Emergence, Singularities, and Symmetry Breaking

Robert W. Batterman*
University of Western Ontario, University of Pittsburgh

July 16, 2010

Contents

1 Introduction

In most philosophical discussions the concept of emergence is intimately related to the following notions: antireductionism, unpredictability, and novelty. Emergence in these contexts is also typically associated with parts and

1

wholes. The idea being that a phenomenon is emergent if its behavior is not reducible to some sort of sum of the behaviors of its parts, if its behavior is not predictable given full knowledge of the behaviors of its parts, and if it is somehow new—most typically this is taken to mean that emergent phenomenon displays causal powers not displayed by any of its parts.[1] In addition to irreducibility, unpredictability, and novelty, it is often asserted that emergent phenomena are inexplicable—they defy explanation in terms of the behaviors of their components. And, as "explanation" is typically understood to be explanation by a particular theory, this means that the behavior of the emergent whole is not fully explained by the theory that governs the behavior of the component parts.

A number of philosophers of science have imposed these philosophical conceptions upon physical theory in an attempt to address the issue of emergence in physics. While I believe there are some benefits to this methodological approach, on the whole I think it is better to turn the process on its head. We should look to physics and to "emergent relations" between physical theories to get a better idea about what the nature of emergence really is. Trying to impose a conceptual framework designed primarily to deal with the problem of the *mental's relation to the physical* is by and large unhelpful.

There have been discussions of emergence in the "modern" physics literature (though not necessarily always using that term) at least since the time of the development of kinetic theory and statistical mechanics. The questions addressed in these discussions focus on how it is that the macroworld of our experience can arise out of behavior of the microconstituents of the everyday objects of our experience. As with the philosophical discussions, most of these more historical questions presuppose a mereological or part/whole picture. However, relatively recent developments in physics show that this presupposition is not key to a proper understanding of emergence in physics. Rather, as I will try to argue, the most important features for understanding (at least some key examples of) emergence in physical theory are mathematical. In particular, one should take seriously the singularities and divergences that appear as one tries to understand the relationship between macro and micro theories, or more generally, between phenomenological theories at some energy scale and "more fundamental" theories at higher energy scales.

---

[1]Most philosophical discussions focus on emergent properties. As the discussion below will show, in the context of physical theory, I think this is a mistake and prefer to speak rather loosely of emergent "phenomena."

The next section discusses the not-so-clear-cut distinction between so-called "fundamental theory" and phenomenological theory. A theory may be fundamental, in that it gets a system's ontology correct, yet at the same time fail to be fundamental in other, particularly explanatory, ways. Section 3 begins to develop the theory of critical phenomena and the renormalization group. The main aim here is to begin to motivate the reasons for the explanatory inadequacy of ontologically correct theories for dealing with phase transitions and critical phenomena. In section 4 I discuss similar claims that have been made in the quantum field theory literature. In particular, this brief section presents the point of view of Roman Jackiw in his very suggestive paper "The Unreasonable Effectiveness of Quantum Field Theory." [Jackiw(1999)] Section 5, entitled "Necessary Singularities" makes the case, through a more detailed, yet still qualitative discussion of the renormalization group, that the infinite thermodynamic limit and the accompanying divergence of various quantities in that limit at criticality are absolutely essential for explanatory purposes. The idea is to argue that a potential finite theory, free of the infinities that plague quantum field theory and condensed matter physics, cannot explain the existence of emergent protected states of matter. Finally, the conclusion wraps up with some very brief remarks about the nature of scientific theory and the desideratum that such theories, to be acceptable, need to be consistent.

2 Phenomenological vs. Fundamental Theories

There is a sense in which classical thermodynamics is less fundamental than statistical mechanics. In classical thermodynamics a system such as a gas in a box is a continuous blob of stuff. The theory describes (and to some extent explains) the observable behavior of various systems such as gases and fluids while remaining completely agnostic about the internal make-up of those systems. Historically, of course, the developments of thermodynamics and kinetic theory/statistical mechanics were intertwined. [Brush(1983)] However, conceptually, the quantities and properties of state in orthodox thermodynamic equations appear largely to be independent of any specific claims about the ultimate constitution of the systems described. Similar claims hold for hydrodynamics—a continuum theory governed by the Navier-Stokes equa-

tions. It, too, remains largely agnostic about the molecular make-up of the fluids it describes.

On the other hand, as physical and philosophical orthodoxy would have it, the more fundamental theories—statistical mechanics and molecular dynamics—respectively, reduce and explain the phenomenological theories of thermodynamics and hydrodynamics. It is widely believed that statistical mechanics justifies the classical thermodynamic theory. (Indeed, in philosophical circles, the reduction of thermodynamics to statistical mechanics is taken to be the paradigm example of successful intertheoretic reduction.) Reduction in this context typically is taken to mean that the laws of thermodynamics (the reduced theory) are derivable from and hence explained by the laws of statistical mechanics (the reducing theory). [Nagel(1961), Batterman(2001), Batterman(2006)][2] To a certain extent this can in fact be done. The reductive project (again, in certain contexts) has been extraordinarily successful.

