, the quantum mechanical connotations of the term 'operator' are due to its origins in QFT, and must be firmly ignored). We look at each term in the Hamiltonian density $h_{LGW}$ and ask whether it is a relevant or irrelevant operator, depending on how it is transformed by application of the recursion relation. We can also consider possible additions to $h_{LGW}$, and ask whether these operators would be relevant or irrelevant.

The point is that if we can show that an operator is irrelevant, it will mean that it can have no effect on the critical behaviour of the system. For recall the discussion of §3.4.2. As we can see from Equation 3.52, if we repeatedly apply a recursion relation to a Hamiltonian containing an irrelevant operator, that term will be made smaller and smaller. Yet the recursion relation has been designed in such a way that applying it repeatedly should not change the critical behaviour. Thus the origin of the term 'irrelevant' operator: since they will be irrelevant to the critical behaviour. Relevant operators, on the other hand, will almost certainly have an effect on critical behaviour, and marginal operators may have an effect — often one that is rather hard to calculate.

So if we look at the relevance of an extra operator treated as an addition to the

| Scaling properties of operator $O_n$ in $d$ dimensions | | | | |
|---|---|---|---|---|
| $d$ | $n=2$ | $n=4$ | $n=6$ | $n=8$ |
| 5 | 2 | -1 | -4 | -7 |
| 4 | 2 | 0 | -2 | -4 |
| 3 | 2 | 1 | 0 | -1 |
| 2 | 2 | 2 | 2 | 2 |

Table 3.3: Scaling properties for operators of the Gaussian model. For a table entry $x$, the rescaled free energy scales as $b^x$. Therefore, if the table entry is positive (negative), the operator $O_n$ is relevant (irrelevant).

LGW Hamiltonian, we will be able to see whether the critical behaviour is likely to be affected by that operator. The answer turns out to depend on the dimension of the system, and the analysis again depends on the technical apparatus of Feynman diagrams. But as an illustration, we can consider a far simpler example, and examine the relevance of various even operators to the Gaussian model (that of Equation 3.66).

$$h_G = \frac{1}{2}\alpha^2 \left|\nabla\phi(\mathbf{x})\right|^2 + O_2 + O_6 + O_8 \tag{3.73}$$

(where we have been taking $O_2$ to be $\frac{1}{2}\mu^2\phi(\mathbf{x})^2$ in both the Gaussian and LGW models, and $O_4$ to be $\frac{\lambda}{4!}\phi(\mathbf{x})^4$ in the case of the LGW model.) Table 3.3 summarises the results.

This analysis indicates why the classical methods might fail below $d = 4$, but succeed in higher dimensions. At higher dimensions, only the operators of the Gaussian model are relevant, so we observe the Landau exponents. But at lower dimensions, some of the higher $n$ operators become relevant, and this can start to have a more complex effect on the critical behaviour, giving rise to the more complex LGW analysis.

Of course, we want to perturb the LGW model, not the Gaussian one, and this

is possible with a great deal more theoretical work (see for example: Binney et al., 1992, §14.2). And we must also consider the relevance of many different types of operators (not just the even ones) as well as additions to the LGW Hamiltonian that may have effects we cannot account for using these perturbative methods. Clearly, the generality of any conclusions will be dictated by the range of operators we have shown to be irrelevant. What we can say, however is this: if we can show that an operator is irrelevant, then any model that has an effective Hamiltonian differing from the LGW only by that operator, will have the same critical behaviour as LGW. And it turns out that one can prove irrelevance of large classes of operators based on their order, and their symmetry properties (for example, one can be sure that all even-$n$ operators of higher order than $n = 6$ will be irrelevant in more than three dimensions).

In my view, such results offer a fully satisfactory explanation of universality, and also demonstrate why universality classes should be dependent on both the symmetry of the order parameter (reflected in the symmetry of the operators), and the dimension of the system.

### Physical Interpretation

I think the division drawn here between Kadanoff's and Wilson's approach is a useful one, but it is not the only one possible, and the names I have chosen are not entirely apposite. For Wilson contributed much to the techniques and understanding of real-space techniques, and Kadanoff did the same for LGW and related approaches. And there are many other broad divisions that can be drawn amongst the variety of renormalisation methods.[42] But the points I want to press home are the differences in physical interpretation between the 'Kadanoffian' and 'Wilso-

[42]Wilson himself divides the methods according to a four-part categorisation (Wilson, 1975, 776), which adds to these two categories the Wegner formal theory of fixed points, and some older techniques in quantum field theory.

nian' approaches. For although the procedure of transforming between models appropriate to different scales is common to both, the significance of the model considered at each scale is very different.

First, let us consider the strong parallels: changing our cutoff value in the LGW model, is analogous to a rescaling in the Kadanoff picture. For we can consider $\Lambda$ as physically corresponding to the lattice spacing, the cut-off reflecting the fact that we do not take into account any fluctuations on a smaller scale. Then a change of $\Lambda$ can be compared to applying a 'blocking' transformation to produce a new Hamiltonian, and the changes in the parameters $\alpha$, $\mu$ and $\lambda$, can be compared to the changes in the couplings $[K]$ in order to keep the large-scale physics the same.

The main differences come when we ask how we should interpret the Hamiltonians that are transformed. The Kadanoff 'blocking' procedure is applied to a specific microphysical model, which is at least an attempt to represent microphysical interactions. When we apply the recursion relation, we generate a new Hamiltonian, which we recognise as an *effective* one, in that unlike the original, it is not intended to represent any microphysical properties directly. The effective interactions generated are then truncated at a certain order so they have the same form as the old ones. In contrast, the LGW approach, we *start* with an effective Hamiltonian, and since we choose it to have a renormalisable form, we can be confident that no new terms will be generated. Another way of looking at the difference is that the Kadanoff approach takes a realistic microphysical Hamiltonian, ignores the fact that it is almost certainly not renormalisable, and — regardless — truncates the extra terms that should be generated. In opposition, the LGW approach operates with a non-realistic, non-microphysical, but renormalisable Hamiltonian.

Translating this into the micro-macrophysical language we used earlier: the LGW Hamiltonian need not aim to represent a microphysical model at all. Rather,

it aims to represent the relevant directions associated with a *fixed point* in the parameter space.[43] Thus our conclusions about relevant and irrelevant directions associated with the LGW Hamiltonian will not apply to just a particular microphysical model, but will concern *any* model that has a flow that may come within the basin of attraction of the LGW fixed point(s).

But now it looks as though the Wilsonian approach does no better than the classical approaches in satisfying Batterman's criteria for an explanation of universality. For in choosing the LGW form of Hamiltonian density, Wilson's approach is to impose from the 'top down' a form for the Hamiltonian, thus violating Batterman's desideratum (A). This is exactly the same point that disqualified Landau theory as a satisfactory explanation for universality — it is *not* based in the microphysical details of the models in question, but is imposed by external, macrophysical constraints. The LGW Hamiltonian is an *effective* form, which we adopt on the basis of macroscopic symmetry considerations of the phase transition in question. True, the LGW theory works with a Hamiltonian density, not with the free energy; and it is more sophisticated and empirically successful than Landau's original, but it ignores the lower-level theory just as much.

## 3.6 Explaining Universality

To sum up the conclusions of the last two sections: there are various 'classical' theoretical approaches to phase transitions, which pre-date renormalisation, but their empirical success is only partial, and they do not conform to Batterman's desiderata for a new approach to explain universality. Of the two we have examined: Landau theory imposes a common set of higher-level properties, in a single

[43] In fact, the LGW form can encompass more than one fixed point, both the so-called "Gaussian" fixed point, which we obtain with $\lambda = 0$, and the Wilson-Fisher fixed point, which can appear with non-zero values of $\lambda$.

form for the free energy; and MFT found (or just assumed) properties common to the microphysics of the models, idealising each of them as a mean-field form of microphysical interaction. Thus, as one would expect, neither of them fulfilled Batterman's desiderata for a genuine explanation of universality, and thus of multiple realisability.

But in the last sections, we saw that renormalisation techniques share the deficiencies of the classical approaches: Kadanoff's argument imposes microphysical elements in common; Wilson's approach imposes a general form of effective Hamiltonian due to macrophysical considerations. So we are left with a problem: it looks as though RG techniques are no advance on the classical ones in explanatory terms, even if they do go beyond them in empirical prediction.

In this section and the next, I offer an explanation as to how Batterman's opposing view could arise. We return to basics; first: what exactly we are aiming to explain (§3.6.1); second: what is the explanatory strategy of renormalisation (§§3.6.2 & 3.6.3). This leads to §3.7, where I discuss how a conflation of different approaches could lead us to the illusion that Batterman's criteria can be satisfied by renormalisation.

### 3.6.1 What is Universality?

Batterman's strategy of identifying multiple realisability as a variety of universality is attractive in that while the former is a controversial philosophical term, the second is a widely-used term in physics, and can therefore be expected to have a precise technical meaning. And Batterman does suggest that the terms 'Universality or Universal Behaviour' are used by physicists in 'special' and 'technical' senses (Batterman, 2000, 116, 120). But the only definition he gives is contained in the following passage: 'Are there parameters that somehow characterise the dominant features of systems of widely different types? If there are, physicists often call the

observed behaviour and the dominant features 'universal'.' (ibid., 120). Clearly, this definition is unsatisfactory, for it covers properties far outside the scope of RG techniques; both trivial and non-trivial examples of multiple realisability. And even within the scope of RG analysis, it counts properties shared across all universality classes (sometimes called *trans*-universal) such as the fulfilment of the Widom equalities.

In practice, Batterman uses the term "universality" to refer to those properties that he believes that an RG analysis can be applied to, an approach in common with many expositions presented in physics textbooks. And other philosophers' explanations of RG techniques appear to rest satisfied with similarly vague ideas of universality:

> ... the fact that whole classes of radically different systems and phenomena shared the same critical exponents is known as the universality, which has no explanation in Thermodynamics. The systems and phenomena can be as different as between fluids for the transition from the one-phase state of gas to the two-phase state of gas and liquid and magnets for the transition from the paramagnetic to the ferromagnetic phase. (Liu, 2001)

Batterman states that he has taken his lead from the physicist Sir Michael Berry:

> ... saying that a property is a "universal feature" of a system is the slightly pretentious way in which physicists denote identical behaviour in different systems. The most familiar example of universality from physics involves thermodynamics near critical points (Berry, 1987, 185).

Yet as Batterman himself points out, Berry's characterisation implies that familiar cases such as the period of a pendulum being independent of its microstructure would be "universal" just as much as the critical phenomena. This is something that Berry clearly does not intend, or his 'most familiar example of universality' could be taken from contexts a great deal more familiar than critical phenomena.

Less trivially, classical equilibrium thermodynamics holds for a wide variety of systems with very different microphysics, and properties such as pressure, temperature and entropy could be understood as universal in this sense. But as already mentioned, the widespread appearance of properties such as "temperature" are usually explained by the systems possessing common microphysical properties. That is, they count as "trivial" examples of multiple realisability, in the sense of §3.1.1, and do not require an novel explanatory strategy.[44]

It is also worth repeating that the 'classical' analyses of critical phenomena predicted critical exponents even more universal than those predicted by RG. Rather than a dependence even on minimal features of the systems, such as dimensionality, they predicted values universal to almost all near-critical systems. In this respect, RG techniques predict strikingly *non*-universal quantities, when it (correctly) separates systems into a finer-grained set of universality classes which depend on the symmetries of the order parameter, and the dimensionality of the lattice.

I think it revealing that while philosophers focus on the universality of the critical exponents as the important innovation of the renormalisation techniques, physicists do not. In Fisher's historical review of the subject he writes:

> If one is to pick out a single feature that epitomizes the power and successes of RG theory, one can but endorse Gallavotti and Benfatto when they say "it has to be stressed that the possibility of nonclassical critical indices (i.e., of nonzero anomaly) is probably the most important achievement of the renormalization group." (Fisher, 1998)

As already mentioned, the anomalous dimensions encode the deviation of the critical values from the Landau values, provided perplexing theoretical puzzles before the advent of renormalisation. For it is not that without RG analysis, universality would be a mysterious phenomenon. Rather, it is that there was no clear way

[44]Batterman (1998) suggests that an approach similar to renormalisation analysis might also given insight into the success of equilibrium thermodynamics, but he also caution that it is inappropriate to apply it in all cases (Batterman, 2000, 130).

of obtaining the correct exponents, nor the observed separation into universality classes.

Thus 'universality', at least in the vague definition offered here is not the real explanandum for RG. Rather, the facts to be explained are the exact ways in which systems fall into their universality classes, and the values of the exponents which define those classes.

### 3.6.2 What is Renormalisation?

There are several ways of looking at the techniques collected under the 'Renormalisation' banner, which suggest different interpretations of the explanatory strategy employed. I would like to separate three popular approaches.

The first is often presented as a fundamental insight underlying the renormalisation approach, and tends to be especially emphasised in introductory and popular expositions. This is the idea that the structure of a near-critical system repeats itself on different scales. I shall call it the *self-similarity view*. Though it is widely reported otherwise, I do not believe that this insight does a great deal of explanatory work, and it certainly does not lie at the heart of the RG strategy. The second I call the *sensible sum view* and is based on the attitude taken by Kenneth Wilson in his papers which put the RG techniques on firm foundations, and was elaborated in his subsequent review articles. The third *asymptotic view* has been developed by more recent workers, who have extended and generalised the RG techniques to applications such as chaotic phenomena and turbulence.

It is rather striking how utterly different these attitudes appear. I suggest that while all three of these viewpoints have some validity, they each arise from a focus on particular aims and on particular applications of RG techniques, but do not illuminate the explanatory strategy itself.

### The 'Self-Similarity' View

Writing about the physical basis for the RG techniques in calculating critical exponents, Sang Wook Yi outlines the justification for its application as follows:

> Figuratively speaking, the system 'looks' the same whether we look at the whole system or [a] certain part of it provided the correlation length is large enough. To put it differently, a system with a large correlation length looks the same whether we see it closely or a bit from a distance. As the Ising-2D system approaches its phase transition point, its correlation length gets larger and larger. So we can expect that in its critical region the Ising-2D system exhibits so-called scale invariance: the invariance of the critical properties of the system under the (length) scale change. (Yi, 2000, §5.3.2)

This idea — that it is self-similarity at a critical point that allows us to calculate numerical values using RG techniques and explains universality is too often put in careless terms. I hope enough has been said to see that unqualified statements of this sort are — at the least — highly misleading. Slogans such as 'Looking the same at different length scales' needs to be cashed out, but the clearest sense available is the effect of applying the recursion relation, which relates descriptions of systems at different scales. But the two most obvious interpretations are mistaken: First, if by 'looking the same', one means that it can be expressed in terms of a Hamiltonian of the same form but renormalised parameter values, then this is true of *every* point in parameter space, for we truncate any extra terms generated (at least for Kadanoff's methods, which Yi is discussing). But on the other hand, if one means that $\mathcal{H}$ stays *exactly* the same — not only the form, but the values of the parameters remaining unchanged — then it is only at the *fixed* points in parameter space that we find systems self-similar in this sense. Elsewhere on and close to the critical surface, (i.e., at or close to each system's critical point) the parameters *are* altered by the recursion relation — if they were not, there would be no flow toward the fixed points, and there would be no explanation for the universality of

the critical exponents.

There is one notion of self-similarity which is appropriate, and this is the claim that the thermodynamic quantities associated with near-critical systems take on the Widom scaling form (i.e., they satisfy Equation 3.20). But this is not the explanation *of* universality, it is the explanandum. We must explain why this scaling form is shared by many near-critical systems, and more particularly, why members of universality classes share particular values of the exponents. This is what the RG is intended to do, by showing that flows from near-critical systems will exit the critical surface along the relevant directions.

There are other, more nebulous senses in which critical systems do "look the same" at every scale; for example, the correlation length is infinite, and there are fluctuations of all sizes, with no characteristic scale. But we cannot put these facts at the basis of an explanation of the universality of critical exponents, for it does not follow that all critical systems in a universality class show similar behaviour. This is easily seen by the existence of the many *non*-universal properties of critical systems: any adequate explanation for universality must also explain what determines whether a property is universal or non-universal, and here a vague appeal to the fact that the system "looks the same" cannot help us.

### The 'Sensible Sum' View

In his two articles that showed how RG techniques could be developed into a powerful method for analysing critical phenomena, Wilson presented his techniques as a purely technical solution to a purely technical problem.[45] The technical problem is to apply statistical mechanics to some particularly tricky cases. Calculating the

[45]Maris and Kadanoff (1978) state that they shall assume that 'the RG approach is a practical but approximate way of calculating the sums involved in $\mathcal{Z}$.' But they also suggest that this viewpoint might be limiting, and miss the 'deep physics of the renormalization group'. Sadly they give no clue of where this 'deep physics' might be found.

critical exponents involves evaluating the free energy and partition function near the critical point. Unfortunately, the usual methods for the evaluation of the sums and integrals are defeated by the multi-scale fluctuations which occur as the correlation length $\xi \to \infty$. The RG provide a solution by reorganising the calculation of the free energy and partition function — i.e., by doing the integrals or sums in a cleverly chosen order:

> The renormalization-group approach is to integrate out the fluctuations in sequence, starting with fluctuations on an atomic scale and then moving to successively larger scales until fluctuations on all scales have been averaged out. (Wilson, 1982/1993, 584)

This attitude is repeated in his other review articles (e.g., Wilson, 1975): RG techniques are seen as a set of methods for gradually integrating out the degrees of freedom, starting at the smallest scales, and working upwards.

I take it that this 'Sensible-Sum' attitude is the interpretation that Chuang Liu has in mind when he describes the RG as consisting of a 'systematic — e.g. a re-iterative — set of transformations which coarse-grain out the degrees of freedom that are tied up by the long-range order' (Liu, 2001). It is also explicitly expressed by Bruce and Wallace, when they characterise the 'essence' of the RG as a method to 'reorganise sums [...] so as to deal successively with the contributions made by configurational features of larger and larger scales' (Bruce and Wallace, 1989, 243).

On this view, we are engaged in an exercise of standard statistical mechanics, deriving all the thermodynamical quantities from the partition function or free energy in the usual way. Where we meet difficulty doing the sums or integrals, they must be reorganised. Of course this suggests that the RG method may be extended to other physical problems involving a large number of degrees of freedom, where a similar reorganisation may be useful. Wilson again:

> The basic problem causing the difficulties in understanding critical phenomena is the problem of the infinite number of degrees of freedom.

> This problem is also the bottleneck in quantum field theory and in many of the stubborn problems of solid-state physics. The methods of this paper are methods for attacking the problem of the infinite number of degrees of freedom regardless of where this problem arises. (Wilson, 1971b, 3185)

So on this interpretation, RG techniques add nothing to the explanatory power of the theories to which they are applied. Statistical Mechanics (or for that matter, Quantum Field Theory) is just supplemented with some extra mathematical apparatus so that these theories can be applied in some rather challenging circumstances.

### The Asymptotic View

A third attitude to RG techniques is to view them as a particular method of *asymptotic analysis*: the study of how some large-scale features of a system stabilise as a quantity is taken to a limit. All agree that this is an achievement of the RG method, but on this reading, we regard it as fundamentally a study of limits. The simplest form of asymptotic analysis is elementary dimensional analysis, where all the quantities associated with the problem scale as their physical dimensions; more sophisticated versions include the powerful techniques of Barenblatt (1979). In addition, it has been claimed that there are strong analogies between asymptotic analysis and renormalised perturbation theory (Chen et al., 1989, 1994). On the view of these authors, 'the essence of the renormalization group method is to extract structurally stable features of a system which are insensitive to details' (Chen et al., 1994, 376). Bruce and Wallace also hint at this attitude (apparently as an alternative to the sensible-sum view) when they speak of those quantities 'sufficiently deep-seated to survive coarse-graining and which together serve to define the system's universal critical behaviour' (Bruce and Wallace, 1989, 248).

This is the interpretation championed by Batterman;[46] he characterises the essence of RG techniques as follows:

> The RG type analysis *demonstrates that many of the details that distinguish the physical systems from one another are irrelevant* for their universal behavior. At the same time, it allows for the determination of those physical features that are relevant for that behavior. Finally it allows for the quantitative determination of the nature of the universal behavior by enabling one to identify the critical exponents. (Batterman, 2002a, 42, italics in original)

So on this view, the RG techniques are in essence techniques for sorting large-scale behaviour common to many systems from the large-scale behaviour that they do not share, as well as being a quantitative calculational technique.

At first sight the self-similarity, sensible Sum and asymptotic viewpoints look completely different, and rather hard to reconcile. Nigel Goldenfeld, a proponent of the last of these, sees the opposing views as obstructing the application of RG techniques to new areas of physics. He laments of his education in the sensible sum view: 'like most condensed matter physicists, I thought that the renormalization group was, at heart, concerned with "integrating out degrees of freedom". It was difficult to see the relevance of this to non-equilibrium systems.' (Goldenfeld, 1992, 1). A reconciliation is — I think — possible, though none of the viewpoints seem to capture what the RG techniques are about, 'at heart'.

### 3.6.3 A Reconciliation: RG as an Iterated Idealisation

It is possible to put forward a slightly deflationary attitude to the strategy embodied by RG techniques, that might also explain how each of the aforementioned views express important aspects of renormalisation, and yet appear so different from one another.

[46]The thesis is emphasised in all of his articles on RG, but in particular: Batterman (1992, 1995, 2002b) and his book: *The Devil in the Details* 2002a.

When we apply RG methods it is often commented that the physical insight appears only in designing an appropriate recursion relation for the system under question (e.g., Maris and Kadanoff, 1978, 654). And a well-designed recursion relation is simply a rule by which we replace a very complex model of a system with a large number of degrees of freedom, with a less complex model — the large-scale effects of the extra degrees of freedom being absorbed into renormalised parameters.

But there are a great number of designs for recursion relations, which apparently have little in common beyond simplifying the problem in question while retaining the large-scale properties of the system. Focussing on this thought, each application of the recursion relation can be seen as a mathematical realisation of a technique ubiquitous in science — that of *idealisation*. I discuss varieties of idealisation in a little more detail in the next chapter (especially §4.3.3), but here I mean simply the practice of replacing one theoretical model with another, this replacement accompanied by some justification that the replacement should make no (or little) difference to the quantities that we are interested in modelling. In the specific cases of the recursion relations we have been examining, this replacement plays a very particular role: we replace a Hamiltonian which takes account of small-scale interactions, with a new one that does not, but we compensate for neglecting them by making changes to the parameter values of the Hamiltonian. Thus we move from a complex model to a simpler one, but with good reason to believe it will still give us insight into the original. When we iterate the application of the recursion relation we simply iterate the "idealisation" many times.

This view is quite close to Kadanoff's original stated aim, which was to transform an Ising system near criticality, into an Ising system far away from criticality, where orthodox methods might be applied to calculate the critical exponents (Kadanoff, 1966). The modern view of both real-space and Wilsonian renormali-

sation has shifted. Rather than only looking at the systems at the end-points of the flows, we look at the flows themselves. In other words, we look at the ways in which we can reach different idealisations, not just the idealised models themselves.

Looking at the RG strategy as a formalised process of iterated idealisation suggests that the other views mentioned may not be in conflict with one another: rather they just focus on different aspects of RG techniques. First, the Self-Similarity view focusses on a particular feature of near-critical systems that is sometimes illuminated by the process of iterated idealisation. When we look at a Hamiltonian near criticality, we find that the idealised versions share certain characteristics with the original (in particular, they share an overall form). And the RG can illuminate exactly how the features change with scale: finding that near-critical systems share interesting features in this respect. But this is not the essence of the RG, nor is it the explanation of universality, rather it is something that the RG can be used to study.

The Sensible Sum viewpoint concentrates on a particular set of mathematical techniques which we implement when we apply our iterated idealisation to the expression for a free energy or partition function, for a near-critical system. Here, we are indeed faced with a large intractable sum (or integral) when we attempt to solve the original problem. And when we iterate the recursion relation, we do indeed end up splitting this sum according to fluctuation scales. For in renormalising, we absorb the effects at each scale, starting with the shortest, and taking into account their combined effects on the longer. (Kadanoff's methods order the sum by a real-space scale, whereas Wilson's methods do so on wavelength.) However, it seems misleading to think of this procedure just in terms of the mathematics that must be done; it is not just a re-ordering of sums, but a consideration of a new model at each stage, each with its own effective Hamiltonian, partition function and free energy.

Finally, the Asymptotic viewpoint focusses on a particular *application* of RG techniques: to identify universal phenomena associated with critical fixed points, as those will be stable under changes in the microstructure of the system. Certainly this is *one* application of RG techniques (and a most fruitful one) but it plays down the method by which it is achieved. Taking the asymptotic view also ignores the fact that the same methods can be profitably used to analyse *non*-critical phenomena, where no such universal stable quantities are found as the limit is taken (as we emphasised at the end of §3.4.2).