However, there are very good reasons to deny that all thermodynamic (and hydrodynamic) phenomena are reducible to "fundamental" theory. Consider phase transitions and critical phenomena. Everyday (and not so everyday systems) can undergo *qualitative* changes in state. Water in our freezers goes from its liquid phase to its solid phase (which then sublimates in frost-free freezers). Magnetic materials undergo transitions from ferromagnetic states to paramagnetic states. Such qualitative changes of state, as I will argue below, cannot be reductively explained by the more fundamental theories of statistical mechanics. They are indeed emergent phenomena. The reason for this (rather dramatic) negative claim has to do with the fact that such changes require certain *infinite* idealizations. From the point of view of the underlying fundamental theory whose proper focus is on the interactions of a *finite* number of molecular components of the macrosystems, these qualitative changes are genuinely novel.[3] The upshot is that the statistical mechanics of finite systems is explanatorily insufficient. While it gets the ontology of blobs of gases and fluids right, they are composed of a finite number of interacting molecules, there remain macroscopic phenomena—universal patterns

---

[2]Things are actually much more involved and complex than this brief description allows. See [Batterman(2002), Sklar(1993)] for a detailed discussion of issues of intertheoretic reduction.

[3]One starts all investigations, at least from the point of view of logical priority, by writing down the partition function for the N components of the systems. Such a partition function for finite N will never exhibit the nonanalytic behavior necessary to represent qualitatively different states of matter. See [Kadanoff(2000), pp. 238–239] for a discussion.

of behavior—that cannot be explained by this fundamental theory.

The term "fundamental" here is, therefore, ambiguous. A theory may be fundamental in that it properly characterizes the detailed constitution of the systems it studies, but can fail to be fundamental in its ability to explain and provide understanding of the systems it correctly describes. This point of view is expressed nicely in David R. Nelson's book *Defects and Geometry in Condensed Matter Physics.*[4]

> The modern theory of critical phenomena has interesting implications for our understanding of what constitutes "fundamental" physics. For many important problems, a fundamental understanding of the physics involved does *not* necessarily lie in the science of the smallest available time or length scale. The extreme insensitivity of the hydrodynamics of fluids to the precise physics at high frequencies and short distances is highlighted when we remember that the Navier-Stokes equations were derived in the early nineteenth century, at a time when even the discrete atomistic nature of matter was in doubt. [Nelson(2002), p.3]

3 Condensed Matter Theory

Contemporary references to "emergence" in the physics and popular science literature accord with Nelson's assessment of what counts a fundamental theory. Laughlin and Pines [Lauglin and Pines(2007)] use the term "protectorate" to describe domains of physics (states of matter) that are effectively independent of microdetails or of high frequency/short distance scales. A (quantum) protectorate, according to Laughlin and Pines is "a stable state of matter whose generic low-energy properties are determined by a higher organizing principle and nothing else." [Lauglin and Pines(2007), p. 261] Laughlin and Pines do not say much about what a "higher organizing principle" actually is, though they do refer to spontaneous symmetry breaking in this context.[5]

---

[4]See also Philip Anderson's famous paper "More is Different." [Anderson(1972)]

[5]For instance, they consider the existence of sound in a solid as an emergent phenomenon independent of microscopic details: "It is rather obvious that one does not need to prove the existence of sound in a solid, for it follows from the existence of elastic moduli at long length scales, which in turn follows from the spontaneous breaking of translational and rotation symmetry characteristic of the crystalline state." [Lauglin and Pines(2007),

In this section and the next I would like to suggest a way of understanding emergence in this sense (or at least a way of understanding the concept of a protectorate) that goes beyond mere claims to the effect that symmetry breaking occurs. I want to argue that one can tell a fairly persuasive story about why and how protectorates exist—one that goes beyond unexplicated appeals to higher organizing principles.

While not using the term "protectorate", Nelson also speaks of stable states of matter in a general way that directs us to where to find the details of the story. Instead of "protectorate" he uses the term "general fixed point." Here is a passage in which he describes this situation:

> It turns out that not just critical points but entire phases of matter are described by a "universal," coarse-grained, long-wavelength theory.... One can make similar statements about the hydrodynamic laws derived for fluids in the nineteenth century. Upon systematically integrating out the high-frequency, short-wavelength modes associated with atoms and molecules, one should be able to arrive at, say, the Navier-Stokes equations. ... Ignorance about microscopic details is typically packaged into a few phenomenological parameters characterizing the "fixed point," such as the density and viscosity of an incompressible fluid like water in the case of the Navier-Stokes equations. [Nelson(2002), p. 3.]

I suggest that in order to better understand what these protectorates are and how they are possible, we should look to the spectacular success of condensed matter theory in explaining the universality of critical phenomena. Protectorates or Nelson's generalized "fixed points" are intimately related to the concept of universality and the renormalization group theory developed by Kadanoff, Fisher, and Wilson. [Kadanoff(2000), Fisher(1983), Wilson and Kogut(1974)] It is to a brief discussion of this theory that we now turn. More details will follow in section 5.

Renormalization group theory provides an explanation for the remarkable similarity in behavior of "fluids" of different molecular constitution when at their respective critical points.[6] Experimentally it is found that a certain

---

p. 261]

[6]I put "fluids" in scare quotes because really there are many different kinds of systems that exhibit the same critical behavior including magnets and because, as we will see below, "fluids" can exhibit behavior that pre-theoretically we would *not* associate with fluids, i.e., liquids.
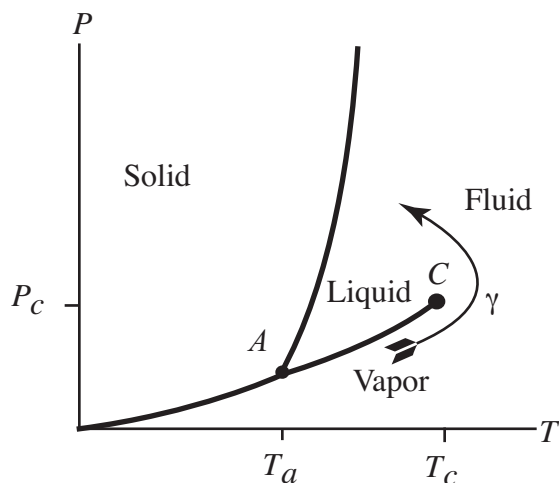
Figure 1: Temperature-Pressure Phase Diagram for a "Fluid"

dimensionless number—a critical exponent—characterizes the virtually identical behavior of diverse systems at their respective (and different) critical points. One would like to explain this remarkable universality of behavior or, in philosophical terms, how this *multiply realized behavior* is possible.[7] Let me briefly introduce some of this terminology.

Consider the temperature-pressure diagram for a typical "fluid" in figure 1. The bold lines represent thermodynamic states in which two distinct phases of the fluid (pairs of solid, liquid, and vapor) can coexist.[8] Thus, the lines represent states in which the system is undergoing a (first order) phase transition. For instance, along the line between points A and C one will find both liquid and vapor in the container—just what we see when the tea kettle boils on the stovetop. At point C, the critical point, something strange happens. Below the critical temperature T_c and below the critical pressure P_c, the fluid exists in its vapor phase. Further, it is impossible, below T_c and above T_a, to change the system from its vapor phase to the liquid phase by increasing pressure without crossing the line A–C. In other words, below T_c the system cannot pass from the vapor to the liquid phase without going through a state in which both vapor and liquid are simultaneously present in

---

[7]See [Batterman(2000)] for a discussion of the relationship between universality and multiple realizability.

[8]Point A is actually a triple point—a state in which all three phases can coexist.

7

the container. Above the critical temperature, $T > T_c$, it suddenly becomes possible to do this. Simply increase the temperature beyond T_c, increase the pressure beyond P_c and then decrease the temperature below T_c. This will be a path, γ, through states in which the "fluid" changes from vapor to liquid without ever being in a state where both phases coexist. The critical point indicates the existence of a qualitative change in the behavior of the system. Below T_c the distinction between liquid and vapor makes sense; above T_c, it apparently does not. Thermodynamically, the qualitative distinction between different states of matter is represented by a singularity in a function (the free energy) characterizing the system's state.

Thus, mathematical singularities in the thermodynamic equations represent *qualitative* differences in the physical states of the fluid in the container. As mentioned above, at the critical point systems with radically different microstructures exhibit behavior that, mathematically, is described by a particular number: the critical exponent β. The number, β, is the *same* for the diverse systems. At the simplest level, the critical behavior can be characterized in terms of the behavior of a certain quantity, the so-called "order parameter," which for fluids is the difference in densities between the different coexisting phases. So along the line A–C in figure 1 the order parameter, Ψ, is the difference between the liquid and vapor densities:

$$\Psi = |\rho_{liq} - \rho_{vap}|.$$

Below the critical temperature, T_c, Ψ is nonzero indicating the simultaneous existence of both liquid and vapor in the container. Above T_c, the order parameter takes the value zero. Figure 2 exhibits the vanishing of the order parameter Ψ at T_c. The heavy line is the coexistence curve with the vertical "tie lines" indicating that at some specific temperature T', the liquid density is $\rho_{liq}(T')$ and the vapor density is $\rho_{vap}(T')$. Note that at T_c the difference between these two values vanishes.

One universal feature–a feature displayed by all of the distinct fluids at their respective critical temperatures—is the fact that the coexistence curves near each of these T_cs have the same shape. We can introduce a "distance" measure called the "reduced temperature", t, that allows us to describe how far any system is from criticality. This allows us to compare the critical behavior of different systems with different T_cs—different critical temperatures. The reduced temperature is
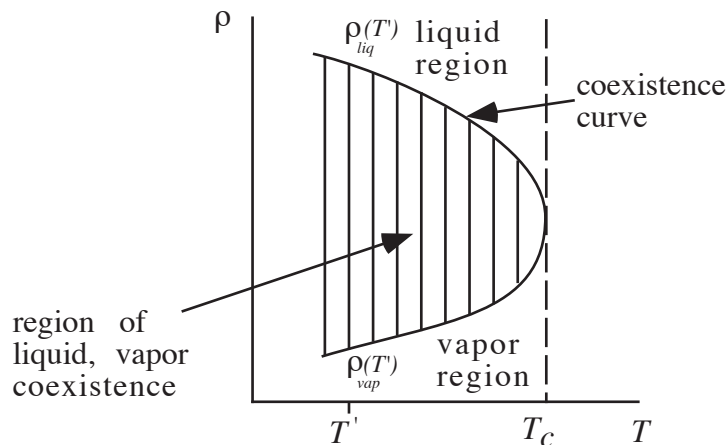
$$t = \left| \frac{T - T_c}{T_c} \right|,$$

8

Figure 2: Coexistence Curve: Density vs. Temperature for a "Fluid"

and the universality claim amounts to the fact that the order parameter, Ψ, for every fluid vanishes as some power β of t:

$$\Psi = |\rho_{liq} - \rho_{vap}| \propto t^{\beta}. \tag{1}$$

The number β characterizes the shape of the coexistence curve in the neighborhood of T_c. For instance, if the curve were a parabola, then β would be $1/2$—a result erroneously predicted by mean field theory. Experimentally, it has been found that the β is not $1/2$ but rather some number close to 0.33.