One may wonder whether anything of substance hangs on the choice of interpretations discussed in the last two sections. For practical purposes, there is much to be said for adopting the interpretation that best suits the application in hand. However, I think that when discussing the philosophical significance of the RG strategy we must focus on a minimal core shared by RG techniques, and not bring in either the specific technical means by which it might be implemented, nor the aims it may achieve when applied. And for these purposes, there are good reasons for preferring the Idealisation viewpoint.

### 3.6.4 Iterated Idealisation: An Original Explanatory Strategy?

If we take the position recommended in the preceding section, and hold that renormalisation is a process of iterated idealisation, there are two obvious places to search for a strikingly original aspect of the technique: it could be in the nature of the idealisation implemented by a recursion relation, or it could be in the fact that it is iterated.

Taking the first of these: techniques of idealisation are endemic to physics, and have been explicitly recognised since at least the time of Galileo. And while there are doubtless many different varieties used in the sciences, which can be separated

from one another and analysed,[47] we can be confident that the particular variety of idealisation implemented by a recursion relation need not be original to RG techniques. At least the process of rescaling and renormalising parameter values in order to absorb microphysical details — or more generally, to reduce the number of variables involved in a problem — is certainly nothing new. A simple example is provided by the well-known technique of replacing the problem of calculating the motions of a two-body mutually gravitating system, with the problem of single body moving in a central field. The change in parameter values can be seen in the replacement of the masses of the two bodies by the single *reduced mass*: $\mu = \frac{m_1 m_2}{m_1 + m_2}$. Or, for an example closer to the condensed matter context, consider the self-consistent Debye-Huckel theory of electrons. According to Coulomb's law, the potential at a distance $r$ from one electron is given by: $V(r) = \frac{q}{r}$, where $q$ is the electron charge. But in a cloud of electrons, say in an electrolyte or plasma, a cloud of charge around each electron *screens* the Coulomb potential, and accordingly, the Debye-Huckel potential is given by: $V(r) = \frac{qe^{-\frac{r}{l}}}{r}$, where $l$ is the Debye-Huckel length, which depends on the number density of electrons and the temperature. This is exactly the same form as the Coulomb potential, but we have replaced the charge parameter $q$ with $qe^{-\frac{r}{l}}$. Even the language used is the same: we say that the "bare" charge has been replaced with a "renormalised" charge.

If it is not the variety of idealisation that is original to the RG techniques, perhaps it is the fact that we iterate it, in the repeated application of the recursion relation. This is how I read the conclusions of Lesne (1998), who argues that we should distinguish two senses of the word "Renormalisation". The first is more associated with field theories, and refers to a one-shot procedure of imposing a cut-off — also called *regularisation*. This is similar to the Debye-Huckel procedure

[47] McMullin (1985) explores different varieties of idealisation in Galileo's own work. Nowakwa and Nowak (2000) is an outline of one of the most sophisticated general approaches to a theory of scientific idealisation, and draws on many historical case-studies.

of replacing a "bare" quantity with a "renormalised" one. Lesne judges this as less interesting than the second sense, which is the one we have been focussing on, the examination of RG flows in parameter space to extract useful empirical predictions. Thus, on her view, the interesting sense of the term turns on the fact that we have managed to extract empirical conclusions by an iterated process of idealisation, not by idealising a specific model.

But there is nothing new in iterated stages of idealisation, nor indeed in the idea of taking a limit of such a series. For such a practice is common and can be traced back at least to Galileo's consideration of a series of falling objects encountering progressively less air resistance (for discussion see McMullin (1985) and Laymon (1991)).

Lesne's idea that there is something qualitatively different about extracting predictions from the iterated process of idealisation, rather than idealised models, is also somewhat misleading. For while it is true that RG techniques extract predictions from the flows induced by the recursion relation, and not the models themselves, we tailored the recursion relation to the model to which it is to be applied. In both Kadanoff and Wilson's approach, the construction of the relation is determined by the nature of a Hamiltonian model, be it an effective or a microphysical one. And this is just as well, since we should not be able to draw conclusions about the critical exponents 'from the armchair' with no empirical support.

These considerations suggest that the explanatory strategy of renormalisation adds nothing to familiar forms of scientific explanation, but I must admit uncertainty as to where to place it in the usual classifications (that is, as causal-mechanical, unificatory, deductive-nomological, etc). For taken in their context within condensed matter physics, RG explanations combine idealisation, statistical mechanics, and limit operations — all of which are seen as controversial in the

literature on scientific explanation. Once again, I seek safe ground. If we restrict ourselves to the pure philosophical issue of whether RG involves a novel *explanatory strategy*, then the foregoing discussion shows that the strategy of the RG involves nothing original over and above classical techniques for critical phenomena. (Of course, it involves many originalities (and empirical triumphs) from a physicists' point of view.) As these classical techniques are agreed to be standard as to their explanatory strategy, I shall assume that the RG strategy can be accommodated under the one of the standard headings in the philosophy of scientific explanation, and leave it undecided which one.[48]

## 3.7 Concluding Remarks

We began in §3.1 by challenging two claims:

(1) The Renormalisation Group explains the phenomenon of universality.

(2) The explanatory strategy of the Renormalisation Group is an original one, profoundly different to the old and more familiar attempts to explain either multiple realisability or universality.

As to (1), we have seen that Kadanoff's approach cannot explain universality in the requisite sense. It can account for the values of the critical exponents for a given microphysical model, but not why their values divide them into universality classes. Wilson's approach does a great deal better. It demonstrates that the irrelevant operators of interaction Hamiltonians make no difference to the critical

[48]If forced to categorise the RG explanations of universality, I would count both Wilson's and Kadanoff's approaches as causal-mechanical. Both give a broadly mechanical account of how the critical exponents arise by showing how the Widom scaling forms of the free energy arise near critical points. The issue of causation is less clear, but this is a generic feature of approaches based on statistical mechanics.

behaviour, and so explains how universality can arise amongst models falling under the influence of a fixed point.

As to (2), the notions of 'originality' and 'profound difference from the past' have been made precise in terms of Batterman's criteria (A) & (B), and both approaches found deficient, though for different reasons.

In §3.7.1, I shall look at how the belief in the profound originality of renormalisation could have arisen, suggesting that it may be due to a conflation of Kadanoff and Wilson's approaches. In §3.7.2, I shall look at how the "Iterated Idealisation" view of renormalisation can help us locate the central distinction between the two approaches, not as a difference in the *explanatory strategy* that they implement, but in the *theoretical models* that they are applied to. I conclude in 3.7.3, with a look ahead to Chapter 5 where I suggest that renormalisable models have a representational role not covered by orthodox accounts within physics or philosophy.

### 3.7.1 A Diagnosis

In opposition to the conclusions of the previous sections, philosophical commentators such as Batterman, Liu and Yi do seem to view renormalisation as explaining universality in a profoundly original way. From the discussion so far, we can identify at least two distinct and important misunderstandings about RG techniques.

First, I suggest that the practice of introducing RG techniques by Kadanoff's procedure, and treating his approach as the royal road to thinking about renormalisation, has led to confusion between the two methods presented here. For if we conflate Wilson's and Kadanoff's approach, we might think that there exists a procedure which fulfills both of Batterman's conditions. Specifically, we might imagine a procedure that is both (a) wholly microphysically based — like Kadanoff's real-space procedure for the Ising Model, and also (b) a demonstra-

tion that a variety of microphysically heterogenous Hamiltonians may exhibit the same critical exponents — like Wilson's analysis of the LGW model. This *would* constitute an explanation of universality which would fulfil both of Batterman's requirements. Unfortunately, such a procedure is unknown in renormalisation theory.

Second, it is possible that the technical details of the LGW Hamiltonian might lead us astray, and make us think that it was intended as a direct representation of microphysical properties. Unlike Landau's approach, which imposed a form on the free energy itself, and thus wore its macrophysical pedigree on its sleeve, we considered an LGW form of the *Hamiltonian*. If we thought that this implied that the LGW Hamiltonian had the same physical status as the microphysical Hamiltonian we considered when setting up the Ising model, then we might be misled into thinking that the LGW approach was based totally in a microphysical theory. The common use of the name: 'LGW model', might also give this impression, for it suggests that the Hamiltonian plays a similar role to that of the Ising, Heisenberg, Potts and other microphysical models. But to emphasise again, it is not — the LGW Hamiltonian expresses an effective form for the order parameter field, summing up the contributions of many microphysical spins.

### 3.7.2 Kadanoff versus Wilson

Viewing RG techniques as processes of iterated idealisation, it is possible to locate clearly a central difference in the approaches of Kadanoff and Wilson. For the explanatory strategy remains the same, but in each case, it is applied to a physical model with very different representational significance.

First, in Kadanoff's approach, the process of iterated idealisation is applied to an original Hamiltonian which is meant to be a direct representation of a microphysical model. It might not be perfect — no-one really believes that the

nearest-neighbour Ising model is a particularly accurate representation of the microstructure of ferromagnets; but the point is that the model is meant to accrue any empirical success from its representation of the microphysical state, even if this representation is imperfect.

In this case, a recursion relation is designed around our microphysical model, and it can be shown that the flows induced give rise to a critical exponent with a certain numerical value. This is a perfectly good explanation of the non-classical values of these exponents for each particular model in turn, and it is firmly microphysically based; but as we saw in §3.4, it cannot be an explanation of universality.

In the Wilsonian approach, the LGW Hamiltonian is an effective one, and though we again design the recursion relation around this Hamiltonian, we hope it will remain justifiable for a wide range of models. In particular, it will be justifiable to apply it to all those that differ from the LGW Hamiltonian only by including irrelevant operators. Wilsonian renormalisation thus provides a good explanation of universality: it explains why the critical exponents for a universality class take the non-classical values that they do, and it explains why the various microphysical models fall into their respective universality classes. (Though as we saw in §3.5, it is not "microphysically based" and so does not fulfil Batterman's criteria.)

Of course the two approaches differ a great deal in their technical details, and in their practical applications. But I think that the foregoing is a good way of looking at a key difference between the two procedures. In both cases, we apply a procedure of iterated idealisation, in the form of a recursion relation: but in the first case, it is applied to the Hamiltonian of a specific microphysical model; in the second case, to an effective Hamiltonian suitable for a wide range of microphysical models. The physical significance both of the procedure itself and the conclusions we can draw from it, are accordingly different.

### 3.7.3 Originality After all?

In its rejection of the original aspects of renormalisation, this discussion might come across as relentlessly negative. Yet I do not mean to play down in any way the insight or the practical power of RG techniques, nor that they allow a deeper understanding of the phenomena in question. I merely claim that the explanatory strategy they exemplify is not a new one to the philosophy of science, and *a fortiori*, cannot offer a radically new approach to explaining multiple realisability.

Here, the "iterated idealisation" view of renormalisation is useful once again. Since we can separate the renormalisation strategy cleanly from the theoretical models, we can turn from a fruitless search for an original explanatory strategy, and look to the significance of the models it is applied to. And here I think that the RG methods have introduced something quite new to the philosophy of science: the concept of a *renormalisable* theory.

In §3.5.4, we used RG analysis to show that a Hamiltonian such as the LGW model, with all the relevant operators associated with a certain fixed point, would allow us to predict the universal quantities, in complete ignorance of any further operators. This is a rather striking situation: we are presented with a theory that is empirically successful (over the scale of energies under which it has been tested), but its empirical success gives no reason to believe that it is close to an accurate representation of the system in question.

At first sight, this appears at odds with orthodox accounts of scientific representation, which typically demand some link of resemblance or isomorphism between a theoretical model, and the system to which it is successfully applied. In particular, the realist demand for a firm link between theory and reality appear threatened.

However, we can already see from the discussion that this is not quite correct. The analysis of the recursion relations is adjusted to the systems in question, even when the LGW Hamiltonian can be shown to be suitable for such a wide

range of microphysical Hamiltonians. We do not derive universal quantities "from the armchair", but by close analysis of the empirical features of the systems in question.

Similar issues have impressed some of the New Emergentist writers, who claim that they demonstrate a sense in which higher-level sciences are independent of the small scale. I shall address both these issues in more detail in Chapter 5.

# Chapter 4

# Phase transitions in Finite Systems

## 4.1 Introduction

At the close of his second Meditation, Descartes is diverted from intellectual sparring with his all-powerful deceitful demon to consider a more mundane issue. Can he clearly and distinctly perceive the nature of something so simple as a piece of wax? As he places it in his fire its sensible properties of colour, texture and solidity, undergo abrupt qualitative changes, yet it remains the same wax as before. He is forced to admit that none of those sensible properties could be in the nature of the wax itself, and thus he had not clearly and distinctly perceived it after all.

In picking his example, Descartes chose well. For there are few everyday phenomena that have confounded analysis as consistently as have phase transitions. It was only late in the last century that physics made good progress in understanding many aspects of these changes: how a single collection of molecules can rearrange themselves in such a way that wax can melt, a liquid become a gas, or a piece of iron can become magnetic, while remaining the same materials throughout. It

has proved possible to build on existing accounts of many-body systems — statistical mechanics and thermodynamics — to give a comprehensive account of the changes. But this theoretical apparatus continues to provide surprises.

### 4.1.1 What is a Phase Transition?

There are clear definitions of what counts as a phase transition within theories of thermodynamics, and there are slightly different definitions within various approaches in statistical mechanics. But it is surprisingly difficult to characterise them in theory-neutral terms. In general, they are marked by abrupt changes in one or more large-scale physical properties of a system, with a small change in some control variable. Descartes' melting wax undergoes a phase transition when its density and other large-scale properties undergo a large change with a small change in temperature. And if we raised the temperature even more, the liquid wax would boil, and the density would undergo an even more dramatic sudden change. We see similar changes in the large-scale magnetic properties of a piece of iron, when we cool it below the Curie temperature and apply a small external magnetic field.

Once we consider phase transitions from within a particular physical theory, we can be a great deal more precise, and can start sorting them into types. Thermodynamics and statistical mechanics both represent phase transitions by non-analyticities in the free energy, as a function of one or more of its thermodynamic variables.

The Ehrenfest categorisation of phase transitions originally took advantage of this definition by grouping phase transitions by the lowest derivative of the free energy which undergoes a discontinuity. Thus first-order transitions (which include melting, boiling and the reversing of magnetic polarisation) have a discontinuity in the first derivative of the free energy with respect to a thermodynamic control

variable. Second-order transitions include those at the Curie point of magnets, where the magnetization itself increases continuously from zero, but the magnetic susceptibility undergoes a discontinuity at the Curie tempature. (The magnetization is the first derivative, the susceptibility the second derivative, of the free energy with respect to the applied field.)

The Ehrenfest classification has now been rejected, since its underlying rationale was based in mean-field theory, which has been superseded. Also, it has been found that some of the more complex phase transitions do not fit comfortably into its categories. However, modern approaches retain some of those original distinctions. They begin by sorting phase transitions into two groups: first-order (as with Ehrenfest) and continuous (all the higher-order Ehrenfest). These continuous transitions are then further subdivided according to the symmetries of the phases each side of the transition.

### 4.1.2 Two Issues

Recently, a debate has arisen in the philosophy of physics literature on the best way to understand the physical modelling of phase transitions.[1] The controversy centres on the fact that non-analyticities in the free energy of a system are central to the theoretical account of phase transitions, and yet statistical mechanics can only accommodate non-analytiticies in a system with an *infinite* number of degrees of freedom — while physical systems such as Descartes' wax, or a boiling kettle of water are surely finite. In my view, the debate comprises two very separate issues, but the literature has tended to have address them simultaneously, and without distinguishing them clearly.

The first issue is about the relation of theory to real physical systems. The most successful treatments of phase transitions can deal with them only as features of

[1] In particular, see Liu (1999, 2001, 2004), Callender (2001) and Batterman (2004).

infinitely large systems, and make a central use of this infinite number of elements in order to derive their results. Naïvely, this seems to imply that we have no successful theory of the phase transitions we see around us, since they occur in finite systems. We need an account of how our existing theories may be related to the concrete systems observed in the world. I shall call this the "Idealisation Problem".

A second issue is about the relation between two different theoretical approaches to phase transitions: those based in thermodynamics and those based in statistical mechanics. This second issue can be viewed as a special case of the question of the general relation between those two theories: in what sense, if any, can one be reduced to the other? It has been argued that their respective approaches to phase transitions pose particular difficulties; so I shall call this the "Reduction Problem".

I feel that the Idealisation and Reduction problems have been conflated in the existing literature, perhaps because of a belief that statistical mechanics uses non-analyticities to represent phase transitions *only* because of some need to follow the lead of thermodynamics in doing so. Of course such a requirement would lead immediately to a requirement for infinite systems. But as we shall see, putting the issue this way is extremely misleading. For the main reasons for representing phase transitions as non-analyticities are internal to statistical mechanics, and are independent of any need to capture the treatment of phase transitions by thermodynamics. In the rest of this section I shall consider the Reduction Problem and argue that it holds little interest on its own. The remainder of the chapter will then be concerned with the Idealisation Problem, which cannot be dismissed so easily.

### 4.1.3 The Reduction Problem

Chuang Liu argued in a 1999 paper that phase transitions provided an example of a genuinely *emergent* phenomenon, in the sense that the treatment provided by thermodynamics cannot be reduced to the treatment of statistical mechanics.[2] In thermodynamics, phase transitions are represented by surfaces, lines and points in the value-space of the relevant thermodynamic variables, at which one or more of these variables is not analytic. Liu judges that 'a rigorous account of phase transitions in purely thermodynamical terms encounters no real conceptual problems' (Liu, 1999, S93). But he also claims that this representation of phase transitions by non-analyticities is 'a feature that any micro-explanations of phase transitions must recover', and goes on to argue that statistical mechanics was wanting in this respect, since under some rather weak assumptions, such non-analyticities cannot appear in a finite system.

Liu notes that the problems can be avoided if we consider statistical mechanics in the 'Thermodynamic Limit' (TD Limit), which involves taking $N$ — the number of particles in the system — to infinity, while carefully preserving their overall density and the interactions between them. In this infinite limit, non-analyticities may appear in the quantities provided by statistical mechanics. But, he contends, the recovery of the thermodynamical representation in statistical mechanics still fails for finite systems. In keeping with his general assumption that a failure of reduction is a signature of 'emergence', Liu calls phase transitions 'emergent phenomena'.

Craig Callender cut off Liu's argument at its first stage, his rebuttal being an instance of the overall thesis of his 2001 paper: that we are often guilty of 'taking thermodynamics too seriously'. In this case, the fault comes when Liu lifts the

[2] In the terms of Chapter 1, Liu takes emergence as defined by a failure of inter-theoretic reduction, an approach considered in §1.5.

definition of phase transitions as non-analyticities out from its thermodynamical context, and demands that statistical mechanics must also treat them this way. Callender denies that statistical mechanics has any such obligation and objects to 'this knee-jerk identification of mathematical definitions across levels' (ibid, 550). Rather, he holds that statistical mechanics must be allowed to deal with phase transitions in whatever way seems fruitful by its own lights.

On this particular point, one must surely agree with Callender. By demanding that the same definition be used in each theory, Liu's argument assumes an over-simple view of theoretical reduction. And in any case, the reduction of thermodynamics to statistical mechanics in the classic Nagelian sense has long been recognised to be a project fraught with difficulty.[3]

If anything, it is surprising that statistical mechanics can even approximately reproduce any part of the thermodynamic representation of phase transitions, whether in the TD limit or otherwise.[4] For phase transitions are among the most subtle and intricate phenomena treated by modern physics, and the hard-won theoretical advances have come mainly in statistical mechanics, where they have far outstripped the 'classical' thermodynamic approaches. For example, the phenomena associated with spontaneous magnetisation and supercooled fluids are now sizeable research areas, and neither can be properly understood without a sophisticated statistical mechanical treatment.

Since it goes so far beyond the treatment of thermodynamics, it seems clear

[3]Nagel (1961) is the classic statement of inter-theoretic reduction, though it has been much-criticised. Sklar (1993) gives a comprehensive survey of problems associated with the thermodynamics/statistical mechanics reduction. And see §1.5.1 for more details and further references.

[4]Amongst other simplifications, taking the TD limit often quenches thermal fluctuations, which greatly helps the identification of quantities between the two theories. This disappearance of fluctuations in the TD limit might appear to conflict with its use in association with phase transitions. At critical points, it is the presence of fluctuations in the order parameter of the system that drives the distinctive critical behaviour. In fact, in these cases the fluctuations are not quenched by the TD limit, because the correlation length association with the fluctuations also becomes infinite.

that statistical mechanics should be free to develop its own definition and analysis of phase transitions. It would be a ridiculous and artificial restriction to insist on a strict adherence to an old definition for the sake of a reduction project that is already in serious trouble. As such, I shall assume from now on that the Reduction issue may be left aside. When developing new, fruitful methods of dealing with phase transitions within statistical mechanics, or in other frameworks, we should feel no pressure to follow thermodynamics in representing them as non-analyticities.

## 4.2 The Need for the Thermodynamic Limit

Despite the claims expressed in the last section, the fact is that almost all of the sophisticated and successful statistical mechanical treatments of phase transitions *do* make use of the TD limit, and they *do* represent phase transitions as singularities. The most successful approaches include mean field theories, Landau's approach, Lee-Yang theory, and renormalisation treatments — all of which use non-analyticies. So even if we follow Callender in dismissing the Reduction issue, we are still confronted with a serious problem. For regardless of how thermodynamics represents phase transitions, statistical mechanics still represents them as non-analyticities in the free energy. And as already mentioned, these non-analyticities do not appear in the free energy of a finite system: only infinite systems can accommodate them. The Idealisation problem is how to relate these accounts to the systems that we see in the world, since they appear to have a finite number of degrees of freedom.

### 4.2.1 The Idealisation Problem

I shall now try to set out a clearer statement of the Idealisation problem, and thus of what would count as a satisfactory solution. I will structure the discussion around three questions:

- A Motivation: Why is the thermodynamic limit so useful for theories of phase transitions? Is it indispensible?
- A Definition: What constitutes a phase transition in a finite system?
- A Justification: If our theory of phase transitions does make an ineliminable appeal to the infinite nature of a system, what justification do we have for applying the theory to finite systems?

In practice, physicists do not worry about these questions, mainly because as a matter of empirical fact, quantitative predictions derived for infinite systems hold very accurately for finite systems of "laboratory scale". But it is unsatisfactory for philosophers of physics to merely note as a surprising fact that analyses can be used to model systems for which they are provably inapplicable. We need an account of how this comes about. There has been some interest in addressing this question from philosophers of physics, and as several authors emphasise,[5] our difficulty does look more serious than the *general* worries about relating clean, idealised systems that the theoretricians can consider, to the messy, realistic ones that the experimentalists must observe. For the definition of a phase transition is an all-or-nothing singularity in the free energy, which in no clear sense can be "approached" as $N$ becomes very large. And it is important to realise that the theories really do require a genuine singularity; vague appeals to "steepness" or an "extreme

[5]Examples are found in Callender (2001, 550), Liu (1999, S100), and Batterman (2004, 13-14). (Batterman is especially concerned with the *singular* limits that appear at continuous phase transitions.)

gradient" will not do. For we can find finite systems with extreme gradients in the relevant thermodynamic variables which do not become a singularity as the TD limit is taken: these do not represent phase transitions.

The issue I called 'Motivation', will be addressed in the remainder of this section, where we shall look at the Lee-Yang theory of phase transitions. Using this example, we will see that existing theories do make essential use of the TD limit, but also that there is no barrier in principle to alternative theories that do not appeal to it. Having seen how theories of phase transitions use the TD limit, we can address Craig Callender's suggestion 2001, 547-552 that there is an outright contradiction amongst four statements that could plausibly be made about such theories (§4.3).

Drawing on the discussion of these points, we will move to the second question of 'Definition', which appears the most urgent; for without a clear definition of phase transitions which applies to finite systems it is difficult to claim that we have a theory of phase transitions at all. In §4.4, I propose a definition, which allows us to address the third question, of 'Justification': how we can justify applying our best theories to finite systems. Finally, in §4.5 we apply some lessons drawn from the classical discussion to a suggestion of Laura Ruetsche (2003), that quantum statistical mechanics can provide guidance of how to interpret quantum field theories.

So to start with, we can look at an example of a statistical account of phase transitions as non-analyticities. As already mentioned, mean field theory, Landau's techniques and renormalisation methods each make use of the TD limit, but perhaps the clearest example of the ineliminability of the infinite nature of the models is to be found in Lee-Yang theory (Yang and Lee, 1952). The details are found in standard textbooks (e.g., Thompson (1972, 85-89), Reichl (1980)), and only an outline sketch need be given here.