The explanatory question then is *why* the order parameters for various physically distinct fluids (and even magnets where the order parameter is the net magnetization) scale as a specific power law t^{β}. Condensed matter theorists have provided an explanation for this fact that involves the so-called renormalization group.[9] Without going into detail here, one essential feature of the explanation provided is the invocation of the so-called thermodynamic limit. This is the limit in which (roughly speaking) the number of particles of the system, e.g., the number of H_2O molecules in the tea kettle, approaches infinity. And, of course, as noted above, this is an idealization: water in real tea kettles consists of a finite number of molecules. This limiting idealization is essential for the explanation because for a finite number of particles the

---

[9]See [Kadanoff(2000)] for an account by one of the founders of the technique and [Batterman(2002), chapter 4] for more a more philosophical discussion.

statistical mechanical analogs of the thermodynamic functions cannot exhibit the nonanalytic behavior necessary to represent the qualitatively distinct behaviors we observe. [Kadanoff(2000), pp. 238–239]

In addition, the thermodynamic limit is required because unless we are dealing with infinite systems, there will always be some characteristic length scale (atomic spacings for instance) that will differ from system to system. The renormalization group explains the universal behavior at criticality essentially by exploiting the divergence (blow up to infinity) of the correlation length (a parameter reflecting how far-apart molecules are correlated with one another). The divergence of this parameter allows for the comparison of the different fluids at criticality by allowing for scaling or self-similar solutions. See the discussion below in section 5.

So we have an explanation for the universality of critical phenomena that depends essentially upon infinities and divergences. The explanation in essence delimits a class of macrostates of various systems at the scale of everyday objects (fluids) that are essentially decoupled or independent of their microdetails. These are the protectorates of Laughlin and Pines. [Lauglin and Pines(2007)]

The renormalization group explanation provides principled physical reasons (reasons grounded in the physics and mathematics of systems in the thermodynamic limit) for ignoring details about the microstructure of the constituents of the fluids. It is, in effect, an *argument* for why those details are irrelevant for the behavior of interest. What then of Laughlin's and Pines' higher organizing principles? Recall that the protectorates are supposed to depend upon such principles "*and nothing else.*" It is true, as we can see in figure 2, that as the critical temperature is approached from above T_c there is a breaking of symmetry in that the order parameter undergoes a discontinuous jump from zero to nonzero values at that temperature. But is this a higher organizing principle? And if it is, does appeal to it *by itself*—that is without the entire renormalization group story—suffice for understanding the emergent protectorate with its "insensitivity to microscopics"? It seems hardly satisfactory to appeal to symmetry breaking as an organizing principle independent of microdetails when we have such a profoundly successful story about why the microdetails in fact are largely independent or irrelevant.

Most crucial to the renormalization group explanation is, as noted, the ineliminable appeal to the thermodynamic limit and to the singularities that emerge in that limit. This raises issues about the use of idealized models in explanatory contexts. Many physicists and philosophers apparently believe

10

that singularities appearing in our theories are indications of modeling failures. In particular, they hold that such blowups are signatures of some kind of unphysical assumption in the underlying model. Singularities are, on this view, information *sinks*. We cannot learn anything about the physical system until we rid the theory of such monstrosities. (This attitude is still quite prevalent in the literature on quantum field theory, a little bit of which will be discussed in the next section.) On the contrary, I'm suggesting that an important lesson from the renormalization group successes is that we rethink the use of models in physics. If we include mathematical features as essential parts of physical modeling then we will see that blowups or singularities are often *sources* of information.

Mark Wilson has noted something akin to this in his discussion of determinism in classical physics. He notes that while from a

> *modeling point of view* we are inclined to object to the appearance of singularities ..., from a *mathematical point of view* we often greatly value these same breakdowns; for as Riemann and Cauchy demonstrated long ago, the *singularities* of a problem commonly represent the precise features of the mathematical landscape we should seek in our efforts to understand how the qualitative mathematics of a set of equations unfolds. Insofar as the project of *achieving mathematical understanding* goes, singularities frequently prove our best friends, not our enemies. [Wilson(2009), pp.184–185]

As mathematical features (particularly singular limits) are often part of applied mathematical modeling, it seems to me that we ought to expand our understanding of Wilson's "modeling point of view" to include the mathematical singularities.

The next section looks briefly at infinities arising in quantum field theory with the goal of supporting the view that singularities can be sources of information.