### 4.2.2 The Lee-Yang Theory of Phase Transitions

Although the Lee-Yang theory is a very general approach, we can illustrate it with a concrete example. Consider a simple model system of a large, but finite number of spins in thermal contact with an external reservoir. The total energy $E$ can take values $n\epsilon$, where $n = 0, 1, 2, ..., N$, where $N$ is the total number of energy levels and $\epsilon$ is the interaction energy between the spins. If we write the number of microstates corresponding to the $n$th energy level as $g(n)$, the canonical partition function is given by:

$$Z_N(\beta) = \sum_{n=0}^{N} g(n) e^{-\beta n \epsilon} \tag{4.1}$$

where $\beta = \frac{1}{kT}$, the inverse temperature. We make the change of variable $z \equiv e^{-\beta\epsilon}$, which allows us to factorise the polynomial

$$Z_N(z) = \sum_{n=0}^{N} g(n) z^n = \kappa \prod_{r=1}^{N(V)} \ln\left(1 - \frac{z}{z_n}\right), \tag{4.2}$$

where $z_n$ are the $N$ zeroes of the partition function, and $\kappa$ is a constant which we will ignore in what follows. Since all coefficients in Equation 4.2 are positive, the $z_n$ will lie away from the physical values of $z$, which are on the positive real axis.

To analyse the locations of the zeroes in more detail, we define the complex generalization of the free energy:

$$h_N(z) \equiv \frac{\ln Z_N(z)}{N} = \frac{1}{N} \sum_{n=1}^{N} \ln\left(1 - \frac{z}{z_n}\right) \tag{4.3}$$

and note that a Taylor expansion of $h_N(z)$ around a point $z \neq z_n$ has a finite radius of convergence, given by the distance of the nearest zero from $z$. Therefore $h_N(z)$ can be differentiated infinitely many times in any region that does not contain any

zeroes, which means that within such a region there will be no non-analyticities in the partition function nor the free energy. Therefore, for any point $z_0$, there can be a phase transition only if there is a zero in a region arbitrarily close to it on the complex plane. For finite $N$, there will be a finite number of zeroes, so there will be no phase transitions except at the points $z_n$ themselves. And we have already seen that these lie away from the physical values of $z$ on the positive real axis.

We now want to see how a phase transition might develop in the limit of infinite $N$. Here we assume that the limit:

$$h(z) = \lim_{N\to\infty} \frac{\ln ZN(z)}{N} \tag{4.4}$$

exists, and that we can write

$$h(z) = \int dz' \rho(z') \ln\left(1 - \frac{z}{z'}\right) \tag{4.5}$$

where $\rho(z')$ is the local density of zeroes in the complex plane.[6]

Analysis of this expression can show us where the local density of zeroes is positive. For systems such as our simple model, we obtain a curve $C$ that "snips" the real axis.[7] And so at the region where the curve meets this axis, we have the possibility of a non-analyticity, and therefore a physical phase transition.

What is more, the Lee-Yang method provides a classification of phase transitions, according to the line density of zeroes along this curve $C$. First-order phase transitions are characterised by the line-density remaining non-zero where $C$ crosses the real axis. But for continuous phase transitions the line density goes

[6]Ruelle (1969) examines rigorously and generally the conditions for the existence of the TD limit.

[7]Thompson (1972, 85-9) provides an analysis of the shapes of $C$ for some realistic models. Blythe and Evans (2003) give a pedagogical account of Lee-Yang theory, before considering how the techniques can be extended to nonequilibrium cases.

smoothly to zero as the curve approaches the axis, and we find no discontinuity in the first derivative of the free energy (though there may be discontinuities in higher derivatives).

This sort of analysis of the density of zeroes in the complex plane, and thereby the nature of the singularities in the free energy, tells us a great deal about how the derivatives of the free energy behave in the TD limit. Since in statistical mechanics, macroscopic quantities are obtained from derivatives of the free energy, this in turn yields information about the properties of a substance as it approaches a phase transition. These predictions can be tested against experiment (albeit on a finite system!) and are found to be remarkably successful.

Yet the Lee-Yang analysis also shows clearly that a phase transition — defined as a discontinuity in the free energy — cannot appear in a finite $N$ system of the type considered. What is more, this lack of non-analyticity can be shown under some fairly weak assumptions, which cover a large variety of systems.[8] Yet all the melting and boiling we see around us occur in finite systems. This seems to doom our theoretical account to irrelevance, since it provably does not apply to the phase transitions we observe.

### 4.2.3 Some Roles of the Thermodynamic Limit

If we accept (from §4.1.3) that it is not just from a desire to reproduce the thermodynamic analysis that we model phase transitions as non-analyticities in the TD limit, we must look for other motivations. Discussions tend to focus on a single motivation as *the* reason for its use. But choices vary. Of the authors mentioned so far: Callender focuses on several mathematical conveniences that it allows,[9]

[8]Emch and Liu (2002, Ch.12) have a detailed discussion with a rich resource of references. It should be mentioned that some mean field approaches do evade the assumptions that go into such proofs. Nevertheless, most MFTs do still apply the TD limit in any case.

[9]Callender (2001, 549-552). He also makes some puzzling comments about the vanishing of thermal fluctuations in the TD limit. While the neglect of these fluctuations may be important

Batterman concentrates on its provision of an analysis of singular limits,[10] and Liu on the mathematical rigour it provides.[11]

I feel that settling on one exclusive choice of motivation is a mistake: rather, the TD limit plays several distinct roles in theories of phase transitions. Here I shall separate three, and though I make no claim to comprehensiveness, it appears that much would be lost by ignoring any of them.

### Mathematical Convenience

Those who hold that the TD limit is *practically* essential, but foundationally unimportant, tend to motivate its use from mathematical convenience. They grant that physicists use the thermodynamic limit, and that they appeal to the singularities that arise there, but hold that this is just because it makes their calculations easier. In the first place, it is often easier to cope with infinite sums than large-but-finite ones, and it is often possible to simplify matters further by replacing sums with integrals. It is also possible to identify non-analyticities even when we cannot provide exact solutions; the appeal to the line-density of zeroes we saw in Lee-Yang theory is just one example of the powerful geometrical and topological techniques that can be brought to bear.

If mere mathematical convenience were the only reason for the deployment

for the Reduction problem, their significance is not so clear for the Idealisation problem. For in the case of critical phase transitions it is these very fluctuations that grow in range, cause the failure of mean field theory and necessitate the use of RG techniques — but as we saw in Chapter 3, both the demonstration of this failure, and the RG techniques themselves, also use the TD limit.

[10] Batterman (2004): a focus I think misplaced for the reasons elaborated in §4.3.2.

[11] In Liu (1999) he focusses on the need to rigorously derive singularities, and concentrates on the neglect of fluctuations, though here he seems motivated by the Reduction issue. In a later article (Liu, 2001) he concentrates more on the 'accentuation or exaggeration of the corresponding physical properties by neglecting or filling out negligible differences' by which I take him to mean what I call *seclusion* below. A joint paper with Gerard Emch mentions its use in discarding surface effects (Emch and Liu, 2005), and their jointly authored book separates out several more examples of mathematical and physical motivations (Emch and Liu, 2002, 394-6).

of the TD limit, the whole issue would be of little foundational or philosophical interest (except perhaps as a vivid example of the powers of idealisation, or of the difficulties in inter-theoretic reduction). All that would be required would be a demonstration that a theory using the TD limit gave the same result as the "inconvenient" analysis of the finite case. Experimental success would go a fair way towards such a justification, and for more reassurance, we could compare some 'toy' cases where both finite and infinite results are obtainable. But we are not in this situation; our problem is not that theories of finite phase transitions are mathematically inconvenient or impractical, it is that they do not exist.

**Seclusion**

A very different set of motivations stem from a need to isolate and separate out distinct phenomena occurring simultaneously at the point of a phase transition. We analyse phase transitions by examining the partition function or free energy of a system, but these functions will be made up from many contributions: some related to the effects of the phase transition itself, but many of them unrelated. We want to remove as many irrelevant contributions as possible, so that the phase transition itself can be recognised apart from extraneous effects. I shall call this practice 'seclusion.

This second motivation has long been recognised in the philosophy of science as a species of idealisation distinct from motivations of mathematical power. It has been with us at least since Galileo's separation of the horizontal and vertical components of a projectile in order to exhibit separately its accelerated and inertial motion. (It may well be older, but McMullin (1985) argues that Galileo was novel in applying the technique, which he calls — perhaps misleadingly — 'causal idealization').

One good example of seclusion is that taking the TD limit automatically dis-

cards complicating features such as 'surface' or 'edge effects'. Consider a lattice system of finite extent with contributions to the partition function coming from all sites. The contributions of those at the edges and surfaces will be markedly different to those nearer the centre of the sample, since any short-range interactions will be affected due to their different arrangement of neighbours. In the TD limit, the whole body can be treated as 'bulk', and the surface effects vanish.

We shall see later (§4.4.3) how "cross-over" effects can separate different categories of phase transitions in the TD limit, and many other seclusion effects appear there. It is in infinite systems that we can distinguish the features that are definitive of phase transitions, as well as distinguishing different types in a systematic classification. However, this does not *imply* that a theory dealing principally with the infinite case should be inapplicable to finite systems: it merely suggests that some work needs to be done to fill the gap; and of course, that is one of the problems with which we are concerned.

#### Structure

A third motivation for considering the TD limit is that it introduces new mathematical structure to our theories: a structure which plays a more fundamental role than "merely" increasing convenience. Rather, it introduces qualitative distinctions which would not otherwise be available and may be essential to give an adequate theoretical. For example: Robert Batterman insists that theories must recognise the existence of what he calls 'physical discontinuities' (Batterman, 2004) at phase transitions. This demand can be interpreted as requiring that the clear qualitative distinction between physical phases must be reflected directly by some clear qualitative distinction in our mathematical representation of different phases. And such a distinction is not immediately available in representations of states of

finite systems.[12]

As we have seen, the Lee-Yang theory appeals to a structure which emerges only as the TD limit is taken — the line-density of zeroes in the complex plane. Other theories provide the required structure in very different forms. Perhaps the most striking examples appear in algebraic approaches to quantum statistical mechanics, where distinct phases of an infinite system are represented as unitarily inequivalent representations of the algebra of its observables (Sewell, 1986, Ch.2). Yet for a finite system, the Stone-von Neumann theorem assures us that all such irreducible representations are unitarily equivalent. So again, without the TD limit our theoretical account does not provide enough structure to distinguish more than one phase. We shall return to this in some detail in §4.5.

Again, it is perfectly possible that this extra structure can be supplied in some other way than by introducing an infinite system. For example, Lev Landau's approach to phase transitions involves introducing an *order parameter* that expresses characteristic features of each phase, and it seems that one could appeal to attributes such as its sign (when a scalar) or direction (when a vector) to express a qualitative difference between phases. In fact, Landau's approach does use the TD limit and appeals directly to non-analyticities, classifying phase transitions by the lowest order of derivative of the order parameter that contains a discontinuity (Landau and Lifshitz, 1959, Ch.8), but there seems no *a priori* reason why the alternative would not work.

Each of these three roles — convenience, seclusion, and structure — are significant in themselves. While they appear independent of one another, they are all satisfied by the idealisation of taking the TD limit. Of the three, the first appears to be at most a motivation of 'mere' practicality, (though for a working physicist

[12]But see §4.3.1, where we reject Batterman's further claim that these physical discontinuities must be represented by mathematical discontinuities.

it is of course quite sufficient in itself.) The second and third appear of more foundational importance, and I shall focus on them in what follows. However, it is important to emphasise that none of the roles seems to *require* the infinite limit — alternative approaches are not ruled out. But it is striking that the TD limit is able to fulfill all three roles, and to do so naturally. More significant for what follows, a demonstration that a *single* role can be filled without the TD limit is *not* a solution to the Idealisation problem. We require alternatives for all three.

## 4.3 An Apparent Paradox

Craig Callender puts the Idealisation problem in sharp relief in his 2001 paper by presenting four jointly contradictory statements about phase transitions, reproduced (and re-ordered for later convenience) below:

1. Phase transitions are governed/described by classical or quantum statistical mechanics (through $Z$).

2. Real systems have finite $N$.

3. Phase transitions occur when the partition function $Z$ has a singularity.

4. Real systems display phase transitions.

Once we add the theorem that a partition function $Z$ of a finite system cannot display a singularity, the four are contradictory. And it is possible to classify attitudes to the nature of phase transitions by looking at which of these statements are denied. In the rest of this Section, I shall examine the consequences of denying each of the statements on Callender's list, and whether doing so can be counted a satisfactory solution to the Idealisation problem. Surprisingly (perhaps), it is possible to find advocates for rejecting of each one of his statements.

### 4.3.1 Deny Statement 1: Declare that phase transitions are not governed/described by classical or quantum statistical mechanics

The most straightforward approach is to take our apparent paradox at face value. Ilya Prigogine points to it as yet another failure of reductionistic approaches to complex phenomena: phase transitions are 'emergent properties', a term he uses in a very strong sense, to mean that they are not derivable from known laws of quantum mechanics or classical physics. Accordingly, he holds that phase transitions are insufficiently described by the theories we have been considering, and holds that genuinely new law-like behaviour appears in systems large enough to exhibit such phenomena (Prigogine, 1997, 45).

Since Prigogine has long been engaged in a research project to discover emergent laws across many areas in physics, he draws this conclusion with a great deal more readiness than would most physicists or philosophers. I feel that this straightforward denial of the scope of present physical theories is an extraordinary move to make, given that our paradox started from the striking *success* of statistical mechanics in modelling phase transitions. Our problem is to understand the relation between the infinite methods it uses and the finite systems it models so faithfully. This problem is not touched by trying to deny that any such link exists.

Both Batterman and Liu can be read as considering rather less extreme variations on this position. Liu's conclusion in his most recent article (2004) is that neither the Ising spin models nor the TD limit are 'realistic' in that they do not accurately represent the structure of real systems. But these two deficiencies somehow compensate for one another: 'It seems that by introducing two radically unrealistic idealizations — the Ising lattice and the thermodynamic limit — one is able to do better [than a more realistic approach]' (Liu, 2004, 256). However, it

is not entirely clear that Liu equates the 'unrealistic' treatment of statistical mechanics with a straightforward failure of description. It is equally possible to read his statement as marvelling that such an *a priori* unpromising approach gives such accurate results. In any case, Liu gives no explanation of how the compensation is supposed to work, and however we interpret his position, he gives no clue as to how the theories achieve their empirical success.

Robert Batterman suggests that the existence of phase transitions forces us to accept the existence of what he calls 'physical discontinuities'. These are physical quantities which really do undergo discontinuities 'out there' in the world. So he interprets the singularities suggested by the mathematics in a very direct way. As examples he offers the qualitative difference between phases, and also one which he discusses in greater detail, the breaking of water into droplets. In both cases, he holds that the singularity is 'physical'.

> ... it surely does seem very plausible to describe the breakup of water into droplets as a genuine physical discontinuity. It is true that we do not see the topological change in the phase transition (say when we witness water boiling in a tea kettle) in the same way we see a stream of fluid break apart. But that, by itself, does not show that there is no genuine physical discontinuity in the thermodynamic system.
>
> My contention is that thermodynamics is correct to characterize phase transitions as real physical discontinuities and it is correct to represent them mathematically as singularities. Further, without the thermodynamic limit, statistical mechanics would completely fail to capture a genuine feature of the world. Without the thermodynamic limit, in fact, *statistical mechanics is incapable even of establishing the existence of distinct phases of systems.* (Batterman, 2004, 12-13, italics his)

Batterman appears to be siding with Prigogine in insisting that statistical mechanics is actually *wrong* in failing to accommodate phase transitions as physical distincontinuities in finite systems. For he seems to claim the following: a physical discontinuity can be 'seen' as the density of water boiling in the tea kettle

instantaneously jumps from one value to another, and so statistical mechanics is inadequate if it fails to represent this discontinuity in any finite system.

I admit to some uncertainty as to how to interpret Batterman's 'physical discontinuities', and it is helpful to disambiguate his thesis into two claims, one markedly stronger than the other. The weaker claim is that there are important features of the physical world that are represented by discontinuities (or more precisely, by non-analyticities) in the mathematics used by our best theories. Since our ability to represent these features would be lost if we gave up using the TD limit, the limit is essential. The stronger claim adds to the first that the features represented by the mathematical discontinuities are 'real physical quantities' undergoing 'real physical discontinuities'.

In so far as I understand Batterman's terminology, the stronger claim appears untenable. For it is possible to recognise the existence of distinct phases and the transition between them without being forced to accept that any physical quantity undergoes a discontinuity. The most straightforward alternative would be to recognise that phase transitions take a finite time, during which the water belongs in no definite phase. This could be implemented in a variety of ways, but one would be the following. Let us consider a finite sample of water, and examine its density as a function of time as it boils. Certainly, there is a range of densities over which we are happy to call it a liquid, and a range of densities over which we happy to call it a gas, and further that there is a discontinuous 'gap' between these two ranges of densities. We can certainly recognise a 'physical discontinuity' in this sense. However, this does not mean that there is any particular sample of water whose density must 'jump' across this gap instantaneously when the water boils.

Further details would vary according to inclination, but one option would be to hold that for the sample of water as a whole, neither the terms 'liquid' and 'gas'

are relevant. Alternatively, we can hold that there is local variation while boiling, with some areas being liquids, some gases (until we get down to some scale, at which point these terms fail to apply).

I should clarify that I see no *a priori* reason to deny the possibility of physical discontinuities. Indeed space, time or any other physical quantity may be discrete, in which case they might be very common indeed. However, their existence should not be demonstrable by observing — say — water boiling in a tea kettle. We need an account of phase transitions that would accommodate this behaviour, even if the world contained only continuous physical quantities.

If we reject Batterman's stronger claim, then we are still free to consider the weaker one on its own. His claim then is that without discontinuities, the orthodox statistical mechanical techniques fail to 'capture a genuine feature of the world'. The weak claim leaves aside the issue of whether these 'genuine features' are themselves discontinuities in physical quantities, but they are certainly *something* physically significant, and they need to be represented by any empirically adequate theory. The most obvious way to implement this would be to replace the orthodox procedure for constructing a partition function for a finite system. Rather than sum over a finite number of degrees of freedom, we replace any finite sums by infinite ones. This approach might be in keeping with the actual practice of physicists, but it leaves unanswered all the mandatory questions identified in §4.2.1. Namely: How can it be justified? What definition can we give for a phase transition in a finite system? As such, Batterman's weak thesis on its own is not an answer to the Idealisation question, and we need to choose another path.

### 4.3.2 Deny Statement 2: Declare that real systems do not have finite $N$

A second way of dealing with our problem is to hold that, contrary to appearances, real systems really do have infinite $N$. That is, physical systems undergoing phase transitions either have an infinite number of parts, or in some other way acquire the infinite degrees of freedom necessary to support a discontinuity in the free energy.

A direct approach would be to point out that since no real system is completely isolated, the modelling of a finite system with one of infinite extent reflects the fact that the system is always coupled to the rest of the universe.[13] Emch and Liu (2005) can be read as advocating such an approach, with the justification that it is the "least of all evils". They point to the infinitely insulating walls or infinite distances that are implicit when we consider a perfectly isolated finite system, and conclude: 'unless specific surface or boundary effects are of interest, taking the infinite limit is (from a philosophical point of view) more sensible than not taking it'. I would suggest that only from a *practical* or *physical* point of view is it more sensible. From a philosophical point of view, taking the infinite limit represents a far more radical step. For the idealisations involved in postulating completely insulating walls involve neglecting small external perturbations that become smaller as we consider increasingly isolated systems. This limiting procedure is in contrast to the singularities, which are not 'approached' in any simple way as the system becomes larger. In any case, it is not an acceptable defence of a position merely to point out that an alternative also has its problems.

[13] If we want to reserve judgment about the size of the universe (not to mention considerations of locality), an alternative route to the same goal would to claim that any system can acquire an infinite number of degrees of freedom from interaction with a continuous field of some kind. Classically, this could be the electromagnetic field, but we could also point out that at a quantum level, all matter ultimately has a field-like nature.

More generally, it is hard to see how such suggestions can provide a satisfactory solution to the Idealisation problem. For example, to provide a definition of a phase transition we do not merely need to show the possibility of non-analyticity, but also that non-analyticity appears at the points of phase transitions, and *only* at those points. Coupling the system to a large number of degrees of freedom makes discontinuities in $\mathcal{F}$ *possible*, but for an adequate theory, they must appear at the right places and times. For example: consider two kettles next to one another: in one the water boils, in the other it does not. We must guarantee that the non-analyticity is in some way associated with the first kettle. It cannot 'leak' over to the second kettle, since there is no phase transition occurring there. That is, our non-analyticities must be tied to the finite systems involved in each phase transition. It is difficult to see how such a satisfactory justification could be conjured from an appeal to coupling to the rest of the universe, or to some electromagnetic or quantum field.

But we can modify our denial of the finite nature of the systems to yield a more subtle suggestion. Batterman's weaker suggestion identified in §4.3.1, might be read as recommending such an adjustment. He might not actually doubt that real systems are made up of a finite number of particles, but he certainly demands that they be 'characterized' using an infinite $N$.

> I want to champion the manifestly outlandish proposal that despite the fact that real systems are finite, our understanding of them and their behavior requires, in a very strong sense, the idealization of infinite systems and the thermodynamic limit. (Batterman, 2004, 9)

Perhaps the best way to implement Batterman's proposal in Callender's terms would be to modify Statement 2 to: "Real systems have finite $N$, but to describe them, statistical mechanics must model them as infinite."

I have some sympathy with an alteration in this spirit, but have difficulty with the motivations that Batterman gives. He argues that the ineliminability of infinite

limits stems from situations in which the relevant limits are *singular*, that is: when a small parameter being taken to zero describes behaviour that becomes markedly different as the limit is approached.[14]

However, these singular limits that Batterman appeals to do *not* appear in all phase transitions, only in some cases such as critical points. There is little space here to discuss the relationship between non-critical and critical phase transitions, and to do so would be to wander from our main theme.[15] But singular limits cannot be the key to understanding why the TD limit must be used. For the theories which adequately represent phase transitions (both critical and non-critical) make a central appeal to the TD limit, whether the limit is singular or not.

Instead, let us agree with Batterman that the use of the TD limit is ineliminable, while reserving judgment on his diagnosis of singular limits as the root cause. This brings us back to the main problems, to definine phase transitions in finite systems and to justifying their application of our best theories. Batterman says little about the details of a definition, but the rest of his article is concerned with providing a justification for infinite idealisations. He illustrates these with a set of examples drawn from hydrodynamics. When a stream of water breaks off

[14]Batterman begins his article (§§1-3) by considering the relation between theories (the Reduction problem), and there he appears to use singular limits only as a metaphor for a new approach he recommends for intertheoretic reduction. However, his arguments are addressed to the Idealisation problem in the remainder of the paper, and there he appeals to singular limits in a literal sense.

[15]For more details, see the discussion of different varieties of phase transition in Chapter 3 (especially §§3.2 & 3.1.2) but a few brief comments seem in order. At a *critical* phase transition, a micro-structural property called the *correlation length*, $\xi$, also diverges to infinity. This $\xi$ is a measure of the range of the correlation of fluctuations. Since mean field theories neglect the correlation of such fluctuations, they fail near critical points (in less than four dimensions for the Ising model), and a mean field account of the critical phenomena is seen to be inadequate. Renormalisation introduces a systematic scheme for reducing the degrees of freedom, utilising a transformation that 'traces them out' while accounting properly for the diverging correlation length, and that does not fail near the transition, but instead gives accurate predictions for quantities such as the critical exponents. The main point for our purposes is that the failure of the mean field approximations is *not* directly connected to the necessity for the infinite limit, but rather to the diverging correlation length.

into drops, the shapes of the "neck" and of the drops are *universal*: they remain the same across many different kinds of fluid, which nonetheless differ in their microscopic structure. Batterman argues that if the microscopic details of the fluid make no difference to their large-scale features, it might be justifiable to replace a particulate body by a continuum.

For the special case of critical phase transitions, there appears to be hope of a similar approach. Some critical phenomena are also universal: the same large-scale effects occur in systems with widely varying micro-physical details. As we saw in Chapter 3, renormalisation techniques can provide an account of this universality, and could lead to an argument that a sufficiently judicious replacement of a finite spaced lattice with a continuum might leave untouched the large-scale physics that we observe in phase transitions.