4 Quantum Fields

Quantum field theory is without a doubt one of the most successful frameworks ever developed in physics. Field-theoretic ideas pervade physics at

various energy and length scales and have resulted in some of the most spectacular agreements with experiment imaginable. The standard model of particle physics is one such example, as is the use of field-theoretic techniques in condensed matter theory. Quantum field theory has predicted and explained various unexpected phenomena such as fractional quantum numbers in the Hall effect, and various aspects of superconductivity. Nevertheless, (and this is still widely discussed) the theory has been deemed by many to be foundationally suspect. Those who hold that a successful theory should yield predictions from-first-principles, as it were, independent of experimental/phenomenological input believe that it cannot be the final theory. And so the search for a deeper theory—perhaps, string theory—continues. But more upsetting to many is the fact that quantum field theory when actually used for calculations and predictions typically engenders all kinds of divergences. With these monsters ever present, it is claimed that there must be something wrong with the foundations of the theory.

Most worrisome are the divergences that appear in high energy regimes—the so-called ultraviolet divergences. As Roman Jackiw notes, such ultraviolet infinities

> do appear to be intrinsic to quantum field theory, and no physical consideration can circumvent them; unlike the infrared [low energy] divergences, ultraviolet ones cannot be excused away. But they can be 'renormalized'. This procedure allows the infinities to be sidestepped or hidden, and succeeds in unambiguously extracting numerical predictions from the standard model and from other 'physical' quantum field theories .... [Jackiw(1999), pp. 149–150]

Jackiw's goal in this paper is to persuade us that "the divergences of quantum field theory must not be viewed as unmitigated defects; on the contrary they convey crucially important information about the physical situation, without which most of our theories would not be physically acceptable." [Jackiw(1999), p. 150] This fits well with the remarks at the end of the last section; in particular, with Wilson's claim that singularities are often our best mathematical friends.

One crucial and obvious feature of the world is that different things—particles, organisms, etc.—exist or appear at different scales. The world is not scale invariant with respect to space and time. Jackiw puts the point as follows: Dilating (or contracting) spacetime coordinates will generally change

the units of measurement for space and time. But, if a theory possesses no intrinsic dimensional space and time parameters, then that theory will be scale invariant.

> Our quantum chromodynamical model (QCD) for quarks is free of such dimensional parameters, and it would appear that this theory is scale invariant—but Nature certainly is not! The observed variety of different objects with different sizes and masses exhibits many different and inequivalent scales. ...Fortunately, scale symmetry is quantum mechanically broken, owing to the scales that are introduced in the regularization and renormalization of ultraviolet singularities. [Jackiw(1999), p. 155]

If it were not for the singularities that appear in our theories and models we would have no understanding of the emergence at different scales of distinct and apparently "protected" states of matter.

Most interesting, from a philosophical perspective, is that the infinities appear to arise from the mathematical description employed in the theories themselves. Of course, this raises the question of whether the infinities are merely artifacts of our field theoretic formulation. Perhaps, some other framework—maybe algebraic quantum field theory or, more radically, a correct theory of everything (string theory?)—will provide an infinity-free theoretical account of our non-scale invariant world? Jackiw expresses this worry as follows:

> ...[F]or me chiral and scale symmetry breaking are completely natural effects, but their description in our present language—quantum field theory—is awkward and leads to extreme formulations, which make use of infinities. One hopes that there is a more felicitous description, in an as yet undiscovered language. ...
>
> The relation between mathematical entities and field theoretical anomalies was realized 20 years ago and has led to a flourishing interaction between physics and mathematics, which today culminates in the string program. However, it seems to me that now mathematical ideas have taken the lead, while physical understanding lags behind. In particular, I wonder where within the completely finite and non-local dynamics of string theory we are

to find the mechanisms for symmetry breaking that are needed to describe the world around us. [Jackiw(1999), pp. 157–158]

Jackiw here nicely expresses the tension felt by many about the appearance and role of infinities in quantum field theories. On the one hand, infinities seem necessary for establishing and understanding the different stable states of matter—the different protectorates. On the other hand, they seem "infelicitous," awkward, and extreme. This latter attitude, I believe, comes from a deep feeling that infinities or divergences are indicators of the ultimate *inadequacy* of our theoretical constructs.[10] I think this feeling is inapt and in the next section I provide positive reasons to believe that singularities are necessary for our understanding of emergent protectorates.

5   Necessary Singularities

In this section I would like to *assume* that emergent protectorates exist— that is, that there are stable states of matter that are in effect decoupled and largely independent of physics at shorter length/higher energy scales. The question then is: How are such protectorates possible? Can a more fundamental theory (such as string theory) answer this question without appeal to singularities and infinities? While I'm not remotely an expert on string theory, I think there are some general reasons to think that more fundamental theories without divergences are incapable of answering certain why-questions about such protectorates.

To begin, let's consider what Laughlin and Pines say about this in their paper "The Theory of Everything." [Lauglin and Pines(2007)] They point out that even if we had a theory of everything the victory of reductionism would only be pyrrhic: "We have succeeded in reducing all of ordinary physical behavior to a simple, correct Theory of Everything only to discover that it has revealed exactly nothing about many things of great importance". [Lauglin and Pines(2007), p.260] Their reasons here depend upon the fact that *ab initio*, from-first-principles, deductions of these "things of great importance" such as macro-behavior of ferromagnets and superconductivity al-

---

[10]In many instances, I think investigators hold that the existence of these infinities indicate that the theory is inconsistent. I believe this is a mistake. Singularities *may* be an indication that something is not right, but they need not entail the logical inconsistency of the theory. Michael Berry has long advocated a role for singularities in our understanding of physical theory. [Berry(1994), Berry(2002)]

14

most always fail. One needs experimental data, and approximate calculations are achieved, in their words,

> not [from] first-principle deductions, but are rather *art keyed to experiment,* and thus tend to be the least reliable precisely when reliability is most needed, i.e., when experimental information is scarce, the physical behavior has no precedent, and the key questions have not yet been identified. [Lauglin and Pines(2007), p.260, My emhapsis.]

However, as we've seen above Jackiw allows the possibility that a completely finite, nonlocal string theory may actually be the correct theory—one that might provide a more "felicitous" description and, possibly because of that, a means for from-first-principles deductions.

Recall that the emergent protectorates are, according to Laughlin and Pines "regulated by higher organizing principles [and] have a property, namely their insensitivity to microscopics, that is directly relevant to the broad question of what is knowable in the deepest sense of the term." [Lauglin and Pines(2007), p. 261] Our brief discussion in section 3 suggests how to understand this "insensitivity to microscopics" or better, the *universal* behavior of lower energy protectorates.

The key explanatory feature of the renormalization group argument for the universality of critical phenomena is the demonstration that broad classes of physical Hamiltonians (corresponding to distinct microstructures, e.g., different real fluids) all belong to the same universality class. That is, in an abstract space of Hamiltonians, there are fixed points toward which different systems at criticality flow under an imposed renormalization group transformation. The class of such systems that flow to the same fixed point—the basin of attraction of that fixed point—is the universality class. Crucial to establishing this fact (simplifying somewhat) is that in the thermodynamic limit the correlation length diverges. This means that there is a loss of characteristic length scale (say the atomic spacing) that can distinguish different systems from one another. This loss of scale enables one to demonstrate the equivalence in the behavior of different systems (specifically the fact that their order parameters exhibit power law behavior with the *same* exponent, β as in equation (1)) at criticality.

In fact, *without* the loss of a characteristic length scale, there would be no way to compare different systems to one another; no way to demonstrate the similarity in their behaviors. It is worthwhile going into a bit of detail

15

about how this takes place. (For more details and for a thorough discussion see Kadanoff [Kadanoff(2000), Chapter 12].)

We have seen in section 3 that the order parameter Ψ for a fluid is the difference in densities of the vapor and liquid in the container. Suppose we consider some region in the container and ask about the spatial extent of a 'droplet' of near liquid density in the surrounding sea of vapor. Because of fluctuations in density, this region may contain a liquid droplet of larger than average size. Nevertheless, for a fluid in equilibrium at a given temperature away from criticality, there is a well-defined average size for the droplets. One can now introduce the correlation length, ξ, which, roughly speaking, characterizes the spatial extent of the average droplets of liquid. (Put slightly differently: The correlation length is the typical distance over which the behavior of one microscopic variable or degree of freedom can be correlated with the behavior of another.) If all parameters other than temperature are fixed, ξ is a function of temperature, $\xi(T)$. A crucial fact is that at the critical temperature, the correlation length diverges to infinity:

$$\xi(T) \to \infty \quad \text{as} \quad T \to T_c.$$

As another example consider a lattice of spins above the critical temperature T_c and in zero external magnetic field. Neighboring spins on the lattice interact in a way that tends to align them parallel to each other. This accounts for the ferromagnetic nature of the lattice. At high temperatures, however, thermal energy tends to randomize the directions of the spins. Thus, while the spin-spin interactions tend to correlate the spins, at higher temperatures this tendency can be overridden, resulting in a loss of ferromagnetism. So, at high temperatures regions (or blocks) of spins will have a very small correlation length, and the net magnetization will be zero. As the temperature decreases towards the critical or Curie temperature, the spatial extent of the blocks with correlated spins will increase. The correlation length, ξ, is a measure of the size of the correlated regions or blocks. As for the fluid, the correlation length is a function of T which diverges at the critical point—the point of transition from the paramagnetic phase (above T_c) to the ferromagnetic phase.

So near the critical temperature, the correlation length becomes extremely large. This means that an enormous number of degrees of freedom are coupled together. In other words, the range of the interaction among the degrees of freedom (e.g., the different spins in a correlated block of spins) is large. These correlations make solving the governing equations extremely difficult.

The idea behind the renormalization group method is, in part, to relate an intractable problem involving a large correlation length to a more tractable problem characterized by a correlation length reduced by some factor. One thereby reduces the number of coupled degrees of freedom and demonstrates scaling or self-similar behavior. For example, suppose one has a system of spins on a lattice of spacing 1.0, as shown in figure 3a) (round dots). Let us now group the spins into blocks of four spins. Next suppose that we have a way (via some kind of summation or partial integration) of replacing the grouped spins with a single new kind of spin—a 'block' spin—in the center of the block (square dots). The block spins, in this example, are now located on a lattice of spacing 2.0. Next, figure 3b), one transforms the lengths so that the new spins are arranged on the same lattice sites as the original spins. In other words, all lengths are contracted by a factor of two. Finally, figure 3c), one transforms the new spin variables so that the 'new' system of spins will be as much like the original system as possible.