But worries immediately arise. First, we do not seem to dispense completely with an infinite system. As we mentioned in §4.2, renormalisation techniques *start* with infinite $N$ models, and then demonstrate that the physics below a certain scale can be neglected.[16] This suggests only that details below a certain scale are irrelevant, not that the infinite nature of the system can be disregarded completely. A second concern is that the phenomenon of universality occurs only near critical points, not at all phase transitions; and even there, not all properties associated with phase transitions are universal. For example, the temperature at which a critical phase transition appears is *not* universal. As far as I understand Batterman's suggestions, his justification for 'screening off' small-scale physics would not hold for such properties, so there would be no reason to trust the predictions of the theory.

[16]The particularly interesting case of finite-size crossover is addressed in §4.4.2.

### 4.3.3 Deny Statement 3: Declare that phase transitions do not only occur when the partition function has a singularity

This is the course recommended by Craig Callender, and is also one of the options considered by Chuang Liu. Callender diagnoses our temptation to accede to the statement as stemming from our wish to directly import definitions from thermodynamics to statistical mechanics, but as he points out: 'the fact that thermodynamics treats phase transitions as singularities doesn't imply that statistical mechanics must too.' But in diagnosing this motivation, Callender is engaging only with the Reduction problem. Simply denying Statement 3 leaves the theory of statistical mechanics without a good account of finite phase transitions, indeed it leaves us unable to recognise their existence. Callender holds that finite phase transitions must be governed by analytic partition functions, which 'in some sense approximate a singularity', but gives little idea of what this sense is to be.

The most obvious move would be to assume that a sufficiently extreme gradient in the free energy could represent a phase transition. However, this cannot be the whole story. The Lee-Yang theory, in common with other treatments, requires a genuine discontinuity, not just an extreme gradient in the free energy. We can easily construct finite systems with extreme gradients in their free energy that do not develop into discontinuities when the TD limit is taken; these do not signify genuine phase transitions.

Liu (2001, 2004) is also worried by the difficulties presented in filling out Callender's 'some sense' of approximation, because he feels that this only makes sense when one can point to an asymptotic limit which ends at a phase transition. Liu draws a contrast with how we might justify 'applying calculus to *almost* continuous bodies (i.e., those which contain a large number of very small particles of finite

mass)'. He judges that we are correct to do so, because it is meaningful to say that $\delta M/\delta V$ approaches $dM/dV$, as $\delta M \to 0$.[17] Thus there is a clear sense in which one quantity approximates the other for 'almost continuous bodies'. However, this is in contrast to the case of the thermodynamic limit:

> Until the limit is reached, the pressure or free energy of any macrosystem is analytic. The non-analyticity is not at all asymptotically reached by the process of taking TL [the Thermodynamic Limit]. At no stage of this process is non-analyticity (representing a phase transition) roughly or approximately defined. (Liu, 1999, S103)

Here Liu brings up an important consideration, but a further distinction needs to be made. Consider Liu's 'almost continuous' body. There are at least two ways to regard his justification for applying calculus to it. The first is the one that Liu discusses: we consider the quantity $\delta M/\delta V$ for the particulate body, which can be directly interpreted as a physical quantity. Postulating a differentiable function $M \equiv M(V)$ and considering its derivative $dM/dV$, we can then show that $\delta M/\delta V$ is very well approximated by $dM/dV$ as $\delta M$ becomes very small. This provides us with a justification for making the substitution $\delta M/\delta V \approx dM/dV$ throughout our equations. In this first approach we aim to represent a realistic physical system, but then approximate some of the terms in our mathematics.

A second way is to argue that in regard to the behaviour of the body that we are interested in, it would be as well to consider a continuous substance rather than one made up from particles. We therefore consider this idealised physical situation rather than the original: modelling a continuous body rather than an atomistic one. In this case, the quantity $dM/dV$ is treated as representing a physical quantity, albeit one connected to a fictional, unrealistic situation. To justify this procedure, we need some argument that the substitute system will share the behaviour we are interested in. To make such an argument there is no sense in which the particulate

[17] Of course, this is not a rigourous approach to justifying the application of calculus to physical systems, but it serves to illustrate Liu's argument.

body need 'get asymptotically closer' to the continuous one. And this is so, even if we do appeal to the fact that $\delta M/\delta V \approx dM/dV$ for small $\delta M$.

This distinction is a well-known one, and versions have been drawn in many different contexts, and under many different names. Butterfield (2006a, 24-5) and Teller (1979, 348-9) agree in calling the first 'approximation' and the second 'idealization'. McMullin (1985, 264n) separates similar procedures under the headings: 'construct' idealization and 'causal' idealization.[18] I shall continue with Butterfield and Teller's terminology in what follows.

In many cases, the choice between idealisation and approximation makes little practical difference, because we often find that the justification we need to supply for each procedure is very similar. We have already seen in Liu's example that the small difference between $\delta M/\delta V$ and $dM/dV$ may be used to justify either approximation or idealisation. Or consider an almost force-free body such as a hockey puck moving on ice. We might either consider dropping some of the smaller force terms from our equations (approximation) or replacing the whole problem with consideration of a genuinely force-free one (idealisation). In both cases the equations we consider will be the same, and the justification for each comes from the negligible differences that these force terms make to the results we are interested in. For these cases, the idealisation/approximation distinction is of no importance to how a calculation proceeds.

But sometimes the two come apart: and in cases associated with the TD limit, the distinction is often crucial. Callender's 'in some sense' cannot be filled out by considering an approximation: it must refer to an idealisation. In other words: Liu's absurdity of attempting to 'approximate an analyticity with a non-analyticity' (whatever that might amount to), can be avoided if instead we idealise

[18]Other authors draw similar distinctions, Cartwright (1989) uses 'abstraction' for what I am terming 'idealisation', but she is also interested are interested in wider connotations, such as a distinction between "partially true", and "strictly false" idealisations. I would like to avoid these additional considerations.

a finite system with an infinite one. For it is legitimate to replace a finite system with an infinite one, and *then* to consider what sort of features of the finite systems might correspond to the non-analyticities that appear in the free energy of its infinite replacement. Of course we must face the questions of the physical significance of this procedure. Most importantly: what justification can we provide that the replacement will give us insight into the phase transitions of the original? I shall try to fill some of these gaps in §4.4.2.

Before we move to consider the next option, we must mention the specialist research programme studying phase transitions in 'small' systems, or those involving gravitational effects, where the TD limit cannot be applied. One advocate of this approach, Dieter Gross, argues that many first-order phase transitions are *not* best understood in the TD limit. He works with an alternative set of definitions, based around the statistical mechanics of the microcanonical ensemble (Gross, 2001). This is a far more straightforward denial of Statement 3 than we have so far considered: Gross simply rejects the whole non-analyticity approach, and suggests an alternative way in which statistical mechanics can model phase transitions. Unfortunately, it appears that these replacement definitions do not give rise to theories with as wide a range of applicability as the orthodox ones. So until Gross' programme develops further, his approach must be seen as a complement, rather than a replacement.

### 4.3.4 Deny Statement 4: Declare that real systems do not display phase transitions

Perhaps surprisingly, this is the option that appears to be taken by many physicists, even those who have made great contributions to theories of phase transitions. For example, Leo Kadanoff was the first to propose that the 'scaling' relations seen

in critical phenomena could be analysed using a 'blocking' procedure.[19] In his textbook on critical phenomena he writes:

> The existence of a phase transition requires an infinite system. No phase transitions occur in systems with a finite number of degrees of freedom. (Kadanoff, 2000, 238)

Similar statements can be found in many texts, usually with an associated plea that as the systems we see around us have very large $N$, they are 'effectively' or 'nearly' infinite. (Of course, as we discussed in §4.3.3, a discontinuity does not appear at *any* finite $N$, nor do we get "nearer" to one as $N$ increases.)

Perhaps even more striking is the attitude expressed in a text on the computer modelling of phase transitions (Mouritsen, 1984). These techniques model only a finite array of spins, and one might expect that theoreticians would welcome these as providing a more direct route to real physical transitions than do orthodox techniques, since they reproduce the finite nature of the physical systems. Not in the least; Mouritsen follows Kadanoff in declaring that phase transitions cannot occur in finite systems, and continues:

> Nevertheless, finite systems have *reminiscences* of phase transitions, and systematic [computer] studies of these pseudo-transitions as functions of system size may reveal information about the phase transition in the infinite system. (Mouritsen, 1984, 20, italics his.)

So even in those situations where it is mathematically convenient (indeed, necessary) to model a finite system directly, there is no suggestion that these models can show anything other than a 'pseudo-transition'. Instead, we look at our results as functions of system size, and attempt to extrapolate them to the infinite case. Only when Mouritsen is satisfied that the infinite case would exhibit a non-analyticity, does he consider that we really are dealing with a true phase transition,

[19] Chapter 3 gives more details of his proposals, and how his line of thought was later honed into the techniques of the Renormalisation Group by Kenneth Wilson.

and only then does he attempt to categorise it and draw further conclusions about its properties (Mouritsen, 1984, 22-26).

The fact that these theoreticians of phase transitions flatly deny that they take place in finite systems, may seem surprising. Presumably, they see kettles boil no less clearly than the rest of us. But the very fact that they feel that their denials are uncontroversial suggests that we are dealing with a matter of words only. The modelling of phase transitions with non-analyticities is a part of their theoretical toolbox, and the term 'phase transitions' has come to refer only to phenomena that fit this definition. Of course a change in reference of the term 'phase transition' does not solve the Idealisation problem, but just changes the form of words in which it must be stated. In usual terms the problem was to relate phase transitions in finite systems to a definition available only in infinite systems. Now we must state it as a problem of relating "nameless phase-transition-like phenomena" in finite systems to "phase transitions", now defined so as to only appear in infinite systems. Likewise, the theoretician's pleas that large systems are 'effectively' or 'nearly' infinite, translate to Callender's statement that the phase transitions must 'in some sense' approximate a singularity. The problems remain the same: the fact that we trust the infinite results to tell us about the finite sized phenomenon remains unchanged, and we still need an account of when this trust is well-founded.

## 4.4 Phase Transitions in Finite Systems

Let us review the issues still outstanding. We have a definition of phase transitions common to almost all successful theories of those phenomena. This definition, as well as those theories, apply only to infinite systems. Yet they describe well the phase transitions (or "phase-transition-like" behaviour) of finite systems, indeed

this is the main standard by which we judge their success. To avoid a simple contradiction, we could choose to deny any of the four statements presented in §§4.3.1-4.3.4: but by just doing so we may leave the important questions untouched. We have seen some reasons for the utility of the TD limit, but still require (i) a definition for phase transitions that we can apply to finite systems and (ii) a justification for applying the theories to them. In §4.4.1 and §4.4.2I shall address each of these, and then in §4.4.3, return to see how the answers cast light on how the TD limit is able to fulfill the roles we identified for it.

### 4.4.1 Defining Phase Transitions in Finite Systems

Consider a finite system with $N$ degrees of freedom in a particular state $S$. Let us denote its free energy as $\mathcal{F}_N(S)$ and its partition function as $Z_N(S)$. And assume that our system is one in that clearly delimited set for which there exists a well-defined procedure for taking the thermodynamic limit. The free energy of that system when the TD limit is taken we can call $\mathcal{F}_\infty(S)$ and the partition function $Z_\infty(S)$. (Naturally, $\mathcal{F}_N(S)$ and $\mathcal{F}_\infty(S)$ will coincide for a system that is already infinite). We have a well-defined criterion for phase transitions in infinite systems, so there is an obvious definition for the finite case.

**Definition 1.** *Phase transitions occur for a finite system in state $S$ if and only if $\mathcal{F}_\infty(S)$ has a singularity.*

Rather surprisingly, using this definition it is possible to hold on to all of Callender's four statements without contradiction; though only in a Pickwickian sense — it is a "trick" possible only due to his choice of wording. Namely, the singularity referred to in Statement 3 is one not in the partition function $Z_N$ but in $Z_\infty$.

I believe that this definition of a phase transition in a finite system is at least

naturalistically appropriate, in that it is in keeping with the practice of physicists. In particular, it accounts for their uncritical reliance on theories of the infinite to tell us about finite cases. For example, it makes some sense of the otherwise odd practice we saw in §4.3.4, of withholding judgment as to the nature of a phase transition until results are rigorously demonstrated in the infinite case, even if finite results are available.

But from a philosophical point of view the definition looks suspect, for it seems to make the existence of phase transitions a subjective, theory-dependent question, rather than something directly determined by physical facts about a physical system. This worry can be decomposed into two more sharply defined objections: first, that the existence of a phase transition is determined by a counterfactual (indeed, a highly unrealistic counterfactual); and second, that the existence of the phase transition depends on exactly how the system is modelled, or how the TD limit is taken — subjective matters that we can change on a whim. Let us take these in turn.

**Phase Transitions are Actual**

The thought behind the objection is that properties such as 'is boiling' appear to be an intrinsic property of a given sample of water. And as such, the facts we need to decide whether or not it is undergoing a phase transition should be physical facts, about actual states of affairs contained within — say — a kettle. They should not exist only in an idealised model on a theoretician's blackboard.

This objection is based on a version of a general principle often known as the 'Truthmaker Principle', roughly: that any actually true proposition should be made true by actual facts, real goings-on in the world. Put as a general claim, the principle is both disputable and lacks an agreed articulation (and this is not for a lack of attempts, e.g., Armstrong (1997, Ch.8) and Mellor (2005)). But I

shall not dispute the specific version appealed to here. For if we look carefully at Definition 1, we see that all the truthmaking facts *can* be located within the kettle, though we may need to appeal to counterfactuals to apply our detailed theory and decide whether the facts count as a phase transition (and if so, what type). That is, it is facts about the finite sample of water that determine whether or not a phase transition is occurring, and determine whether it is first-order or continuous. However, we find it necessary to appeal to a radically counterfactual circumstance to state these facts in any reasonable way.

There is a strong parallel between this situation and the defense offered by David Lewis in regard to his analysis of counterfactuals, which is held to contravene a similar principle. He defends his analysis, which is made in terms of a "similarity" relation amongst possible worlds, by suggesting that it is the qualitative character of the actual world which determines this "similarity" relation. However, he also defends the ineliminability of the other possible worlds: 'it is only by bringing the other worlds into the story that we can say in any concise way what [qualitative] character it takes to make what counterfactuals true' (Lewis, 1986b, 22). Lewis' memorable example was the following: while it is no doubt possible to describe Buenos Aires to a stay-at-home Australian by describing many intrinsic properties of Buenos Aires, it is so much more concise and illuminating to say 'It's like a Spanish-speaking Sydney'.[20]

Returning to the tea kettle, we should realise that it is the character of the finite sample of water that determine the nature of the infinite system that we then consider. When we draw conclusions about the nature of the phase transitions, they are conclusions about the character of the finite sample, but by reference to the infinite model we can express them in concise and illuminating form.

This consideration also blunts the worry that our definitions makes reference

[20]I thank Jeremy Butterfield for pointing out the parallels with Lewis' position. His 2003 paper contains an interesting note on the genesis of this particular example (24n).

to physically impossible circumstances. So long as we are satisfied that the truth-makers are intrinsic properties of the finite boiling water, the physical impossibility of the TD limit does not seem to pose any *further* problem than does the original counterfactual appeal. There is nothing paradoxical about the fact that our clear, concise expression makes reference to a physically impossible situation. Of course it must be conceded that an uncritical consideration of physically impossible counterfactuals might lead us to all kinds of absurd conclusions. We must be cautious, and this is part of the reason why it is so important to take the TD limit in a well-defined manner.

### Phase Transitions are Objective

While there exist standard procedures for taking the thermodynamic limit, and rigorous methods to determine whether it is well-defined,[21] these procedures are human inventions, and choices could have been made differently. Even when clear rules are laid down, theoreticians will always be willing to break them, if it proves fruitful for the analysis a particular situation. The definition of a phase transition thus seems arbitrary in a disastrous sense: we can choose whether one is occurring or not by modelling it differently, or taking the limit according a different scheme.

To allay this concern, it should be recognised that it was empirical considerations that have led TD limits to be taken in the way that they are. Over the last century, theories have developed to account for differences and similarities in the phase transitions we observe. We have been led to distinguish first-order vs. continuous transitions; and symmetry-breaking vs. non-symmetry-breaking transitions for example. The particular procedure for taking the TD limit has

[21] Ruelle (1969) is the definitive modern analysis of the conditions that allow a TD limit. Emch and Liu (2002, Chs. 11-12) provide a historical and philosophical discussion of the techniques, including the amusing fact that the necessity of the TD Limit was put to a vote at a 1937 conference. The result was not recorded.

developed along with the theories, in such a way that phase transitions are defined in good accordance with experiment, and the categorisation is faithful to empirical similarities and distinctions.

If it was found that a different set of methods for taking the TD limit gave a better classification of phase transitions, more faithfully accounting for the experimental observation, it would be adopted. There are good historical examples of such alterations: for example, we mentioned earlier how the Ehrenfest categorisation of $n$th-order phase transitions ($n$ being the lowest order of derivative of the free energy with a discontinuity) was discarded when it was found to exclude transitions where $\mathcal{F}$ diverges.[22]

One final difficulty must be mentioned. Strictly interpreted, the definition I have offered would allow phase transitions to occur in very small systems indeed. A lattice of four Ising spins laid out in a square might be said to have the infinite two-dimensional Ising model as their TD limit, and so could undergo a phase transition. Personally, I think this bullet can and should be bitten, but it can also be easily dodged. To do so, one might add an additional condition that the original finite system must be sufficiently large, or the gradient of $\mathcal{F}_N$ be sufficiently steep, before any change can qualify as a phase transition. This will make the attribution of phase transitions a vague matter in some cases, but we would have no problem producing clear cases on each side of the divide: four atoms of $H_2O$ would not be enough to boil, but a kettle full of it can.

### 4.4.2 Justifying the Application to Finite Systems

If the account presented in the last section is acceptable, then we have a definition of a phase transition in a finite system. We are now in a position to move to our last question: how can we justify applying our theories of infinite phase transitions

[22]For historical details, see Domb (1996).

to finite ones? A partial answer can be made by appealing to the design of the TD limit: for the whole point of the limit is to produce an infinite system that would tell us about the statistical mechanical behaviour of the original finite one. More support comes from a simple appeal to empirical success. But it would be advantageous to come up with further justification, and in some specific cases we can.

As an example, consider the phenomenon of *crossover*, an important feature of critical phase transitions. As substances approach a critical phase transition, they typically exhibit behaviour characteristic of one of a small number of *universality classes*. For example, beta-brass is placed in the "Three-dimensional Ising universality class", since it shows critical behaviour associated with that of the 3D Ising model. 'Crossover' happens when a substance appears to show behaviour characteristic of one universality class, but then suddenly changes to another as it is brought even closer to its critical point. One of the triumphs of Renormalisation techniques are that they provide a beautiful explanation of crossover, entirely lacking in older techniques such as mean field theories and Landau's approach.

I discussed renormalisation in more detail in Chapter 3, but we can get a feeling for the explanation as follows; (full and clear treatments can be found in good textbooks on critical phenomena such as Cardy (1996, Ch.4) or Goldenfeld (1992, 271-280)). A condensed matter system can be characterised by the parameters of its microscopic Hamiltonian: interaction strengths between molecules, couplings between an external field and the system, and so on. We can represent many different Hamiltonians by constructing a space coordinatised by all of these parameter values, with each Hamiltonian represented by a point in the space. (Typically, the temperature is also represented as a parameter in this space, so each point represents a system at a particular temperature.)

Renormalisation techniques allow us to examine the near-critical behaviour

of such systems by constructing an iterable transformation on this space. The transformation is designed to preserve the large-scale physics of the system, and the resulting system is rescaled so it can also be assigned a point within the space of Hamiltonians. The transformation thus induces a 'flow' through points in the parameter space which preserves all the large-scale physics of the system. There are certain *fixed points* of these transformations, and the properties of some of them (in particular, the *critical* fixed points), can be shown to affect the near-critical behaviour of any system whose flow comes close to them. Crossover occurs when a system's flow approaches one fixed point, but then as it is brought closer, it veers away and as a result, loses the behaviour characteristic of one universality class and takes on another behaviour associated with another.

At critical phase transitions, a quantity known as the *correlation length* diverges. A particular variety of crossover, known as *finite-size crossover*, occurs when the ratio of the correlation length to the system's size determines the fixed point to which the system flows. When the correlation length is small compared to the size of the system, the system's flow is attracted to a fixed point associated with the phase transition of an infinite system. However, as this phase transition is neared, the correlation length grows, becomes comparable to the size of the finite system, and then the flow crosses over to some different fixed point. The upshot is that as a parameter such as temperature is tuned so that a system approaches the critical phase transition, the system exhibits behaviour characteristic of an infinite system, but as the temperature is tuned more closely to the critical point, this behaviour disappears. Usually this is reflected in the 'smoothing out' of the gradients of quantities such as the free energy, so the singularities associated with an infinite system become peaks, which are tall and narrow — but analytic.

Finite-size crossover thus provides a very neat explanation for how a theory of phase transitions which only describes infinite systems, might model finite systems

to accuracies well within experimental error. For a correlation length will only approach a system's physical size as it is brought incredibly close to the phase transition itself, well within the tolerance of experimental measurement. And until we enter this regime the critical behaviour will be as though the system was infinite in extent. So powerful is this effect, physicists can be confident that experimental results on finite physical systems should match the theoretical ones taken in the TD limit.

We can give a measure of how striking this effect can be, when we consider the process of extrapolating the results of computer models of a finite number of spins to the TD limit. Computer models of systems as small as $L = 10a$ (i.e., ten spins per side of a square lattice,) have been shown to give results so close to the theoretical results obtained in the TD limit for 2D and 3D Ising systems (Mouritsen, 1984, 21) that little extrapolation is needed.

Here I should admit to some double-standards. In §4.3.2 I criticised Batterman for offering general accounts of phase transitions which apply only to specific cases. I admit that finite-size crossover also only provides an example of how we might justifying the application of our theories of phase transitions to near-critical systems — it is not a general prescription. And in fact, justifications tend to be made on a case-by-case basis. Where all else fails, large computer simulations of specific systems can give much reassurance that as $N$ gets large for a specific system, quantities such as the free energy converge to the values obtained in the TD limit. These cannot give general proofs, but can provide justification for the use of the theories on a model-by-model basis.

### 4.4.3 More on the Roles of the TD Limit

Back in §4.2.3, we identified three separate roles which the TD limit fulfils. The first, of mathematical convenience, is no doubt dispensible in principle, and with

vast increases in computing power, perhaps sometimes even in practice. But it is the other two roles that are of foundational significance: the need to isolate and *seclude* the phenomena we are interested in, and to provide the mathematical *structure* in order to distinguish them into their natural varieties and disregard any unrelated effects. As we shall see in this section, these are not merely practical considerations, and the TD limit fills both roles simultaneously.

Let us start with 'seclusion'. Real, finite, phase transitions are invariably hybrid beasts, with contributions to their partition function or free energy coming from many different effects, and even from different types of phase transition. A satisfactory theory of phase transitions must be able to tease these apart, provide a categorisation, and some understanding of how the effects may combine to produce the effects we see.

One kind of hybrid phase transition occurs in systems very much larger in some spatial directions than in others, and here we can appeal to another type of crossover phenomenon, known as *dimensional crossover*. When we apply the definition offered in §4.4.1, our first task is to take the TD limit, but in these cases it may be unclear in which dimensions it should to be taken. For example, we might be trying to analyse the transitions of a very thin film of some substance, just a few atoms thick. In this case, it might not be obvious whether we are to take the infinite limit of a two-dimensional or three-dimensional system. But dimensionality has an extremely pronounced effect on phase transitions. Typically, choosing between two and three dimensions is amounts to a choice as to the type of phase transition, or even whether it takes place at all. Thus, we revisit the "subjective" worry of §4.4.1.

Consider the parameter space once more. Both two and three-dimensional infinite systems will have their own fixed points. As the correlation length approaches a scale similar to the thickness of the film, the flow can "cross over" from three to

two-dimensional behaviour. Let us say for the sake of argument that both fixed points represent phase transitions, but of different varieties. What we would observe in this case is a substance behaving at first as though were undergoing a three-dimensional phase transition, but then as it is brought closer to the transition point, switching to behaviour associated with a two-dimensional system. To the question of whether the substance is *really* undergoing a phase transition of the two or three dimensional variety, the most natural answer would be that it showed behaviour associated with both.

The crossover phenomenon goes some way to reassuring us that the limited arbitrariness of how the infinite limit should be taken is not a problem — to speak figuratively: even the substance itself seems unsure and crosses over between the two options! But the main point is as follows: the composite behaviour of the finite system can be understood theoretically by looking at the two basic types of transition in the infinite limit, and then look at how they interfere with one another in crossover. Dimensional crossover thus provides further illustration of the indispensibility of the TD limit, both from the need to *seclude* the phenomena associated with phase transitions and also by providing the additional *structure* to separate them cleanly into classes associated with the fixed points (both of the italicised terms being taken in the sense of §4.2.3).

## 4.5 Putting Unitary Inequivalence to Work?

There is a striking feature of the interpretation of quantum field theories: the lack of a unique Hilbert space representation of their canonical commutation relations. This situation prompts various interpretative questions, among them, we must decide which of these representations should be given direct physical significance, whether any are privileged above others. It also suggests that there is no unique

quantisation of a given classical field theory.

In a 2003 paper, Laura Ruetsche takes an original approach to such problems. She starts from the observation that the lack of uniqueness arises from the infinite degrees of freedom of the field, and observes that this feature is also present in quantum statistical mechanics (QSM) when the thermodynamic limit is taken. But in the QSM context, there is a better established understanding of the representational work done by inequivalent Hilbert space representations, and she proposes that we should be guided to analogous conclusions in the interpretation of quantum field theories.

In this section, I want to challenge Ruetsche's approach: not because I utterly reject the interpretation she arrives at, but because the reasoning that takes her there is directly undermined by the position taken in the rest of this chapter.

### 4.5.1 Algebras and their Representations

The standard way of looking at quantum theories of finite systems is as a *Hilbert space theory*. Such a theory is characterised by the following features:[23]

- The *observables* are associated with the set $\mathfrak{B}(\mathcal{H})$ of all bounded self-adjoint operators acting on a Hilbert space $\mathcal{H}$.
- The *states* of the system are represented by the set $\rho(\mathcal{H})$ of all positive normalised trace-class operators on a separable Hilbert space $\mathcal{H}$.
- For a state $\hat{\rho}$ and an observable $\hat{A}$, the theory associates an *expectation value* $Tr(\rho A)$.
- The *dynamics* of the system can be represented by the unitary operator $\hat{U}_t = e^{\frac{-i\hat{H}t}{\hbar}}$, where $\hat{H}$ is the energy observable. In the Heisenberg picture,

[23] In what follows, I take on the notation and approach of Ruetsche (2003), with additional material and discussion taken from Sewell (1986) and Emch (2006).

this operator acts on the observables, giving the evolution of a state $\hat{\rho}$ over time $t$ as a transformation $\hat{A} \rightarrow \hat{A}_t$, where $\hat{A}_t = \hat{U}_t \hat{A} \hat{U}_t^*$.

The standard way of obtaining such a structure is to quantise a classical theory. In fact, there is a rather successful algorithm which we can use to convert a classical theory cast in Hamiltonian form into a Hilbert space theory. This *quantisation algorithm* takes the canonical position and momentum variables $(q_i, p_i)$ and converts them to symmetric operators $(\hat{q}_i, \hat{p}_i)$ acting on a separable Hilbert space $\mathcal{H}$, obeying canonical commutation relations (CCR's) which are related to the classical Poisson bracket. We call a Hilbert space theory that satisfies the scheme, a *representation* of the CCR's, and write one as $(\mathcal{H}, \{\hat{O}_i\})$.

Although it is still orthodox to think of quantum theories as Hilbert space theories, there is an alternative view which arose from the investigation of the realisations of CCRs in quantum field theory. It can be shown that each Hilbert space representation of CCR's gives rise to an abstract $C^*$ algebra, called its *Weyl Algebra*, and that this algebraic structure was independent of the particular Hilbert space representation chosen.[24] The suggestion is that this Weyl algebra could be considered as a theoretical framework in its own right, directly representing the states and observables of the quantum system. And because this alternative representation is unique, it might even provide a 'cleaner' interpretation of a quantum theory, than the Hilbert space theory obtained by the orthodox quantisation algorithm. This *algebraic approach* identifies bounded[25] quantum observables directly with self-adjoint elements of a $C^*$ algebra $\mathcal{A}$, and physical states with functionals

[24] In a little more detail: a $C^*$ algebra is an algebra $\mathcal{A}$ over the field $\mathbb{C}$ of complex numbers, with an involution and a norm. The involution $^*$ satisfies: $(A^*)^* = A, (A+B)^* = A^* + B^*, (\lambda A)^* = \lambda^* A^*$ and $(AB)^* = B^* A^*$ for all $A, B \in \mathcal{A}$ and all $\lambda \in \mathbb{C}$ and where $\lambda^*$ denotes the complex conjugate. Calling an element self-adjoint means that $A^* = A$. The norm, satisfies $||A^* A|| = ||A||^2$ and $||AB|| \leq ||A||||B||$. Clifton and Halvorson (2001, §) provides a philosophically oriented exposition and discussion.

[25] For an explanation and justification of the restriction to *bounded* observables, i.e., those associated with the bounded self-adjoint operators in $\mathcal{H}$, see Sewell (1986, 13-15).

$\omega$ on $\mathcal{A}$, specifically the linear functionals $\omega : \mathcal{A} \to \mathbb{C}$, which are normed and positive for all $A \in \mathcal{A}$.

The links between the two approaches are strong, for we can *represent* an algebra $\mathcal{A}$ in a particular Hilbert space, by setting up a structure preserving map $\pi : \mathcal{A} \to \mathfrak{B}(\mathcal{H})$. A state $\rho$ in a Hilbert space $\mathcal{H}$ then naturally gives rise to the algebraic state $\omega(A) = Tr(\rho\pi(A))$ for all $A \in \mathcal{A}$. And the GNS construction (named after Gel'fand, Naimark and Segal) allows us to move in the opposite direction. For any algebraic state $\omega$ in $\mathcal{A}$, there is a representation and a Hilbert Space, $(\pi_\omega, \mathcal{H}_\omega)$ of $\mathcal{A}$, together with a cyclic[26] vector $|\Psi_\omega\rangle \in \mathcal{H}_\omega$, such that $\omega(A) = \langle\Psi_\omega| \pi_\omega(A) |\Psi_\omega\rangle$ for all $A \in \mathcal{A}$. The GNS construction guarantees that for every abstract algebraic state there is a Hilbert space representation, and also guarantees that it is unique (up to unitary equivalence). Although we discussed this algebra as having arisen from a previously given Hilbert space theory, we are quite free to consider it in abstraction — as a theory in its own right, without any Hilbert space underpinnings.

For finite systems, the Hilbert space and Algebraic approaches co-exist peacefully, for the Stone-von Neumann theorem assures us that any two irreducible[27] representations of CCRs associated with a finite dimensional configuration space, will be *unitarily equivalent*. This means that for any two Hilbert space representations $(\mathcal{H}, \{\hat{O}_i\})$, and $(\mathcal{H}', \{\hat{O}'_i\})$, there exists a unitary map $U : \mathcal{H} \to \mathcal{H}'$ such that $U^{-1}O'_iU = O_i$ for all $O_i$. Unitary equivalence is generally interpreted as implying the equivalence of Hilbert space theories (in the sense that the two theories will deliver the same expectation values for corresponding observables; for further discussion on the link between unitary and physical equivalence, see Clifton and Halvorson (2001, §2.2-2.3)). However, for infinite systems, the Stone-von Neu-

[26]Cyclic means that $\pi_\omega(\mathcal{A}) |\Psi_\omega\rangle$ is dense in $\mathcal{H}_\omega$. All representations can be expressed as direct sums of cyclic representations.

[27]A representation of $\mathcal{H}$ is *irreducible*, iff there are no non-trivial subspaces of $\mathcal{H}$ that remain invariant under the action of all operators in the representation.

mann theorem does not hold, and in general there exist unitarily *in*equivalent representations of the CCR's. Yet the associated Weyl algebra is well-defined, and representation independent. The question immediately occurs: are the differences between unitarily inequivalent Hilbert space representations imbued with any physical significance? Or are they merely alternative formulations of the same physical theory?

Ruetsche sets up two extreme interpretative positions. The 'Hilbert space chauvinist' believes that the Hilbert Space approach is fundamental, and as such we have to choose a privileged representation as physical: all alternative representations are judged to be superfluous. In opposition, the 'algebraic chauvinist' treats the self-adjoint elements of the Weyl algebra as fundamental, and any distinctions amongst particular Hilbert space representations are distinctions without a physical difference.

An adherent to either of these extreme positions soon meets difficulties when they try to interpret quantum field theories. The algebraic chauvinist cannot admit many unbounded operators that we would like to invest with physical significance, since they only appear in a particular representation (e.g., position, momentum and the number operator). The Hilbert space chauvinist is forced to choose a particular representation and to deny that any states not in this privileged Hilbert space possess physical significance. Each of Ruetsche's two positions are extreme, and might well be said to play the roles of straw-men (though she presents them as realistic, and associates her algebraic chavinist with quotes from Segal). But the continued debates in the interpretation of field theories demonstrate the difficulty of defining the best intermediate position — see for example: Ruetsche (2002), Clifton and Halvorson (2001), Kronz and Lupher (2005).

### 4.5.2 QSM and the Thermodynamic Limit

Ruetsche looks to quantum statistical mechanics for guidance on her interpretative position. She notes that neither chauvinist can accommodate the representational work that unitarily inequivalent representations do within QSM, at least once we demand that such a statistical theory must accommodate phase transitions. For as in the classical case, in quantum statistical mechanics we need to take the thermodynamic limit before it is possible to model phase transitions. And there the unitarily inequivalent representations *do* attain a very firm physical significance. We should look briefly at how this arises.

Let us consider a QSM system (finite or infinite) as modelled in the $C^*$ algebraic framework. The system can be represented by the pair $(\mathcal{A}, \alpha_t)$: a $C^*$ algebra $\mathcal{A}$, and a one parameter group of automorphisms $\alpha_t$, which for all states $A \in \mathcal{A}$, represents their evolution through a time $t$. Let the system have inverse temperature $\beta = \frac{1}{kT}$. Now consider an algebraic state $\omega$ which satisfies the following condition:

$$\omega\left[A\alpha_{i\beta}(B)\right] = \omega(BA) \text{ for all } A, B \in \mathcal{A} \tag{4.6}$$

We say that a state satisfying this condition is a KMS state with respect to the automorphism group $\alpha_t$ at inverse temperature $\beta$.[28] For a finite system this KMS state exists, and is unique. And when the state is given a particular Hilbert space representation, by the correspondence scheme described above, it matches that given by the usual density matrix of Hilbert space Gibbsian statistical mechanics,

[28]The states are named after Kubo, Martin and Schwinger who first pointed out the thermodynamic properties of states satisfying Equation 4.6. (Being a little more careful, we say that KMS states must satisfy the condition in a dense subalgebra of $\mathcal{A}$.) For more discussion and a precise formulation see, for example Emch (2006, §10.4), which also gives details of the rich stability properties of KMS states.

viz,

$$\hat{\rho} = \frac{e^{-\beta \hat{H}}}{Tr\left[e^{-\beta \hat{H}}\right]} \tag{4.7}$$

(where again $\beta = \frac{1}{kT}$, and the dynamics will be given by the family $\hat{U}_t = e^{-i\hat{H}t}$ of unitary operators generated by the Hamiltonian $\hat{H}$). The Gibbs state is the unique Hilbert space state associated with the thermodynamically stable state of a quantum system with the Hamiltonian $\hat{H}$, so this encourages us to identify the KMS state as the algebraic version of a thermodynamically stable state. But when we want a theory of phase transitions, the uniqueness of the equilibrium state for each temperature becomes a curse.

One way to get away from the uniqueness is to consider the TD limit, but now the picture starts to break down, for the Gibbs state of Equation 4.7 is not well-defined for infinite systems. Fortunately KMS states are more general, and we can find states of infinite systems that satisfy Equation 4.6. And it is possible to motivate the identification of these KMS states with thermodynamic equilibrium states, even without the motivation of a corresponding Gibbs state. For example, where KMS states are unique, they satisfy both local and global thermodynamic stability conditions against perturbations of state, including invariance under the dynamical group $\alpha_t$ (Sewell (2002, 113-123) and Emch (2006, §10.4)). It is also a theorem that if there is a unique $(\alpha_t, \beta)$-KMS state, then that state is a *factor state*, and these are characterised by features such as a lack of long-range correlations and fluctuations, leading us to identify them with pure thermodynamic phases. (There is more motivation for this identification; Sewell (1986, §4.4)).

But in the infinite limit, we can also find automorphism groups $\alpha_t$ and inverse temperatures $\beta$ such that there are a plurality of $(\alpha_t, \beta)$-KMS states. But each algebraic state $\omega$ in this plurality can be decomposed uniquely into extremal KMS

states (Emch and Liu, 2005, 158-160). Let us suppose that there are two such: $\omega_1$ and $\omega_2$. Besides, these extremal states can be shown to be pairwise *disjoint*, which means that no algebraic state that we can express as a density matrix on Hilbert space representation $(\mathcal{H}_{\omega_1}, \pi_{\omega_1})$ can be expressed as a density matrix on $(\mathcal{H}_{\omega_2}, \pi_{\omega_2})$. In other words, the unique decomposition gives us extremal states that are not representable in the same Hilbert Space, for they live in unitarily inequivalent representations. And these extremal states are again factor states, which we have seen are associated with pure thermodynamic phases.

We are led to the following picture: QSM represents states of thermodynamic phase by KMS states. In finite systems, these are always unique and correspond to the Gibbs state, but in infinite systems they are not unique. But they can be decomposed into extremal, factor states, which can only be represented in unitarily inequivalent Hilbert spaces. One final piece of information is needed to complete the picture: the algebraic framework represents temperature as a "global" super-selected observable, which is a constant number for each representation.

Ruetsche points out that both chauvinists are in trouble. For the difference between inequivalent representations is the only aspect that distinguishes the different states. The algebraic chauvinist is committed to consider such a distinction unphysical, since it is not represented in their semi-local algebra. On the other hand, the Hilbert space chauvinist is meant to select a single representation, and consider all others to be unphysical. And this would commit them to holding that only one equilibrium state was physically possible with only one temperature, and so lose their theory of phase transitions.

### 4.5.3 Lessons Applied

Considering these and similar problems Ruetsche is led to advocate "grades of physical possibility". One thinks of the algebraic states as picking out the broad-

est sort of possibility: the totality of states available to a given system. Then, by choosing a particular Hilbert space representation we narrow down possibility, fixing physical contingencies such as temperature and representing only those states consistent with them. Ruetsche (2003, 1339-41) admits that her approach is only a sketch of a position, and grants that the idea of 'treating physical possibility as a matter of degree' is one that needs to be developed in more detail.

Since Ruetsche's position is at present only an outline, it is hard to estimate its promise. But I am concerned by her belief that if the position can be filled out it can be hoped to provide a general understanding of quantum theories with infinite degrees of freedom (ibid, 1341). For it appears to me that she has drawn general lessons about representational significance from just those features which are bound up with peculiarities specific to the TD limit.

This can be seen if we repeat the question of why we take the TD limit, but this time for quantum rather than classical statistical mechanics. Again, the statistical theory is intended to model a finite system, but we appeal to the TD limit nonetheless. Of the three motivations separated in §4.2.3, let us ignore the consideration of mathematical convenience, for the algebraic approach is fairly difficult to use for everyday calculations in any case. And at the level of generality we are considering, the motivation of seclusion is less important, for we could run the same discussion with a very "clean" example of a single type of phase transition, which needs no further isolation.

We are left with our motivation for taking the TD limit, as providing the theory with enough *structure* to provide a satisfactory theory of phase transitions. And this can be identified with the fact that it is only in the TD limit that the KMS state can become non-unique and non-trivially decomposable into extremal states — exactly the structure needed to provide a theory of phase transitions.

The clearest way of appropriating this structure requires us to recognise phys-

ical significance in the unitarily inequivalent representations of states. But then Ruetsche's point that neither of her 'chauvinists' are able interpret the new structure in a straightforward manner, is hardly surprising. As we have just seen, it is also difficult to interpret the structure provided by the TD limit in the classical case. But this suggests that both the Hilbert Space chauvinist and the Algebriac chauvinist can follow the lead of the classical discussion, and make use of the newly available mathematical structure, without compromising their position. Let us briefly sketch how this can be done.

First, the Hilbert Space chauvinist can consider the KMS/Gibbs equilibrium state to provide the most direct representation of any finite system. But to analyse phase transitions, he must consider the TD limit, in particular he must ask whether the KMS state in that limit is unique and extremal. He can hold that it is facts about the original finite KMS/Gibbs state that determines how the infinite KMS state should be constructed; and thus it is the finite system state that determines the nature of any infinite system states, which describe phase transitions. The fact that he cannot consider more than one representation of this infinite KMS state as directly representational becomes irrelevant — he does not need to.

The Algebraic chauvinist is equally free to consider the finite KMS/Gibbs state as the directly representational one (though he will feel that it is the KMS algebraic state that is primary, as opposed to its particular Hilbert space representation as a Gibbs state). And when he looks to the infinite KMS state in the TD limit, he is not limited to the semi-local algebra when he looks for a temperature observable. The GNS construction gives him the structure he needs to distinguish temperatures, and he can hold that this construction is encoded in the original (finite) algebraic state, even if it did require a diversion through the infinite limit, and through the structure of inequivalent Hilbert space representations to use it.

We can see that both Ruetsche's original problem (the inadequacies of the

resources available to the chauvinists), and its solution (as outlined above) arise directly due to the pecularities of the representational role of the TD limit, in exact parallel with the classical cases. So Ruetsche's suggestion that we should take interpretative lessons across to a situation in which the infinite nature of the system is to be interpreted realistically — such as a field theory — appears at least questionable.

At the end of her article, Ruetsche briefly considers the objection that she may 'have rested interpretative conclusions on the consideration of a setting which is a hotbed of manifest falsehoods and extreme idealisations' (Ruetsche, 2003, 1342). Her response is simply to deny the charge, holding that her interpretative conclusions are resting on 'those facets of the thermodynamic limit that appear to do representational work'; a contention she supports by pointing out their indispensibility to the structure of the theory. My response should be clear: while the TD limit is indeed indispensible, this does not mean it plays a direct representational role, and certainly not that we should draw on it to inform wide-ranging interpretative doctrine.

For vividness, we can push the parallel with the classical case a little further. Imagine that we looked at a classical field theory such as electromagnetism, and wondered how to interpret any non-analyticities in the fields. The analogue of Ruetsche's argument would be to look to the TD limit of classical statistical mechanics, note that it made a central and ineliminable use of non-analyticities to represent an element of physical reality, and argue that electromagnetism must admit them as well.[29] Using this classical statistical mechanics to guide our interpretation of a classical field theory would — I suggest — be as misguided as using quantum statistical mechanics to guide our interpretation of quantum field theory.

[29]It might be objected that the representational role of the free energy function in statistical mechanics is not sufficiently analogous to that of the electromagnetic field. A closer parallel would be to the order parameter field, which we introduced in §2.4.2. This is defined over space-time points, and will have singularities at many varieties of phase transition.

**Effective Field Theories**

All this having been said, there is an oft-discussed interpretative position, within which QSM *would* be relevant to the interpretation of unitary inequivalence. One contemporary approach to understanding many of the pecularities of QFTs, such as the existence of infra-red and ultra-violet divergences, and their renormalisability, is that they should be understood as "effective" field theories. That is, they are to be understood as valid only for a particular energy scale, absorbing the effects of small-scale/high-energy interactions into the definition of the field parameters. In particular, there have been suggestions that the understanding of renormalisation techniques as applied to phase transitions in condensed matter physics, show that renormalisability is a "signature" of an effective field theory. So the renormalisability of present field theories should make us hesitate to interpret them as directly representing any fundamental furniture of the world.

I discuss this view of quantum field theories in the next chapter (§5.5.3), but it bears on Ruetsche's thesis. For if we accept the "effective field" view, with an added hypothesis that the unknown higher-level theory is one with a finite dimensional state space, then lessons from the interpretation of QSM start to look a great deal more relevant. Objects such as the order-parameter field would be seen as a variety of effective field, in that they summarise the effects of many lattice interactions into a continuous field. And this treatment also breaks down at certain energy scales (i.e., those that probe distances comparable to the lattice spacing). In this way we might consider the infinite nature of QFTs as having arisen in a way analogous to the TD limit. But there are at least two extremely controversial presuppositions that would have to be made before adopting this position. First, we would have to hold that all present quantum field theories are effective; second, that the higher energy theory, of which they are merely a low-energy limit, is a discrete theory with finite degrees of freedom. If Ruetsche is committed to both of

these theses she does not give any hint of it. In contrast, she writes throughout as though she wants to interpret quantum field theories as straightforward a manner as possible, directly describing fundamental features of the world.

## 4.6 Concluding Remarks

We have arrived at the following picture. Only in the infinite limit can we find the mathematical apparatus used to describe and categorise phase transitions. The density of zeroes in the complex plane (Lee-Yang), the orders of discontinuities in the free energy (Ehrenfest and modern variants), and the nature of the fixed points (renormalisation), all give a fruitful analysis of different types of phase transitions. On the other hand, the phase transitions that are actually observed and against which these theories are tested take place in systems that appear to be finite.

The natural position would be one that endorses the indispensability of the TD limit for our successful theories of phase transitions, but without denying the evident fact that they occur in the finite systems we see around us. The definition in §4.4.1 is a proposal that allows us to occupy exactly such a position, allowing us to define the finite instances as fully genuine phase transitions (albeit sometimes "messy" or "hybrid" examples).

Naturally, the proposal offered is not the only way in which we can occupy such a position. But I feel that it must be occupied in some way, for the alternatives are untenable. To argue that it is only for the sake of simplicity in our mathematics that we cling to the infinite limit is to fail to recognise that it makes available a taxonomy and structure which are essential to an adequate theory of phase transitions. And if we admit its indispensibility, but try to deny that finite systems exhibit "true" phase transitions, then we face a dilemma. Either we hold that all empirical phenomena are in fact aspects of infinite systems, or we hold that

the phenomena represented by the theory are not those empirical phenomena we observe. If we take the first option then we are committed to the view that we have evidence of the infinite nature of physical systems every time we see a kettle boil. But taking the second option is to hold that our most successful theories cannot be applied to the phase transitions we test them against. Unless we can provide a coherent story of how our infinite theories can acquire empirical significance nonetheless, their success becomes a mystery.

# Chapter 5

# The New Emergentist Theses

The previous chapters addressed some of the New Emergentists' central themes and tried to make them precise as philosophical theses. Briefly, these were: the existence of emergent properties (Chapters 1 & 2), the phenomenon of universality (Chapter 3) and the necessity of the infinite limit (Chapter 4). In these concluding sections I shall look at how the New Emergentists are placed in regard to various issues where emergence is claimed to have relevance.

I first consider how the New Emergentist position bears on the 'emergentist' sciences of Complexity Theory, which is often spoken of as a development of Anderson's ideas (§5.1). Next, I look at a theme not prominent in *More is Different*, but much emphasised Anderson's later work, and by other New Emergentists: the phenomenon of "protected" properties (§5.2). This takes us to a debate between Anderson and Steven Weinberg as to the foundational place of high-energy and particle physics (§5.3). In §5.4, I consider how the New Emergentist position helps illuminate the status of "emergentist" positions in the philosophy of mind, of quantum mechanics and the autonomy of the special sciences. I close with some comments on the significance of renormalisable theories in the context of condensed matter and the parallels that have been seen in quantum field theory.

## 5.1 Emergent Properties and Complexity Theory

Since the appearance of *More is Different* in 1972, many active fields of research have appeared, collected under names such as 'Complexity', and 'Self-Organisation'. Despite many differences of approach, they each claim to be specifically devoted to the study of emergent properties and to be the natural descendents of Anderson's ideas.[1] But these fields study a variety of phenomena, not all of which seem to be directly related to the specific conception of emergent properties defended in Chapters 1 & 2. In this short section I suggest that we should be cautious in relating all of these new ideas directly to the New Emergentist theses.

De Wolf and Holvoet (2005) present an impressively wide survey of 'emergentist' literature, and offer several insights on the relations between the fields. They argue that the two notions of 'Emergence' and 'Self Organisation' can be clearly separated from one another (though they catalogue many instances where they have not), whereas the term 'Complexity' is often used as umbrella term. They offer 'working definitions' of each:

> A system exhibits emergence when there are coherent [properties][2] at the macro-level that dynamically arise from the interactions between the parts at the micro-level. Such [properties] are novel w.r.t the individual parts of the system.
>
> Self-organisation is a dynamical and adaptive process where systems acquire and maintain structure themselves, without external control. (De Wolf and Holvoet, 2005, 3 & 7)

[1] Laughlin (2005), Waldrop (1992, 79-81) and Solomon (2005) all identify Anderson as a complexity pioneer, and *More is Different* as a founding document. Solomon even dates the birth of complexity science to its publication. Anderson (1989) is both encouraging and cautionary when he gives his own recommendations for extensions of notions in condensed matter physics to new areas.

[2] De Wolf and Holvoet use their own term 'emergents' here, which they define as 'a general term to denote the result of the process of emergence: properties, behaviour, structure, patterns etc.'