As mentioned, each system is represented by a function—its so-called Hamiltonian. This function characterizes the kinds of interactions between the degrees of freedom (e.g., between the spins) as well as any effects of external fields. Thus, the Hamiltonian can be considered to be a function of various system parameters—in particular, the temperature, the coupling constants characterizing the spin-spin interactions, and the external fields.[11] So, figure 3 in effect schematizes a transformation from one Hamiltonian to another in which, among other things, the correlation length of the new, transformed, Hamiltonian is smaller by the spatial rescaling factor which in this case is 2.

If one performs this 'renormalization' transformation (call it 'τ') repeatedly, one gets a sequence of Hamiltonians all of which describe systems with the same lattice spacing, but where the correlation length gets smaller and smaller with each iteration. As a result, the new Hamiltonians describe systems with fewer and fewer coupled degrees of freedom within their correlation lengths. Thus, the systems are farther and farther from criticality.[12]

---

[11]Strictly speaking, since this function is explicitly a function of the temperature, it is not truly a *Hamiltonian*. The idea is that temperature is just another parameter describing the system. It does not play any special role. So, whenever we speak of a system in this context, we mean a state of a system at a given temperature.

[12]This process introduces all sorts of new couplings between the various degrees of freedom. Dealing with this increase of complexity is part of the 'art' of the method. It is not necessary to discuss, here, the various technicalities involved.

a)

spatial | rescaling

1.0

2.0

b)

1.0

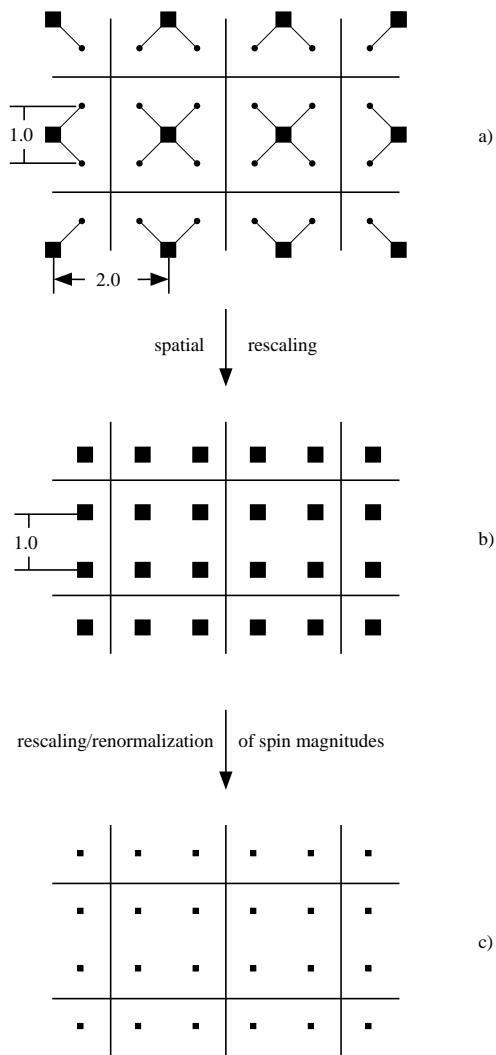rescaling/renormalization | of spin magnitudes

c)

Figure 3: The operations of the renormalization group: a). Block formation (partial integration); b). Spatial rescaling/contraction; c). Renormalization of parameters.

Now, let us consider an abstract space whose coordinates are the parameters appearing in the various Hamiltonians of the systems. In effect, every point of this space corresponds to a possible Hamiltonian, so the space should be thought of as the space of all possible Hamiltonians. Consider, in particular, a lattice system undergoing a ferromagnetic phase transition. Suppose that the external fields and all of the coupling constants are fixed, and that the temperature is the only parameter being varied. As the temperature approaches the critical temperature T_c, the point representing the system undergoing the phase transition moves about in the abstract space of Hamiltonians.[13] Following the discussion in [Pfeuty and Toulouse(1977), pp. 12–15] let us call this path the 'physical line' of the system.

Each point on the physical line represents the system at a given temperature with a correlation length considered here to be solely a function of temperature: $\xi(T_i) = \xi_i$.
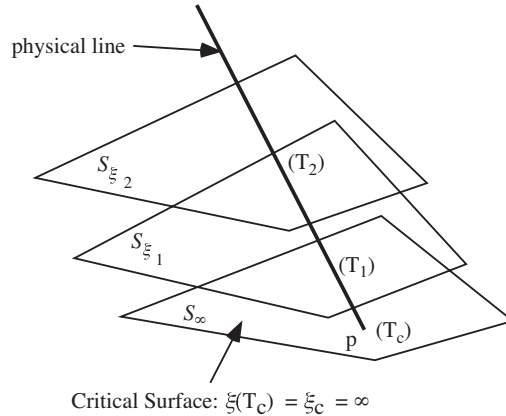

Figure 4: Surfaces of constant correlation length. $\xi(T_1) = \xi_1$, $\xi(T_2) = \xi_2$ where $\xi_1 > \xi_2$.

One can now imagine this space of Hamiltonians divided up into surfaces of constant ξ. Figure 4 illustrates the general idea.[14] The surface, S_∞, is the 'critical surface' corresponding to parameter values (points in the space, i.e., Hamiltonians) having infinite correlation length. Under the renormalization

---

[13]See footnote 11.