Although these definitions are not precise, the first is consistent with the position we have explored in earlier chapters, but the second is clearly a separate notion. De Wolf and Holvoet give cases both of emergence without self-organisation (e.g., stationary states of thermodynamics) and self-organisation without emergence (e.g., certain classes of flocking behaviour where an overall plan is possessed by each agent). But they concede that many interesting phenomena described as 'complex' contain elements of both emergence and self-organisation.

Some rival approaches to complexity can be distinguished by an implicit choice to promote either emergence or self-organisation as a cause, the other left as an effect. For example, Kauffman (1996) treats emergent properties as the result of self-organising processes, whereas Prigogine (1997) appears to reverse this. There are also those who see a common cause behind both, often the non-linear character in the underlying microphysical laws (e.g., Scott, 2003).

Yet looking again at De Wolf and Holvoet's definitions, it is clear that self-organisation is not the theme of Anderson's 'More is Different', nor a natural development of it. To link emergence and self-organisation, we would have to produce some argument that all such mechanisms for emergence are non-centralised and maintain structure in an adaptive manner. True, symmetry-breaking itself is associated with phase transitions, where we might argue that some aspects of the definition of self-organisation are fulfilled, but the link must be shown in all cases. But symmetry-breaking was only a single example of a mechanism for emergence, and we should leave open the possibility that alternatives need not have any relation to self-organisation.

Like De Wolf and Holvoet, I think it important not to conflate emergence with self-organisation. They criticise several authors for doing so (ibid, 3-6), but it is particularly confusing when one is demonstrated, and the other assumed to follow. For example, consider the claims of Roberts (2005) that economics is an

'emergent science'. He notes that no one person controls markets, house prices or commuting times and moves directly to his conclusion with the claim that 'economics is the study of such emergent phenomena'. This is a popular move: to show self-organisation, by demonstrating how a stable structure can arise without any external control, and then move to the conclusion that the structure must therefore have emergent properties. This is typically done without making it clear exactly which of the properties in question are held to be emergent, or what variety of 'novelty' (if any) is being claimed.

## 5.2 Protectorates

In contrast to our focus on emergent properties, and on the infinite limit, we have neglected a strong theme in the New Emergentist writings. Not only do emergent properties look importantly different to the microphysical ones that underly them, some also display an interesting robustness, or insensitivity to these microphysical details.

This phenomenon, often called universality, was the theme of Chapter 3, where we looked at how it might be explained by the renormalisation group. But the New Emergentists are not so interested in its explanation as in the implications of the existence of such states and properties themselves. Since universal properties are "protected" against many changes to their microphysical structure, they call the states 'protectorates', and the associated universal properties 'protected properties'.[3]

> "[P]rotected" properties of matter — generic behaviour that is reliably the same from one system to the next, regardless of details. There

[3]These terms were coined by physicists Robert Laughlin and David Pines (2000) and embraced by Philip Anderson (2000). An extended popular discussion of protectorates is given by Laughlin (2005), who explicitly identifies them with fixed points of a renormalisation transformation, although the concept is clearly meant to also extend beyond areas in which the RG techniques are usually applied.

> are more sophisticated ways of articulating the idea, such as the stable fixed point of the renormalisation group, but these all boil down to descriptions of behaviour that emerges spontaneously and is stable against small perturbations of the underlying equations of motion.[4] (Laughlin et al., 2000, 32)

The New Emergentists claim the protected properties are 'directly relevant to the broad question of what what is knowable in the deepest sense of the term' and 'fundamentally at odds with the reductionist beliefs central to much of physics'.

Drawing on the discussion of Chapter 3, we should be wary of such wide-ranging claims. And as in the discussion of emergent properties, we should be careful to separate methodological and epistemological claims from metaphysical doctrine. Again, much of the New Emergentist discussion of protectorates is taken up with suggestions for their implications for the methodology of physics. These vary from uncontroversial statements: e.g., that many working in higher-level sciences need not worry themselves about the cutting edge of particle physics (Laughlin and Pines, 2000, 30); through specific suggestions of the barriers to progress in understanding high temperature superconductivity (Anderson, 2000); to highly speculative suggestions about principles governing molecular organisation on the mesoscopic scale (Laughlin et al., 2000). I shall again leave these methodological suggestions to those to whom they are addressed — working scientists who can assess their fruitfulness in theory and experiment.

But one of these suggestions, I think we should dismiss. Laughlin and Pines assert that when we have theories which give us reliably accurate knowledge of large-scale properties to a high level of precision, they *must* be theories of protected properties. The precise details of their argument is unclear, but the following is their fullest exposition, where they refer to an underlying microscopic theory as a

[4]Note that this definition brings in requirements worryingly reminiscent of self-organisation. I shall ignore these overtones, since it does not appear to affect their arguments, either here, or against Weinberg.

'Theory of Everything':[5]

> Simple electrical measurements performed on superconducting rings determine to high accuracy the quantity of the quantum of magentic fluc $hc/2e$ ... Four-point conductance measurements on semiconductors in the quantum Hall regime accurately determined the quantity $e^2/2$ ... These things are clearly true, yet they cannot be deduced by direct calculation from the Theory of Everything, for exact results cannot be predicted by approximate calculations ... Experiments of this kind work because there are higher organizing principles in nature that make them work. The Josephson quantum is exact because of the principle of continuous symmetry breaking. The quantum Hall effect is exact because of localization. Neither of these things can be deduced from microscopics, and both are transcendent, in that they would continue to be true and to lead to exact results even if the Theory of Everything were changed. Thus the existence of these effects is profoundly important, for it shows us that for at least some fundamental things in nature the Theory of Everything is irrelevant. (Laughlin and Pines, 2000, 28-9)

If I understand this argument correctly, viz. as moving from the accuracy with which a large-scale systemic quantity may be predicted, to asserting that it cannot depend sensitively on the details of the microphysical laws, then it must be rejected. For it is plausible that there are many quantities that could be predicted accurately for a large system, but would depend very sensitively on the details of a microphysical theory. A trivial example would be a total mass-energy of a large system, which would depend very sensitively on the rest-masses of the individual particles, but this places no barrier on it being measured or predicted to a high degree of accuracy.

Other New Emergentist suggestions about protectorates share a frustrating feature with statements about emergence: they are often equivocal between epistemological and metaphysical claims. When they introduce the idea that it is only

[5]Similar pronouncements are found in (Laughlin et al., 2000, 32) and (Laughlin, 2005). The term 'Theory of Everything' is an ironic one, as they warn us: 'the Theory of Everything is not even remotely a theory of every thing.' (Laughlin and Pines, 2000, 28)

at a small length scale or high energy that an effective theory breaks down, they warn: 'The nature of the underlying theory is unknowable until one raises the energy scale sufficiently to escape protection' (Laughlin and Pines, 2000, 29). But these epistemological cautions are mixed with explicitly metaphysical claims such as this: 'these [protected] properties are actually caused by collective organizing principles that formally grow out of the microscopic rules but are in a real sense independent of them' (Laughlin et al., 2000, 32). So, I shall keep with my earlier assumption that their ostensibly metaphysical claims are to be taken seriously. In fact, in this context the New Emergentists have good motivation for making serious metaphysical claims. For as we shall see in §5.3 they make these causal statements with a specific target in mind.

## 5.3 The Anderson/Weinberg Debate

Anderson's original Emergentist arguments found their most outspoken opponent in Steven Weinberg, who claimed that there was a clear sense in which the properties he studied as a particle physicist had a status above those studied by a condensed matter physicist. His famous statement of the privileged position of the microscopic, goes as follows:

> In all branches of science we try to discover generalizations about nature, and having discovered them we always ask why they are true. I don't mean why we believe that they are true, but why they *are* true ... Why is nature that way? When we answer this question the answer is always found partly in contingencies, that is partly in just the nature of the problem we pose, but partly in other generalizations. And so there is a sense of direction in science, that some generalizations are "explained" by others. ... There are arrows of scientific explanation, which thread through the space of all scientific generalizations. Having discovered many of these arrows, we can now look at the pattern that has emerged, and we notice a remarkable thing: perhaps the greatest scientific discovery of all. These arrows seem to converge on a common

> source! Start anywhere in science and, like an unpleasant child, keep asking "Why?" You will eventually get down to the level of the very small. (Weinberg, 1987, 434-5)

Weinberg gives few further details of his strongly "objective" conception of explanation; and most criticisms concentrate on that deficiency of his account. And this led to the debate taking a rather unsatisfactory path, with critics accusing him basing his account on a notion of 'objective explanation' which simply cannot be filled out in any detail (e.g., Mayr, 1988), and Weinberg expressing disatisfaction with others' interpretations of the concept (Weinberg, 1988).[6]

We can now see one reason for the New Emergentists to emphasise an objective metaphysical sense in which protected properties are independent of details of the very small. For if they can justify these claims of independence, it looks as though they would be able to cut through all these controversies. That is they could attack Weinberg's claim that his arrows of explanation would point towards the small — whatever his detailed picture of objective explanation.

### 5.3.1 Protected Properties or Protected Quantities?

There is certainly a case to be made here, but first, some of the sweeping claims of the New Emergentists need to be qualified. To see why, we should return to a further ambiguity[7] in the sense of 'microphysical details' that we met in

[6] Weinberg's frustration in being misunderstood is perhaps most strikingly evinced when he even sinks so far as to ask for philosophers for help in making the concept precise (Weinberg, 1987). The response has not been entirely constructive; Cat (1998, 294) merrily confesses of being reminded of Carnap when Weinberg outlines his appeal to a 'Logical Structure of the Universe'.

[7] There is also a second ambiguity in the repeated appeal to 'microphysical details'. Such 'details' could mean either the precise state of the system in question (i.e. where every last atom and molecule is located and its energy state), or it could mean the precise form of the microphysical laws, as encoded in the interaction potentials between these atoms and molecules. But since both only enter the question through the Hamiltonian of the entire system, we can be confident that the fact of universality is demonstrating that either kind of 'details' may be changed, at least to some degree.

Chapters 1 & 2, where we distinguished between attributes as determinables, and as determinate values of determinables, calling the former *properties*, and the latter *quantities*.

Unfortunately, the New Emergentists do not distinguish between protected properties and quantities. Occasionally, they speak of determinate quantities as protected, such as the Josephson quantum (e.g., Laughlin and Pines, 2000, quoted on p.266); but more often they are concerned with determinable properties, listing for example, 'superconductivity, band insulation, ferromagnetism, anti-ferromagnetism' (Laughlin and Pines, 2000, 29). But the plausibility of the cases are very different, as are their implications for Weinberg's argument. Taking just one example from Laughlin and Pines' list, viz. ferromagnetism: we saw in Chapter 3 that it is only a very few quantities in the thermodynamics of magnets that are universal (i.e., the critical exponents). Others, such as the temperature of the phase transitions are strikingly *non*-universal, and vary by material.

The question arises as to whether there is a coherent sense in which some determinable properties can also said to be protected, even though they are not universal. I think the answer is Yes, for as we saw in §3.5, so long as our microphysical Hamiltonian falls under the influence of a particular fixed point, we can write an effective Hamiltonian $\mathcal{H}$ with a particular form (the LGW Hamiltonian in that case). Systems with the same form of $\mathcal{H}$, but different values for the coupling constants may have many properties in common, but not many identical quantities. Two LGW Hamiltonians for example, may well share the property of ferromagnetism by virtue of their common form, but would not have identical quantities of magnetisation.

But even if we have a coherent sense for both protected properties and protected quantities, it is only the latter that provide a solid base for an attack on Weinberg's argument. For if we attacked Weinberg's argument relying on the

widespread "determinable-only" protected properties, Weinberg could easily reply that he wanted an 'objective explanation' for not only the properties associated with ferromagnetism, but their exact values. Only if we restrict the argument to the rarer protected quantities can this move be blocked.

### 5.3.2 A Limited Counterexample

Thus, to restrict Weinberg's room for reply, we will now focus on protected quantities: with the critical exponents serving as our examples. And now it seems that we have all we require to ambush Weinberg. Protected quantities depend only on the dimensionality of the lattice, and the symmetry broken in the transition, as summarised in the order parameter. Both of these are clearly large-scale properties of the whole system (in the terms of Chapters 1 & 2, they may be amongst the microphysical$_1$ but they are certainly not microphysical$_2$). And so, in protected quantities such as critical exponents, we have an apparent case where the arrow of explanation cannot easily point to the very small; and this is one that does not depend critically on how Weinberg fills out his concept of "objective explanation". For we can have counterfactual circumstances with the same protected quantity in question in which all the microphysical details are changed (at least to some degree). To dictate that his 'arrow' must point to the microphysical nonetheless, seems to me to be a notion of "objective explanation" unworthy of the name.

But if we consider all the qualifications we have made to get to this point, the victory over Weinberg does not look so terribly impressive. Truly protected quantities are rare, and only for them can we make a fairly watertight case that Weinberg's arrows do not point in the directions that he needs them to. And even there, perhaps once our 'unpleasant child' asks us why the nature of the order parameter or the dimensionality of the lattice, are the way they are, Weinberg will insist that our answer will be in terms of the very small. And since he explicitly

dictates a principle of 'transistivity of explanation' (in Weinberg, 1988), his arrows might end up pointing to the small after all.

As well as this problem, it is not clear to me that Weinberg really could demand explanations for every quantity as well as the properties. For if we are to trace the value of every quantity in science, there are plenty of basic constants that have no obvious connection with the very small (Einstein's cosmological constant, for example). It is hard to decide these questions without further details of Weinberg's 'objective explanation' — perhaps we have not fully succeeded in side-stepping this issue after all.

## 5.4 The New Emergentists and 'Emergentist' Debates

In Chapter 1, I suggested that approaches to emergence often suffered from the undue influence of interest in other philosophical problems. A philosopher of mind, concerned with problems of the causal efficacy of the mental, would search for a definition of emergence that bore clearly on issues of causation. Or an interpreter of quantum mechanics, would look for some property such as a break-down of unitary evolution, that might cast light on the quantum/classical division.

I suggested that the New Emergentists' approach was superior to these parochial approaches, in that they did not design a definition of emergence so that it could solve some existing problem, but formed their concept by drawing on well-understood physical processes. So let us return to these specific problems to see if the New Emergentist picture can help with any of them. We begin with quantum mechanics.

### 5.4.1 Quantum Mechanics and the Measurement Problem

There are at least two programmes within the study of the foundations of quantum mechanics which could be said to involve emergence in a significant sense. First, we could interpret the well-known decoherence programme of quantum measurement as being engaged in study of a specific mechanism for emergence.[8] Second, it has been proposed that quantum measurement itself can be seen as a species of symmetry-breaking, and that the measurement process can be explicitly modelled using the framework for such processes that has been developed within quantum statistical mechanics.

Let us start with the first proposal. The decoherence programme aims to show how a quantum system interacting with its environment, will quickly evolve into a state in which the non-diagonal elements of the density matrix (in a representation for position, or a position-like quantity) quickly become negligible. In itself, decoherence is not a solution to the measurement problem, but it is an important part of approaches such as Everett-like proposals, where we need an explanation of why only branches with classical-like macroscopic quantities are observed.

It is tempting to look at this process as a possible mechanism for emergence in Anderson's sense. Decoherence is a physical process by which properties of large collections of particles can be shown to behave very differently to small collections. So decoherence might be viewed on a par with symmetry-breaking. But the parallel is not a good one, at least not in the best-known approach: environmentally induced decoherence, where we look at the effect of interactions of a system with some external environment. Here, our contrast is between the properties of an isolated small system, and a non-isolated small system. Or to put it another way, the contrast between properties of small and large systems is

[8] Many workers in decoherence theory freely refer to their search for the process by which classical properties 'emerge' from quantum processes; e.g., Zurek (1991, 37), Wallace (2003, 87).

not made directly, we examine only the properties of a *subsystem* of the larger system. In view of this important difference, the classical-like properties that the decoherence effect can produce should be recognised as importantly different to the emergent properties we have been concerned with.

The proposal that measurement might be seen directly as a symmetry-breaking effect has been made in several forms, more commonly as suggestions frustratingly short on detail, e.g., Laughlin (2005, 47-59), Anderson (1984, 50-51). But some provide detailed models of the measuring apparatus, Emch (e.g., 2006, §10.6), Sewell (2007). These models typically set up two quantum systems: one large, which represents a measuring device; and one small, the system to be measured. Emch's approach uses quantum statistical mechanics to analyse the states of the combined system, and shows how a phase transition can be induced in the combined system, which depends on the original state of the smaller system. It can be demonstrated that the properties of the state of the 'measuring apparatus' that emerge from such an interaction have many 'classical characteristics'.

This programme can provide suggestive results, and the specific models seem to me to give much insight into measurement processes. But it is important to realise the restriction on their scope. We can extract as many classical characteristics from our symmetry-breaking models as we like, but so long as we stick with unitary dynamics, we cannot get rid of the superposition of measurement alternatives. So like the decoherence approach, it cannot solve the measurement problem on its own, but needs the help of a separate interpretation (a point which Emch and Anderson concede,[9] but Laughlin does not seem to accept).

[9]Anderson concedes that a separate interpretation is necessary in a transcribed discussion (2005, 512-513) and reluctantly chooses an Everett picture, complatining that 'there is too much philosophical baggage to carry around, but I can't see how to avoid carrying that baggage.' Yet Adler (2003, 235) points to 2001 survey, where Anderson dismisses a discussion on the quantum measurement process by its not mentioning the word 'decoherence'. Adler takes this as proof that Anderson thinks that the decoherence process solves the measurement problem. An alternative interpretation is that Anderson thinks that decoherence is a necessary part of the solution —

### 5.4.2 Emergence in the Philosophy of Mind

It is almost by definition that the New Emergentist doctrines cannot go terribly far in addressing the issues tangled up in the so-called 'mind-body problem'. For the arguments are based on differences between physical and mental properties: for example, particular causal capacities, a capacity to support consciousness, or some semantic character.

But the New Emergentists' examples all contrast the differences of one set of physical properties with another. And the novelty must be shown to come about by means of well-understood physical processes. As such it is difficult to see how the lessons taken from physical-physical contrasts can be used for anything other than a broad parallel with a physical-mental contrast. So even leaving aside my doubts on Batterman's claim that renormalisation techniques give new insight into the phenomenon of multiple realisability (expressed in Chapter 3; I have doubts about his further claim that the multiple realisation of mental properties by physical properties will be explained by the same strategy.

It would be otherwise if the anti-materialist arguments only held for the non-identity of *micro*-physical properties with the mental.[10] But I do not know of any anti-materialist arguments that admit this limited scope.

As it stands, any New Emergentist argument cannot progress much beyond pointing out the wide differences between microphysical and macrophysical properties, and asking rhetorically whether the differences between physical and mental properties could not be bridgeable, by some similar process. We have no new reply if the anti-materialist persistently holds that it cannot.

---

and thus a necessary part of any decent discussion of the topic — but not the whole story.

[10] In the terms of Chapter 1, I mean microphysical$_2$ properties: properties of systems below a certain size, which they display only in isolation, or when they are parts of systems other than the one in question.

### 5.4.3 The Philosophy of Special Sciences

The outlook is a great deal better for those who look to the New Emergentist thesis to explain the independence and variety of the special sciences. Here, the New Emergentists are able to present a coherent picture of how all of these areas of study can be said to have the same subject of study: physical properties, but still split naturally into realms that are both different from and independent of one another.

For an explanation of the variety of the different special sciences we look directly to emergent properties, and see that fully physical properties of large and small systems can be very different to one another. Mechanisms such as symmetry-breaking show us how the dynamical laws governing the properties of larger systems may look very different from smaller ones, and as such we can see how a new set of methodologies, specialist knowledge and mathematical techniques may be needed to study them.

For an explanation of the independence of the special sciences, protected properties may have some bearing. For in these circumstances, we have an explanation for how we can come up with theories for properties of large systems, even in ignorance of the details of the microphysical interactions. As argued in Chapter 3, these circumstances can sometimes be explained by the apparatus of the Renormalisation Group. But again we must caution that the RG is only one explanation amongst many alternatives. In particular, many special sciences may be engaged with functionally defined properties. To give an extreme example: systems theory explicitly abstracts from all but the input and output of functionally defined 'black-boxes', and its independence from particle physics is explained 'trivially' (in the sense of §3.1.1). And it seems plausible that a similar explanation be suitable for areas of biology, economics and psychology (though this is to disagree with Anderson's more speculative suggestions, e.g., 1972, 393).

### 5.4.4 The Necessity of Infinite Systems

There is a final theme in Anderson's paper that I feel should be interpreted as having little metaphysical significance. In *More is Different*, Anderson characterises the process of one science emerging from the subject matter of another as a shift from 'quantitative to qualitative differentiations' and claims that emergent behaviour is only 'rigorously definable' in the '$N \to \infty$ limit'.

The significance of this 'switch' from a qualitative difference in finite systems, to a quantitative one when the infinite limit was taken, was the subject of Chapter 4. Anderson sometimes appears to connect this feature with the existence of emergent properties, a link I feel is not helpful. While the use of the infinite limit is often connected with theories of emergent properties (both symmetry-breaking and the RG use it), I do not think that the connection is a necessary one, as I argued (contra Chuang Liu) in §2.5.1.

Chapter 4 offered a picture of the infinite limit in which phase transitions and other emergent phenomena could be seen as genuine, qualitative phenomena, but without investing the infinite limit with any mysterious metaphysical significance. I cannot find anywhere a statement from Anderson or his followers, that looks as though it is meant as a full-blooded metaphysical claim. So I suggest that in this case, the statements of the New Emergentists are intended as purely methodological recommendations. I also take this as yet another suggestion that in those areas where they do make explicit metaphysical pronouncements — about emergent and protected properties for example — they are fully aware of what they are doing.

## 5.5 Renomalisability, Realism and Effective Field Theories

Let us consider a system that is well-described by a renormalisable theory (such as the LGW model). Our success, and the details of the RG analysis suggests that the LGW Hamiltonian captures all the relevant operators associated with a certain fixed point, but we also know that there is a vast range of microphysical models that flow within the range of this fixed point. This is a rather striking situation: we are presented with a theory that is empirically successful over the scale of length and energy under which it has been tested, but its empirical success gives no reason to believe that it is close to an accurate representation of the microphysical structure or properties of the system in question.

This last point looks as though it might be highly significant, for the usual realist accounts of the empirical success of scientific theories is in terms of some firm link between the real system and its mathematical treatment in the theory. Of course, there is much controversy over the exact nature of this link; most often it is treated as a variety of isomorphism, or of resemblance Suarez (though for a criticism of these views, see 2002, 2003). And there are also suggestions that theoretical models form an important mediating stage (e.g., Morgan and Morrison, 1999). But however it is made out, there is meant to be a clear link between models or theories, and any real systems that they are meant to give predictions for. But in the case of renormalisable theories/models, it seems that there is no easily recognisable link between the mathematical expression of the Hamiltonian and an element of physical reality.

### 5.5.1 An Anti-Realist Attack

The considerations above suggest that renormalisable theories could provide a counter-argument to scientific realism, both on naïve and more sophisticated versions. Let us start with the classic articulation of the realist intuition: Putnam's response to the empirical success of a theory that it must have got something right about the world, or its success would be a 'miracle'. But if we are presented with a renormalisable theory, the appropriate attitude is to be to be fairly confident that the model has got very nearly *nothing* right about the world (with the usual exceptions of the dimensionality of the lattice and of the order parameter of the system). But we need not appeal to miracles — once we have recognised that our Hamiltonian is renormalisable, we have an explanation of its empirical success, which does not rely on an orthodox realist link.

We must also add that we cannot dismiss this situation as one in which just a few microphysical details are missed. There is no expectation that the "true" microphysical theory will share anything very much with the effective one. It is unlikely to be a field theory (in the condensed matter case, at least), and symmetry-breaking rears its head once more, to remind us that the lower-level theory may have symmetries utterly at odds with the LGW model.

This situation does not seem to change when we consider more sophisticated versions of realism. There are a wide variety of approaches that march under the banner of 'structural realism', but all keep the link between theory and reality. They merely weaken any link of resemblance to some version of a structural correspondence. But the minimal symmetry features that characterise a universality class cannot be enough to constitute any "structural correspondence", unless the account is to collapse to vacuity. And similar remarks could be made about accounts such as the 'Natural Ontological Attitude' (Fine, 1984) and other sophisticated realist approaches (Psillos, 1999). If the correspondence claimed is so weak

as to be satisfied by the minimal quantities that serve to define a universality class, our "realism" would not deserve the name.

However, this type of argument is not original to renormalisable theories. For again and again in the history of science, we have discovered that empirically successful theories were almost entirely wrong about the details of the system in question. Classic examples include the phlogiston theory of heat, aspects of both corpuscle and wave theories of light, and Newtonian theories of gravity. All allowed novel successful predictions to be made, despite their lack of "true" correspondence to the systems in question. Such examples lie at the basis of the so-called Pessimistic Meta-Induction (PMI) argument against scientific realism (Laudan, 1981). Realists have not lacked responses to such attacks. For example (at least one version of) structural realism was formulated in response to them (Worrall, 1989).

The difference between these examples and the renormalisable theories is the temporal order in which we are presented with a non-realistic but empirically successful theory, and the explanation for this empirical success. For example, in the case of the phlogiston theory, the kinetic theory of heat presents us with an explanation for its (limited) empirical success. It can show how molecular motion can be transferred in the manner of a "flow" under some circumstances; how there are many processes in which heat is conserved, and so on. It also shows the limits of the superseded theory, by circumscribing those circumstances under which it will be successful. But in the case of a renormalisable theory, we are presented with the explanation for its effective success before any discovery of a particular microphysical structure underlying it. And we are also presented with limits on its domain of applicability: near to the critical point, or within a certain energy range. We know that it will break down outside that domain, even if we cannot test it there, but we do not know what its break-down will reveal.

So renormalisable theories do not give us a new attack on scientific realism,

though they might be said to sharpen an existing one. We could call the counterexample "PMI plus": for not only does the record of previous theories suggest to us that the present one is likely to be inaccurate, but if it is renormalisable we are also given an explanation of its success that does not rely on its accuracy (at least within a certain regime).

### 5.5.2 A Realist Parry

Fortunately for the realist, an understanding of RG techniques does not only provide an attack on the realist position, it also gives them a way to escape. They may concede that the LGW effective Hamiltonian cannot represent the microphysical structure directly, but simply ask what it does represent. I cannot see that the answer can be anything other than the fixed point of the recursion relation in parameter space. (More precisely, the terms of the Hamiltonian represent the relevant directions of the fixed point). This allows us to recognise that a theory is effective, remain ignorant of the microphysical Hamiltonian, and yet be satisfied with an explanation of its empirical success. We can be confident that it will be successful just so long as we are confident that the theory represents the correct RG fixed point, and has captured all of the relevant operators (plus any marginal ones that may arise to complicate matters).

The representational role certainly looks an odd one, but we can retrieve a realist link by pointing out that the recursion relation is tailored to the specific physical situation we have in hand (whether this is in Kadanoff's approach, tailored to a specific microphysical model, or in Wilson's approach, where it is to a macrophysical model). So we can hold that there *is* a connection between the empirically successful theory and the physical system: an indirect one, mediated by the design of the recursion relation and the fixed point. I feel that this is an appropriate response to the realist worries surveyed above: the theory-reality link

still exists, but it is very different to the orthodox suggestions of resemblance or isomorphism.

### 5.5.3 Effective Field Theories

Thoughout this thesis, I have only alluded in passing to the remarkable analogies that exist between theories in condensed matter and high energy physics. Where I have addressed them specifically, such as the criticism of Ruetsche's thesis in §4.5, I have been largely cautious about their significance. But Wilson's work in condensed matter has been hugely influential in the interpretation and philosophy of quantum field theories; in particular, in the attitudes to the significance of renormalisation. Before the 1970s, the renormalisation procedure was either seen as a rather mysterious, but necessary 'dippy process'[11] one had to apply to quantum field theories before they gave sensible, finite empirical predictions; or it was seen as some deep requirement for theory choice (though also a mysterious one). As Weinberg (1980, 517) puts it: 'I thought that renormalizability might be the key criterion, which also in a more general context would impose a precise kind of simplicity on our theories and help us to pick out the one true physical theory out of the infinite variety of conceivable quantum field theories.'

However, due largely to Wilson's work, there has been a remarkable 'change in attitude' (Weinberg, 1997, 41) to the Standard Model, and Quantum Field Theories in general, viewing them as 'Effective Field Theories' (EFT's) for a given energy scale. That is, a quantum field theory merely expresses the low-energy limit of a different theory, which explicitly includes interactions on a smaller scale which have been absorbed into the parameters of the given field Hamiltonian.

The technical details of quantum field theories are far outside the scope of the present discussion, but a broad conceptual picture can be obtained by replacing

[11] Feynman (1988). He also refers to it as 'a shell game' (1998, 6).

the length scale with an energy scale $E$ (a high energy scale replaces a small length scale) and thinking, not of alternative microphysical Hamiltonians, but of alternative higher energy theories. For the condensed matter case, it is as we approach the lattice scale $a$, that universality starts to break down. In the quantum field case it is as we approach the energy $M$ at which a new particle is likely to be produced. As we approach this scale, non-renormalisable interactions appear (until this point, they are suppressed by powers of $E/M$). So as $E$ approaches $M$ we cannot ignore them any more, and a new physics is required. Cao and Schweber urge that we have reason to expect that the new physics is also described by a quantum field theory, for they point to a theorem of Appelquist and Carazzone (1975), which demonstrates that if a renormalisable QFT includes a high energy particle, there will be a theory of the low energy limit (also a QFT) as if that particle did not exist. They can thus guarantee a smooth transition from one QFT to another, with no problems of the rigourous existence of a low-energy limit.[12]

This view has inspired some philosophical comment, largely as a reaction to the work of Cao and Schweber[13] who view each energy level as 'a quasi-autonomous domain, each with its own fundamental laws, and fundamental ontology'. They argue that this supports 'a pluralism in theoretical ontology, an antifoundationalism in epistemology and an antireductionism in methodology' (Cao and Schweber, 1993, 69).

If we lay aside the issues of epistemology and methodology, we are still left with two separate metaphysical claims. First, there is the idea that if we treat field theories as effective, then each energy level constitutes its own quasi-autonomous

[12]For technical details of the effective theory approach, including simple examples, see Polchinski (1984, 1999) and, as a contrasting approach, Georgi (1994).

[13]Their principal works include: Cao and Schweber (1993), Cao (1997), Schweber (1993). Commentaries on their philosophical claims have come from Huggett and Weingard (1995), Castellani (2002) and Saunders (2003).

domain, within which we have fundamental laws and ontology (and so, presumably, fundamental properties). Second, there is the anti-foundational thesis, that as we increase the energy, we just carry on finding new effective field theories, and as Huggett and Weingard (1995) put it, reality is a 'never-ending tower of EFTs'. Each may be considered separately.

In regard to the first, the emphasis seems very much to be on the *quasi* of the quasi-autonomous, for we see the same determinable/determinant ambiguity as we saw in §5.3.1. That is, the *form* of the field Hamiltonian at each level may be largely independent (decoupled) from the form it may take on higher energies, but many of the actual values of — say — the coupling strengths or, particle masses in the renormalisable theory are dependent on the unknown processes at higher levels. The price we must pay for insisting that each level is 'autonomous', is to leave the values of these parameters unexplained at that level; they can only be fixed by matching to experiment. Thus, it is very much open to debate as to whether each level can be said to be "truly separate" from those at higher energies.

In regard to the second metaphysical claim, the tower of EFTs is certainly a coherent metaphysical picture, and it is certainly true that the physics of renormalisation does licence it as a possibility. At low energies we would be able to use a renormalisable quantum field theory, but as we look at high energies, it would break down in a way that suggests there are unrenormalisable terms. We might deal with this by introducing a new particle, and obtain a renormalisable theory once again. But then as we increase the energy further this would in turn start to look unrenormalisable — and so on.

But one of the points of renormalisable theories is that they conceal the physics above a certain energy level. As such, we have no reason to believe in an infinite tower of EFTs as against any other picture that is equally compatible with what we know of the low-energy regimes. While it is very nice to be able to appeal to

the theorem of Appelquist and Carrozone this theorem shows that a higher-energy renormalisable QFT would be *sufficient* for the observed results at low energy, not that one is *necessary*. The higher energy theory could be a non-renormlisable field theory, or for that matter something that looks utterly new. Indeed, if we really are to take the parallel with Wilson's success seriously, we could even argue that we should expect quantum field theories to turn out to be low-energy approximations to a lattice theory.

## 5.6 Closing Remarks

Where does all this leave the New Emergentists? I think it leaves them with a position on emergence with clear standards for the application of the term, which is fully grounded in standard physics (and physicalism), and able to point to paradigm cases of emergent properties. Most importantly, their position gives very clear guidance on what an appeal to emergence can and cannot do. An emergent property can be very different from microphysical properties, and the laws they obey can be very different from those obeyed by the smaller constituents of a system. And in the case of the protected, or universal properties, large-scale properties can be more similar to one another than to the microphysical.

But the New Emergentist position also makes it clear that emergent properties are not a panacea, to be appealed to whenever we are puzzled by the properties of large systems. In each case, we must produce a detailed physical mechanism for emergence, which rigorously explains the qualitative difference that we see with the microphysical. If such a mechanism is missing, an appeal that a suggested property is 'emergent' does nothing more than give it a comforting name. And for the New Emergentists it would be the wrong name. For unless we have shown that an emergent property really does 'emerge' from the lower level, and that microphysical

supervenience remains unthreatened, we have not shown it is emergent. For all we know, such a higher-level property might not supervene on the microphysical, but "float free" from it entirely. And this should serve as a warning to those who uncritically appeal to the notion to dissolve the mysteries of mental properties, the appearance of classical physics and biological phenomena. For calling these phenomena 'emergent' is no explanation of the nature of such properties, but a commitment to provide one.

# Appendix A

# British Emergence and Broad's Doctrines

To recapitulate from the main text, the basic claim of British Emergence is that it is impossible to predict the properties of a composite system from the properties of the components in isolation, or in other composites which are not of the same form as the one in question. Although the main text focusses on one example of how this might come about — via a configuration law — Stephan (2002) argues that Broad's writings contain two very different ways in which the prediction might be recognised as impossible. Broad himself does not distinguish the two possibilities clearly; (nor for that matter does the historical study given by McLaughlin (1992)). Yet I feel that Stephan's distinction is crucial for a proper understanding of the doctrine.

## A.1 Broad's two Senses: Emergence$_1$ and Emergence$_2$

Broad thought that some properties — in particular the secondary qualities, and those of psychology — *had* to be emergent, because they were not 'behaviourably

analysable'. He is frustratingly vague about this important term, but this quote serves to introduce the idea:

> [R]ed seems to me to be a quality of a certain motion-complex in one sense, and life to be a quality of a more elaborate complex in a very different sense. By saying that a body is living I just mean that its motions and other changes fit into each other and into the environment in certain characteristic ways. The statement is an analysis of its characteristic modes of change. But in saying that a motion is red I certainly do not mean that it is a vibration of such and such frequency. The statement is not an analysis of its characteristic mode of motion; but is the assertion that a property, which is not analysable in terms such as velocity, frequency, etc., that apply directly to motions as such, occupies the same contour as a certain set of motions. (Broad, 1921)

Clearly, there is much to be said about what "behaviour" can be included in the analysis. Stephan (2002) argues that Broad's distinction may be understood as similar to that made in the philosophy of mind between a *functionalisable* and *non-functionalisable* property. If this is so, some of his arguments share features with those that are based on the non-functionalisability of qualia, though with the advantage that 'functionalisability' is fairly clearly defined, whereas 'behaviourably analysable' is not. On this reading Broad's Archangel, which we introduce in the next section, would be a precursor to Jackson's 1982 example of Mary, the brilliant but colour-blind neuroscientist.

I prefer to understand Broad's term "behaviour" in a rather different way, one that is not so dependent on a clear notion of a causal role at a higher level. In this conception of emergence, all the properties of each microscopic component of the system are totally predictable from general laws of composition. Nevertheless, the systemic properties cannot be predicted from all of these properties, which seems to suggest a failure of supervenience (see §1.3.1). Broad's appeal to 'behaviourably analysable' properties is rather an appeal to those determined by the properties possessed by the components. Earlier Broad had identified motion as

a possibility for the 'single fundamental kind of change' possessed by the particles of an ultimate micro-level in his outline of 'The Ideal of Pure Mechanism' — the most strict denial of any emergentist hypothesis (Broad, 1925, 44). He goes on to relax this ideal, allowing different properties and changes to feature in a mechanist theory, but I believe he is referring back to this ideal in his conception of non-behaviourably analysable properties as those that do not (strongly) supervene on the most fundamental kind of change. This reading is textually supported by his reference to 'motions as such', in the quote above.

Regardless of which reading is correct, Broad obviously holds that general laws of composition, as usually understood, address only behaviourably analysable properties. Thus one way in which a property can be emergent is if it is not behaviourably analysable. They would not be contradicted, but would simply miss it altogether. Broad makes much use of this class of emergent properties, which we shall call emergent$_1$:

**Definition:.** *A systemic property is emergent$_1$ just in case it is not behaviourably analysable.*

But Broad also holds that there could be fully behaviourably analysable systemic properties that are nonetheless emergent, in his sense of being underivable from the behaviour of the system's components in other, simpler composites. In these cases, the systemic property cannot be deduced, not because they are not behaviourably analysable, but there are *specific laws of composition* the existence of which we could not have suspected from the behaviour of simpler composites. These properties we can call emergent$_2$:

**Definition:.** *A systemic property is emergent$_2$ if it cannot be deduced from the properties of the components of a system using only general laws of composition.*

Emergent$_2$ properties, together with Broad's examples of specific laws of composition are the cases of British emergence that I focus on in the main text (§1.5.2 onwards), so I shall not repeat that discussion here.

However, it is worth pointing out a further distinction amongst emergent$_2$ properties that is *not* recognised by Broad, nor by Stephan. Let us take a case in which only general laws of composition are in effect, and we can assign each system a particular state space according to those laws. Now introduce a specific law of composition, and consider whether this law leads to an expansion of the state space of the system, or whether it just takes on new values while being defined on the original state space. If the spaces are expanded, then the emergent$_2$ properties are still novel, according to Broad's criteria but they were properties that were available to the system operating under the general laws alone. However, if there are such systems whose state spaces are now larger, it seems that these emergent properties will be even "*more* novel", in that they were not even available to the system when only the general laws were in effect. It is possible that this distinction can be related to the previous division of 'behaviourably analysable' properties, but this leads to further difficulties in interpretation, since this new distinction appears *within* the class of emergent$_2$ properties, which are *all* meant to be behaviourable analysable.

Leaving this problem aside, Broad has provided us with (at least) two ways to distinguish trivial from non-trivial systemic properties, and thus provided two clean senses of 'novelty' which can be used in discussions of emergence.

## A.2 Broad's Mathematical Archangel

Broad introduced an aid to the the imagination in his "Mathematical Archangel", who is not only infinitely mathematically powerful, but 'gifted with the further

power of perceiving the microscopic structure of atoms as easily as we can perceive hay-stacks.'[1] This being 'could no more predict the behaviour of silver or of chlorine or the properties of silver-chloride without having observed samples of those substances than we can at present' (Broad, 1925, 71).

Unfortunately, Broad introduced the Archangel within a discussion that moves without comment between his two conceptions of emergence. As well as being unable to predict the chemical properties of silver chloride (which seem best understood as emergent$_2$) Broad suggested that the Archangel would be unable to predict the smell of ammonia, since such secondary qualities are not behaviourably analysable and thus emergent$_1$ (Broad, 1925, 71-80). Once again, this second argument is rather reminiscent of the modern 'knowledge argument' for the non-functionalisability of qualia, but this is a distraction from our central concerns. We can simply understand the Archangel as Broad's 'acid test' for *either* of his senses of emergence.

Let us take an example to illustrate both. If we create a large system from atoms by adding a single one at a time, the conglomerate will acquire new systemic properties at each stage. However, most of these new systemic properties — such as total mass, or total electric charge — will not be emergent in Broad's sense. They arise from the same general laws of composition that hold when we add the second, third, or millionth atom. The crucial issue for Broad is the possibility of new properties that are not covered under these laws.

> I know no reason whatever why new and theoretically unpredictable modes of behaviour should not appear at certain levels of complexity, or why they *must* be explicable in terms of elementary properties and

[1]Broad introduced his Archangel in the course of a thought experiment in which Sir Ernest Rutherford was imagined to predict the properties of large systems from the properties of their component atoms. Recognising the limits placed by experimental technique and mathematical ability, Broad replaced Rutherford by the Archangel. Replacing someone who famously held that 'all science is either physics or stamp collecting' must have been particular pleasing for an emergentist. (Incidentally, Rutherford was awarded the Nobel Prize for Chemistry).

> laws of composition which have manifested themselves in less complex wholes. (Broad, 1925, 73)

Now these could be emergent$_1$ or emergent$_2$. If an emergent$_1$ property should appear, then the system might *behave* just as if only the general laws of composition were in effect. Nevertheless, there are further, non-behaviourably analysable properties of the system that we (or the Archangel) would not be able to derive from the general laws alone. Broad makes several suggestions about 'vital' properties as emergent$_1$, and also seems to think that secondary properties are examples (Broad, 1925, 80-1).

For emergent$_2$ properties the situation is slightly different. A specific configuration law will actually make the system behave differently to how it would if only the general laws of composition were in effect. We would still not be expecting this behaviour, and neither would the Archangel, since we would have no way to suspect that the specific law of configuration existed.

## A.3 Downward Causation

There is a further feature which is particular to emergence$_2$. We have seen that Broad is happy with microphysical supervenience holding in all configurations (§1.6.2), so their components will also show different behaviour to that which one would expect considering only the general laws — a case of "downward causation".

Kim (1996) claims that British Emergentist thesis automatically implies "downward causation" — a causal influence from macroscopic to microscopic levels of nature. To interpret Kim's claim, we must make several points clear. First, for the purposes of the section, I shall assume that any causal influence relevant to the British emergentist thesis must be law-governed. We must also decide how we

should delimit the microscopic and macroscopic levels,[2] and also whether we are speaking of emergence$_1$ or emergence$_2$.

For emergence$_1$ there seems a fairly clear distinction into "levels": any description that mentions non-behaviourably analysable properties, or those laws that govern them, is called a macroscopic description; any that do not is a microscopic one. So if we look only at microlevel description we will miss the non-behaviourably analysable properties of a large system. But emergent$_1$ properties do not commit us to any interesting form of downward causation. By definition (or at least what we have taken to be the definition) the presence of non-behaviourably analysable properties does not "act-back" on the behaviour of the parts.

For emergence$_2$ the issue of levels, and thereby of downward causation, is rather less clear. Here we must revisit the ambiguities in the terms 'microphysical', and 'micro-level' which were highlighted in §1.7.2. If we take one natural reading — that a micro-level description is one that only mentions properties of microscopic entities and their arrangements — then emergent$_2$ properties provide no separation into "levels" of nature. For the specific laws of composition may be described in microphysical terms. (Example: we just presented the silver chloride example while mentioning only silver atoms and chlorine atoms and their spatial arrangements.) And once the effects of the specific laws of composition have been included, the microphysical description is complete and accurate (assuming we do not have emergent$_1$ properties as well).

To separate the micro- and macro-level using emergence$_2$, we presumably require an understanding of levels such that emergent$_2$ properties only belong to the macro-level and not to the micro. We can of course engineer this: the natural move is to call any description that does not take into account the effects of the config-

[2]In Kim's own sophisticated discussions of emergence, he uses a supervenience based conception of levels, for which downward causation is extremely problematic, apparently threatening the causal closure of the microphysical.

urational laws, a "micro-level" description; and any that do, a "macro-level" description. Thus emergent$_2$ laws would be placed on the macro-level, non-emergent laws on the micro. But on this definition a micro-level description will just be an inadequate description: it will get many properties wrong for any system that exhibits emergence$_2$, whether we describe them in terms of the properties of microscopic entities, or in terms of a macro-level. The micro-macro distinction has been defined such that the "micro-level" description simply misses out certain physical laws.

It is only this possibility of emergent$_2$ properties that implies that some sort of downward causation must take place. But if we understand micro- and macro-levels in such a way that microscopic and macroscopic levels are separated by Broad's criteria, then downward causation becomes an unproblematic consequence of the existence of configurational laws. It just becomes the claim that certain large arrangements of matter have causal influence. There are no issues relating to the causal closure of the physical, or similar issues, since the configurational law appears just as much a part of the physical world as the microphysical laws.

This deflationary conclusion can be taken through the parallels drawn with New Emergentism, where the problems are defused even further. For the New Emergentists do not claim that the behaviour of the system will be any different to that dictated by the microphysical laws. Indeed in this respect, New Emergentism shares aspects with emergence$_1$, though it is its stronger parallels with emergence$_2$ which I have concentrated on in the main text.

# Bibliography


S. L. Adler. Why decoherence has not solved the measurement problem: A response to P. W. Anderson. *Studies in History and Philosophy of Modern Physics*, 34:135–142, 2003.

S. Alexander. *Space, Time and Deity: Gifford Lectures, 1916-1918*. Macmillan, 1920.

P. W. Anderson. Are the big machines really necessary? *New Scientist and Science Journal*, 51:510–521, 1971.

P. W. Anderson. Some general thoughts about broken symmetry. In N. Boccara, editor, *Symmetries and Broken Symmetries in Condensed Matter Physics*. IDSET, Paris, 1981.

P. W. Anderson. Theoretical paradigms for the sciences of complexity. In *Anderson (2005)*. Nishina Memorial Lecture, Keio University, Japan, 1989.

P. W. Anderson. Science: A 'dappled world' or a 'seamless web'? *Studies in History and Philosophy of Modern Physics*, 32(3):487–494, 2001.

P. W. Anderson. *A Career in Theoretical Physics*. World Scientific, 2nd edition, 2005.

P. W. Anderson. More is different. *Science*, 177:393–396, 1972.

P. W. Anderson. *Basic Notions of Condensed Matter Physics*. Addison-Wesley, 1984.

P. W. Anderson. Sources of quantum protection in high-$T_c$ superconductivity. *Science*, 288:480–482, 2000.

T. Appelquist and J. Carazzone. Infrared singularities and massive fields. *Physical Review D*, 11:2856–2861, 1975.

D. M. Armstrong. *A Theory of Universals*, volume 2 of *Universals and Scientific Realism*. Cambridge University Press, 1978.

D. M. Armstrong. *A World of States of Affairs*. Cambridge University Press, 1997.

V. I. Arnold. *Mathematical methods of classical mechanics*. Springer, 2nd edition, 1989.

P. Bak, C. Tang, and K. Wiesenfeld. Self-organized criticality: An explanation of 1/f noise. *Phys. Rev. Letters*, 59:381–384, 1987.

Y. Bar-Yam. A mathematical theory of strong emergence using multiscale variety. *Complexity*, 9(6):15–24, 2004.

G. I. Barenblatt. *Similarity, Self-Similarity and Intermediate Asymptotics*. Consultants Bureau, New York, 1979.

R. W. Batterman. Why equilibrium statistical mechanics works: Universality and the renormalization group. *Philosophy of Science*, 65:183–208, 1998.

R. W. Batterman. Multiple realizability and universality. *British Journal for the Philosophy of Science*, 51:115–145, 2000.

R. W. Batterman. Response to Belot's "Whose devil? Which details?". *Philosophy of Science*, 72:154–163, 2005.

R. W. Batterman. *The Devil in the Details*. Oxford University Press, 2002a.

R. W. Batterman. Explanatory instability. *Nous*, 26(3):325–348, 1992.

R. W. Batterman. Critical phenomena and breaking drops: Infinite idealizations in physics. URL `http://philsci-archive.pitt.edu/archive/00001622/`. PhilSci Archive, 2004.

R. W. Batterman. Asymptotics and the role of minimal models. *British Journal for the Philosophy of Science*, 53:21–38, 2002b.

R. W. Batterman. Theories between theories: Asymptotic limiting intertheoretic relations. *Synthese*, 103:171–201, 1995.

W. Bechtel and J. Mundale. Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of Science*, 66:175–207, 1999.

M. A. Bedau. Downward causation and the autonomy of weak emergence. *Principia*, 6:5–50, 2002.

G. Belot. Whose devil? Which details? *Philosophy of Science*, 72:128–153, 2005.

J. Bigelow and R. Pargetter. A theory of structural universals. *Australasian Journal of Philosophy*, 67:1–11, 1989.

J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman. *The Theory of Critical Phenomena - An Introduction to the renormalization group*. Clarendon Press, 1992.

R. Black. Against quidditism. *Australasian Journal of Philosophy*, 78(1):87–104, March 2000.

N. Block. Anti-reductionism slaps back. *Philosophical Perspectives*, pages 107–133, 1997. URL `http://cogprints.org/229/00/199712002.html`.

R. A. Blythe and M. R. Evans. The Lee-Yang theory of equilibrium and nonequilibrium phase transitions. *The Brazilian Journal of Physics*, 33(3):464–475, 2003.

K. Brading and E. Castellani. *Symmetries in Physics: Philosophical Reflections*. Cambridge University Press, Cambridge, 2003.

C. D. Broad. *Mind and its Place in Nature*. Routledge & Kegan Paul, London, 1925.

C. D. Broad. Prof. Alexander's Gifford lectures. *Mind*, 30(129-150), 1921.

A. Bruce and D. Wallace. Critical point phenomena: universal physics at large length scales. In P. Davies, editor, *The New Physics*, chapter 8, pages 236–267. Cambridge University Press, Cambridge, 1989.

J. N. Butterfield. On symmetry and conserved quantities in classical mechanics. In W. Demopoulos and I. Pitowksy, editors, *Physical Theory and Its Interpretation: Essays in Honor of Jeffrey Bub*. Springer, 2006a.

J. N. Butterfield. On symplectic reduction in classical mechanics. In J. N. Butterfield and J. Earman, editors, *The North Holland Handbook of Philosophy of Physics*. Elsevier, 2006b.

J. N. Butterfield. Some aspects of modality in analytical mechanics. In *Formal Telology and Causality*. Paderborn, 2003.

J. N. Butterfield and C. J. Isham. On the emergence of time in quantum gravity. In J. N. Butterfield, editor, *The Arguments of Time*. Oxford University Press, 1999.

C. Callender. Taking thermodynamics too seriously. *Studies in History and Philosophy of Modern Physics*, 32(4):539–53, 2001.

T. Y. Cao. *Conceptual Developments of 20th Century Field Theories*. Cambridge University Press, 1997.

T. Y. Cao and S. S. Schweber. The conceptual foundations and philosophical aspects of renormalization theory. *Synthese*, 97(1):33–108, 1993.

J. Cardy. *Scaling and Renormalization in Statistical Physcis*. Cambridge University Press, Cambridge, 1996.

N. Cartwright. *Nature's Capacities and their Measurement*. Oxford University Press, 1989.

E. Castellani. Reductionism, emergence and effective field theories. *Studies in History and Philosophy of Modern Physics*, 33:251–267, 2002.

E. Castellani. Symmetry and symmetry breaking. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University, 2005. URL `http://plato.stanford.edu/entries/symmetry-breaking/`.

J. Cat. The physicists' debates on unification in physics at the end of the 20th century. *Historical Studies in the Physical and Biological Sciences*, 28:253–299, 1998.

P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics*. Cambridge University Press, 1995.

D. Chalmers. *The Conscious Mind*. Oxford University Press, 1996.

L.-Y. Chen, N. Goldenfeld, and Y. Oono. Intermediate asymptotics and renormalization group theory. *Journal of Scientific Computing*, 47:355–372, 1989.

L.-Y. Chen, N. Goldenfeld, and Y. Oono. Renormalization group theory for global asymptotic analysis. *Phys. Rev. Letters*, 73:1311–1315, 1994.

R. Clifton and H. Halvorson. Are Rindler quanta real? Inequivalent particle concepts in quantum field theory. *British Journal for the Philosophy of Science*, 52:417–470, 2001.

P. Coleman. Many body physics: Unfinished revolution. *Annales Henri Poincare*, 4:1–22, 2003.

T. Crane. The significance of emergence. In C. Gillet and B. Loewer, editors, *Physicalism and its Discontents*, pages 207–224. Cambridge University Press, 2001.

T. Crane and D. H. Mellor. There is no question of physicalism. *Mind*, 99, 1990.

D. Davidson. *Essays on Actions and Events*. Clarendon Press, Oxford, 1980.

T. De Wolf and T. Holvoet. Emergence versus self organisation: Different concepts but promising when combined. In *Engineering Self Organising Systems: Methodologies and Applications*, volume 3464, pages 1–15. Lecture Notes in Computer Science, 2005.

W. M. Dickson. *Quantum Chance and Non-Locality*. Cambridge University Press, 1998.

C. Domb. *The Critical Point: A historical Introduction to the Modern Theory of Critical Phenomena*. Taylor and Francis, London, 1996.

G. G. Emch. Quantum statistical mechanics. In J. N. Butterfield and J. Earman, editors, *Handbook of the Philosophy of Physics*, volume 2 of *Handbook of the Philosophy of Science*. North-Holland, 2006.

G. G. Emch and C. Liu. *The Logic of Thermo-Statistical Physics*. Springer-Verlag, Berlin, 2002.

G. G. Emch and C. Liu. Explaining spontaneous symmetry breaking. *Studies in History and Philosophy of Modern Physics*, 36(1):137–163, 2005.

P. K. Feyerabend. Explanation, reduction and empiricism. In H. Feigl and G. Maxwell, editors, *Minnesota Studies in the Philosophy of Science*, volume 3, pages 28–97. Dordecht Publishing Company, 1962.

R. P. Feynman. *QED: The Strange Theory of Light and Matter*. Princeton University Press, 1988.

A. Fine. The natural ontological attitude. In J. Leplin, editor, *Scientific Realism*. University of California Press, 1984.

M. E. Fisher. Renormalization group theory. *Reviews of Modern Physics*, 70(2): 653–682, 1998.

J. Fodor. Special sciences, or the disunity of sciences as a working hypothesis. *Synthese*, 28(97-115), 1974.

J. Fodor. Special sciences: Still autonomous after all these years. *Philosophical Perspectives*, 11:149–63, 1997.

P. Forrest. Bradley and realism about universals. *Idealistic Studies*, 14:200–212, 1984.

M. Friedman. Explanation and scientific understanding. *Journal of Philosophy*, 71(1):5–19, 1974.

H. M. Georgi. Effective field theory. *Annual Review of Nuclear and Particle Science*, 43(209), 1994.

N. Goldenfeld. *Lectures on Phase Transitions and the Renormalization Group*. Addison-Wesley, 1992.

D. M. Greenberger. Esoteric elementary particle phenomena in undergraduate physics. *American Journal of Physics*, 46(4), 1978.

R. B. Griffiths. Thermodynamic inequality near the critical point for ferromagnets and fluids. *Physical Review Letters*, page 623, 1965.

R. B. Griffiths. Peierls proof of spontaneous magnetization in a two-dimensional ising ferromagnet. *Physical Review*, 136(2A), 1964.

D. H. Gross. *Microcanonical Thermodynamics: Phase Transitions in Small Systems*. World Scientific, Singapore, 2001.

C. L. Hardin and A. Rosenberg. In defense of convergent realism. *Philosophy of Science*, 49:604–615, 1982.

S. Hartmann. Effective field theories, reductionism and scientific explanation. *Studies in History and Philosophy of Modern Physics*, 32B(2):267–304, 2001.

R. Healey. Holism and nonseparability. *Journal of Philosophy*, August 1991.

G. Hellman and F. Thompson. Physicalism: ontology, determination and reduction. *Journal of Philosophy*, 72:551–64, 1975.

C. G. Hempel. Reduction: Ontological and linguistic facets. In S. Morgenbesse, P. Suppes, and M. White, editors, *Philosophy, Science and Method, Essays in Honor of Ernest Nagel*. St Martin's Press, 1969.

C. G. Hempel and P. Oppenheim. Studies in the logic of explanation. *Philosophy of Science*, 15:135–175, 1948.

J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press, 1992.

J. H. Holland. *Emergence: From Chaos to Order*. Addison-Wesley, Redwood City, CA, 1998.

C. A. Hooker. Asymptotics, reduction and emergence. *British Journal for the Philosophy of Science*, 55(3):435–479, 2004.

T. Horowitz and G. Massey. *Thought Experiments in Science and Philosophy*. Rowman and Littlefield, Lanham, MI, 1991.

D. A. Howard. Reduction and emergence in the physical sciences: Some lessons from the particle physics-condensed matter physics debate. In D. A. Howard, N. Murphy, and W. R. Stoeger, editors, *Reductionism and Emergence: Implications for the Science/Theology Dialogue*. Forthcoming, 2005.

N. Huggett. Renormalization and the disunity of science. In M. Kuhlmann, H. Lyre, and A. Wayne, editors, *Ontological Aspects of Quantum Field Theory*. World Scientific, 2002.

N. Huggett and R. Weingard. The renormalisation group and effective field theories. *Synthese*, 102:171–194, 1995.

R. I. G. Hughes. The Ising model, computer simulation, and universal physics. In M. Morrison and M. S. Morgan, editors, *Models and Mediators: Perspectives on Natural and Social Science*, pages 97–145. Cambridge University Press, 1999.

P. Humphreys. How properties emerge. *Philosophy of Science*, 64:1–17, 1997.

F. Jackson. Epiphenomenal qualia. *Philosophical Quarterly*, 32:127–36, 1982.

F. Jackson. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Clarendon Press, Oxford, 1998.

E. Joos. Decoherence through interaction with the environment. In D. Giulini, E. Joos, C. Kiefer, J. Kupsch, I.-O. Stamatescu, and H. D. Zeh, editors, *Decoherence and the Appearance of a Classical World in Quantum Theory*, pages 35–136. Springer-Verlag, 1996.

L. P. Kadanoff. Scaling, universality and operator algebras. In C. Domb and M. S. Green, editors, *Phase Transitions and Critical Phenomena*, volume 5a, pages 1–34. Academic PRess, London, 1976.

L. P. Kadanoff. Scaling laws for Ising models near $T_c$. *Physics 2*, 1966.

L. P. Kadanoff. *Statistical Physics: Statics, Dynamics, and Renormalization*. World Scientific, Singapore, 2000.

M. Kaku. *Quantum Field Theory: A Modern Introduction*. Oxford University Press, 1993.

S. Kauffman. *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*. Oxford University Press, 1996.

S. H. Kellert. *In the Wake of Chaos*. The University of Chicago Press, 1993.

J. Kim. Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenolical Research*, 52(1):1–26, 1992.

J. Kim. Emergent properties. In T. Honerich, editor, *The Oxford Companion to Philosophy*. Oxford University Press, 1995.

J. Kim. *Philosophy of Mind*. Westview Press, 1996.

J. Kim. *Mind in a Physical World*. MIT Press, Cambridge, MA, 1998.

J. Kim. *Physicalism, Or Something Near Enough*. Princeton University Press, 2005.

J. Kim. *Supervenience and Mind*. Cambridge University Press, 1993.

P. Kitcher. Explanatory unification and the causal structure of the world. In P. Kitcher and W. Salmon, editors, *Scientific Explanation*, pages 410–505. University of Minnesota Press, 1989.

F. M. Kronz and T. Lupher. Unitarily inequivalent representations in algebraic quantum theory. *International Journal of Theoretical Physics*, 44(8), 2005.

F. M. Kronz and J. T. Tiehen. Emergence and quantum mechanics. *Philosophy of Science*, 69:324–347, 2002.

M. Kuhlmann, H. Lyre, and A. Wayne. *Ontological Aspects of Quantum Field Theory*. World Scientific, 2002.

C. Lanczos. *The variational principles of mechanics*. Dover Publications, New York, 4th edition, 1986.

L. D. Landau and E. M. Lifshitz. *Statistical Physics*. Pergamon, London, 1959.

L. Laudan. A confutation of convergent realism. *Philosophy of Science*, 1981.

R. B. Laughlin. *A Different Universe*. Basic Books, New York, 2005.

R. B. Laughlin and D. Pines. The theory of everything. *Proceedings of the National Academy of Sciences*, 97(1):28–31, 2000.

R. B. Laughlin, D. Pines, J. Schmalian, B. P. Stojkovic, and P. Wolynes. The middle way. *Proceedings of the National Academy of Sciences*, 97(1):32–37, 2000.

D. A. Lavis and G. M. Bell. *Statistical Mechanics of Lattice Systems*, volume 2. Springer, 1999.

R. Laymon. Thought experiments by Stevin, Mach and Guoy: Thought experiments as ideal limits and semantic domains. In *Horowitz and Massey (1991)*. Rowman and Littlefield, 1991.

A. Lesne. *Renormalization Methods: Critical Phenomena, Chaos, Fractal Structures*. Wiley, 1998.

G. H. Lewes. *The Physical Basis of Mind*. Problems of Life and Mind. Trübner, 1877.

D. K. Lewis. New work for a theory of universals. *Australasian Journal of Philosophy*, 61:343–377, 1983. Reprinted in Lewis (1999, 8-55), from which page references are taken.

D. K. Lewis. Against structural universals. *Australasian Journal of Philosophy*, 64:25–46, 1986a.

D. K. Lewis. *On the Plurality of Worlds*. Blackwell, 1986b.

D. K. Lewis. Humean supervenience debugged. *Mind*, 103(412):473–490, 1994.

D. K. Lewis. *Papers in Metaphysics and Epistemology*. Cambridge University Press, 1999.

C. Liu. Explaining the emergence of cooperative phenomena. *Philosophy of Science*, 66:S92–106, 1999.

C. Liu. Approximations, idealizations, and models in statistical mechanics. *Erkenntnis*, 60(2):235–263, 2004.

C. Liu. Infinite systems in SM explanations: Thermodynamic Limit, Renormalization (semi-) Groups and Irreversibility. *Philosophy of Science*, 68(S):S325–S344, 2001.

C. Liu. Spontaneous symmetry breaking and chance. http://philsci-archive.pitt.edu/archive/00000609/, 2002.

C. Lloyd Morgan. *Emergent Evolution*. Williams and Norgate, 1923.

B. Loewer. From physics to physicalism. In C. Gillet and B. Loewer, editors, *Physicalism and its Discontents*. Cambridge University Press, 2001.

P. R. Mainwood. Properties, permutations and physics. Master's thesis, University of Oxford, 2003.

H. J. Maris and L. P. Kadanoff. Teaching the renormalization group. *American Journal of Physics*, 46:652–657, 1978.

T. Maudlin. Part and whole in quantum mechanics. In E. Castellani, editor, *Interpreting Bodies: Classical and Quantum Objects in Modern Physics*, pages 46–60. Princeton University Press, 1998.

E. Mayr. *The Growth of Biological Thought*. Harvard University Press, 1982.

E. Mayr. The limits of reductionism. *Nature*, 331:475, 1988.

E. Mayr. *This is Biology*. Harvard University Press, 1997.

W. D. McComb. *Renormalization Methods: A guide for beginners*. Clarendon Press, 2004.

B. McLaughlin. The rise and fall of British Emergentism. In A. Beckermann, H. Flohr, and J. Kim, editors, *Emergence or Reduction*, pages 49–93. de Gruyter, Berlin, 1992.

E. McMullin. Galilean idealization. *Studies In History And Philosophy of Science*, 16(3):247–273, 1985.

D. H. Mellor. Truthmakers. In H. Beebee and J. Dodd, editors, *Truthmakers*. Oxford University Press, 2005.

A. Melnyk. How to keep the 'physical' in physicalism. *The Journal of Philosophy*, 94(12):622–637, 1997.

M. Morgan and M. Morrison. *Models as mediators : perspectives on natural and social science*. Cambridge University Press, Cambridge, 1999.

H. J. Morowitz. *The Emergence of Everthing: How the World became Complex*. Oxford University Press, 2002.

O. G. Mouritsen. *Computer studies of phase transitions and critical phenomena*. Springer-Verlag, Berlin, 1984.

E. Nagel. *The Structure of Science: problems in the logic of scientific explanation*. Routledge & Kegan Paul, 1961.

A. Nelson. Physical properties. *Pacific Philosophical Quarterly*, 66(268-282), 1985.

O. Neurath, R. Carnap, and C. Morris. *International Encyclopaedia of Unified Science*. University of Chicago Press, 1955.

D. V. Newman. Emergence and strange attractors. *Philosophy of Science*, 63(2): 245–261, 1996.

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

B. Nienhuis and M. Nauenberg. First-order phase transitions in renormalization-group theory. *Phys. Rev. Letters*, 35:477, 1975.

J. D. Norton. Causation as folk science. *Philosopher's Imprint*, 3(4), 2003. URL `http://www.philosophersimprint.org/003004/`.

I. Nowakwa and L. Nowak. *Idealization X: The Richness of Idealization*. Rodopi, Amsterdam, 2000.

T. O'Connor and H. Y. Wong. Emergent properties. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University, 2005. URL `http://plato.stanford.edu/entries/properties-emergent/`.

P. J. Olver. *Applications of Lie Groups to Differential Equations*. Springer, 2000.

D. Papineau. *Philosophical Naturalism*. Basil Blackwell, London, 1993.

P. Pettit. Microphysicalism without contingent micro-macro laws. *Analysis*, 55: 253–7, 1994.

P. Pettit. *The Common Mind. An essay on Psychology, Society and Politics*. Oxford University Press, 1996.

M. Polanyi. *Personal Knowledge*. New York, 1964.

J. Polchinski. Renormalisation and effective lagrangians. *Nuclear Physics*, 231B: 269–95, 1984.

J. Polchinski. Effective field theory and the Fermi surface. hep-th/9210046v2, 1999.

I. Prigogine. *The End of Certainty*. The Free Press, New York, 1997.

S. Psillos. *Scientific Realism: How Science Tracks Truth*. Routledge, 1999.

H. Putnam. *Mind, Language and Reality: Philosophical Papers*, volume 2. Cambridge University Press, 1975a.

H. Putnam. *On Properties*. Cambridge University Press, 1975b.

P. Railton. Probability, explanation and information. *Synthese*, 48:233–256, 1981.

M. Redhead. Asymptotic reasoning. *Studies in History and Philosophy of Modern Physics*, 35:527–530, 2004.

L. E. Reichl. *A Modern Course in Statistical Physics*. University of Texas Press, Austin, 1980.

R. Roberts. The reality of markets. *The Library of Economics and Liberty*, 2005.

D. Robinson. Matter, motion and humean supervenience. *Australasian Journal of Philosophy*, 67:394–409, 1989.

H. Robinson. Davidson and nonreductive materialism: A tale of two cultures. In C. Gillet and B. Loewer, editors, *Physicalism and its Discontents*. Cambridge University Press, 2001.

P. A. Ross. The limits of physicalism. *Philosophy of Science*, 66:94–116, 1999.

D. Ruelle. *Statistical Mechanics: Rigorous results*. W. A. Benjamin, New York, 1969.

L. Ruetsche. Interpreting quantum theories. In P. Machamer and M. Silberstein, editors, *The Blackwell Guide to the Philosophy of Science*. Blackwell Publishing, 2002.

L. Ruetsche. A matter of degree: Putting unitary inequivalence to work. *Philosophy of Science*, 70:1329–1342, 2003.

G. S. Rushbrooke. On the thermodynamics of the critical region for the Ising problem. *Journal of Chemical Physics*, 39:842, 1963.

W. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, 1984.

W. Salmon. Four decades of scientific explanation. In P. Kitcher and W. Salmon, editors, *Scientific Explanation*. University of Minnesota Press, 1989.

S. Saunders. Critical notice: Tian Yu Cao's "The conceptual development of 20th century field theories". *Synthese*, 136:79–105, 2003.

E. R. Scerri and L. McIntyre. The case for the philosophy of chemistry. *Synthese*, 111:213–232, 1997.

K. Schaffner. Reductionism in biology: Prospects and problems. *Proceedings of the Philosophy of Science Association, 1974*, pages 613–632, 1976.

S. S. Schweber. Physics, community and the crisis in physical theory. *Physics Today*, pages 34–40, November 1993.

A. Scott. *Nonlinear Science: Emergence and Dynamics of Coherent Structures*. Oxford University Press, 2 edition, 2003.

G. L. Sewell. *Quantum Mechanics and its Emergent Macrophysics*. Princeton University Press, 2002.

G. L. Sewell. Can the quantum measurement problem be resolved within the framework of Schroedinger dynamics? quant-ph/0512098 - to be Published in *Markov Processes and Related Fields*, 2007.

G. L. Sewell. *Quantum Theory of Collective Phenomena*. Oxford University Press, 1986.

L. A. Shapiro. Multiple realizations. *Journal of Philosophy*, 2000.

S. Shoemaker. Causality and properties. In P. van Inwagen, editor, *Time and Cause*, pages 228–254. Dordrecht, 1980.

M. Silberstein. Reduction, emergence and explanation. In M. Silberstein and P. Machamer, editors, *Blackwell Guide to the Philosophy of Science*. Blackwell Publishing, 2002.

M. Silberstein and J. McGeever. The search for ontological emergence. *The Philosophical Quarterly*, 49(195):182–200, 1999.

L. Sklar. *Physics and Chance : Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge University Press, 1993.

S. Solomon. Complexity - a science at 30. *Europhysics News*, 34(2), 2005.

R. Spencer-Smith. Reductionism and emergent properties. *Proceedings of the Aristotelian Society*, 117, 1995.

A. Stephan. Emergentism, irreducibility and downward causation. *Grazer Philosophische Studien*, 65:77–93, 2002.

D. Stoljar. Physicalism. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Stanford University, 2005. URL `http://plato.stanford.edu/entries/physicalism/`. The Stanford Encyclopedia of Philosophy (Winter 2005 Edition).

M. Suarez. Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science*, 17(3):225–244, 2003.

M. Suarez. An inferential conception of scientific representation. *Philosophy of Science (Supp.) PSA*, 2002.

G. J. Sussman and J. Wisdom. *Structure and interpretation of classical mechanics*. MIT Press, 2001.

B. Taylor. On natural properties in metaphysics. *Mind*, 102:81–100, 1993.

J. C. Taylor. *Hidden Unity in Nature's Laws*. Cambridge University Press, 2001.

P. Teller. *An Interpretative Introduction to Quantum Field Theory*. Princeton University Press, 1995.

P. Teller. Quantum mechanics and the nature of continuous physical quantities. *Journal of Philosophy*, pages 345–361, 1979.

C. J. Thompson. *Mathematical Statistical Mechanics*. Princeton University Press, 1972.

M. M. Waldrop. *Complexity: The Emerging Science at the Edge of Order and Chaos*. Simon & Shuster, 1992.

D. Wallace. Everett and structure. *Studies in the History and Philosophy of Modern Physics*, 34:87–105, 2003.

S. Weinberg. Newtonianism, reductionism and the art of congressional testimony. *Nature*, 330:433–437, 1987.

S. Weinberg. Conceptual foundations of the unified theory of weak and electromagnetic interactions. *Reviews of Modern Physics*, 52(3):515–523, 1980.

S. Weinberg. A reply to Mayr's 'The limits of reductionism'. *Nature*, 331:475, 1988.

S. Weinberg. Reductionism redux. *New York Times*, page A.27, July 8 1994.

S. Weinberg. Changing attitudes and the standard model. In L. Brown, M. Riodan, and M. Dresden, editors, *The Rise of the Standard Model*. Cambridge University Press, 1997.

J. A. Wheeler and K. Ford. *Geons, Black Holes and Quantum Foam: A Life in Physics*. W. W. Norton, 1998.

J. Wilson. Supervenience formulations of physicalism. *Nous*, 39:426–459, 2005.

K. Wilson. The renormalization group: Critical phenomena and the Kondo problem. *Reviews of Modern Physics*, 47(4):773–841, 1975.

K. Wilson. The renormalization group and critical phenomena. *Nobel Lectures, Physics 1981-1990*, pages 102–132, 1982/1993.

K. Wilson. Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture. *Phys. Rev. B*, 4:3174–3183, 1971a.

K. Wilson. Renormalization group and critical phenomena. II. Phase cell analysis of critical phenomena. *Phys. Rev. B*, pages 3184–3206, 1971b.

K. Wilson and M. E. Fisher. Critical exponents in 3.99 dimensions. *Phys. Rev. Letters*, 28:240, 1972.

K. G. Wilson and J. Kogut. The renormalization group and the $\epsilon$-expansion. *Physics Reports*, 12(2):75–199, 1974.

M. Wilson. What is this thing called 'pain'? The philosophy of science behind the contemporary debate. *Pacific Philosophical Quarterly*, 66:227–271, 1985.

M. Wilson. Theory facades. *Proceedings of the Aristotelian Society*, 104(1):273–288, 2004.

M. Wilson. Honorable intensions. In S. J. Wagner and R. Warner, editors, *Naturalism: A Critical Appraisal*, pages 53–94. Notre Dame University Press, 1993.

S. Wolfram. *A New Kind of Science*. Wolfram Media Inc., 2002.

J. Worrall. Structural realism: The best of both worlds? *Dialectica*, 43:99–124, 1989.

C. Yang and T. D. Lee. Statistical theory of equations of state and phase transitions I. *Phys. Rev.*, 87(3), 1952.

J. M. Yeomans. *Statistical Mechanics of Phase Transitions*. Clarendon Press, Oxford, 1992.

S. W. Yi. *How to Model Macroscopic Worlds: Towards the Philosophy of Condensed Matter Physics*. PhD thesis, London School of Economics, 2000.

D. Zimmerman. One really big liquid sphere: Reply to Lewis. *Australasian Journal of Philosophy*, pages 213–215, 1999.

W. H. Zurek. Decoherence and the transition from quantum to classical. *Physics Today*, 44(10):36–44, 1991.