[14]A very nice, fairly intuitive discussion (upon which this presentation is based), can be found in [Pfeuty and Toulouse(1977), chapter 1].

group transformation τ every point on the physical line gets mapped to another point in the space of Hamiltonians. This yields a trajectory issuing from that point on the physical line. *An important feature here is that the trajectory generated by τ from the point p on the critical surface S_∞ remains confined to the critical surface.*

The reason for this is simply that under the renormalization transformation, lengths are contracted by a spatial rescaling factor 'b'[15]; and

$$\tau(\xi_c) = \frac{\xi_c}{b} = \frac{\infty}{b} = \infty.$$

In other words, the correlation length remains infinite. The transformations of points other than p on the physical line yield trajectories that diverge from the critical surface S_∞ intersecting surfaces S_ξ corresponding to successively lower values of correlation length, thereby effecting the reduction of the number of degrees of freedom within a correlation length. This is illustrated in figure 5.
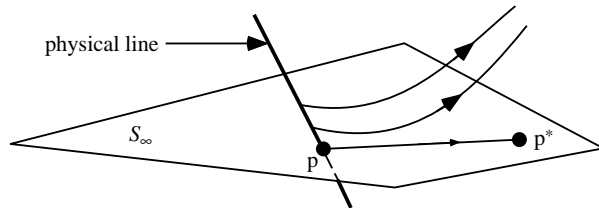

Figure 5: Renormalization group trajectories issuing from points on the physical line. p^* is a fixed point of the transformation τ.

The point, p^*, shown in figure 5 is a "fixed point." This is a point in the parameter space which, under τ, is its own trajectory. That is, it represents a state of a system which is invariant under the renormalization group transformation. Of necessity, such a fixed point has *an infinite correlation length* and so lies on the critical surface S_∞.[16] The singularity/divergence of the correlation length ξ is *necessary.*

---

[15]In figure 3, $b = 2$.

[16]Actually, a fixed point could have zero correlation length, since zero is also a length which is unchanged under division by b. Such fixed points are called 'trivial,' and are usually ignored. See [Kadanoff(2000), pp. 257–258].

We can now explain the universality of critical phenomena. Note that a fixed point of a transformation is a property of the transformation itself. This means that to find them one must solve the fixed point equation:

$$\tau(\mathcal{H}^*) = \mathcal{H}^*.$$

That is, one must determine the fixed point Hamiltonian $\mathcal{H}^* = p^*$ and this is *independent* of any choice of initial Hamiltonian—the details of the individual systems do not matter. Universal behavior is explained by reference to properties of certain fixed points. More precisely, it is related to the stability of the fixed points and to how the renormalization group transformation τ maps points in the neighborhood of the fixed points. To illustrate, consider the simplest visualizable case.

Let us imagine that the critical surface S_∞ is two dimensional. We want to consider the contour structure of the surface. Consider figure 6.[17] One should think of the point b as if it were the top of a mountain. The trajectories of the renormalization group then are lines of steepest descent from the summit. Point c, on the other hand, is like the bottom of a valley. Finally, a represents a pass (or saddle point).
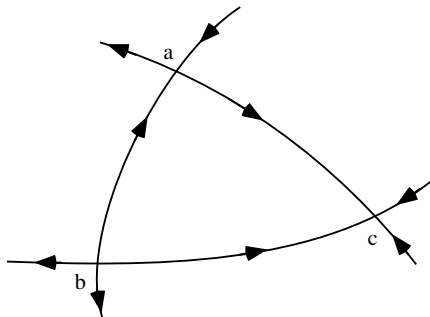

Figure 6: Possible fixed points on a two-dimensional critical surface.

It is clear that these fixed points are characterized by differing degrees of stability. For instance, the peak, b, is the least stable—any point in the parameter space near to, but not identical with b will flow, under τ, away from b. The valley, c, is the most stable—nearby points will be taken towards c by τ. The saddle, a, exhibits some sort of intermediate degree of stability,

---

[17]See [Pfeuty and Toulouse(1977)], pp. 13–14.

being stable in one direction and unstable in the other. These are the most important of the possible fixed points.

One can perform an analysis (linearization) in the neighborhood of such a fixed point which shows how the renormalization transformation acts on points which differ only slightly from the fixed point itself. This yields local axes which determine how lengths are scaled in different directions around the fixed point. In turn, this analysis allows for the calculation of the critical exponents—the number β appearing in equation (1). The fixed point describes, therefore, a specific critical behavior whose salient properties are determined by the local behavior of the renormalization group transformation around the fixed point. *Furthermore,* and this is where we finally get an understanding of the universality of this critical behavior*, corresponding to each fixed point is a domain of attraction. These are all those points in the abstract parameter space (the space of all Hamiltonians) which, under the transformation, τ, eventually flow into the fixed point.*

How does this provide an explanation for the universality of critical phenomena? Figure 7 helps indicate how.[18] Suppose that the manifold/surface, N, of physical Hamiltonians represents different states of nickel near its ferromagnetic critical point.[19] Under the renormalization transformation the points (Hamiltonians) on the manifold are mapped to other points in the parameter space. The physical critical point (corresponding to the nickel at the phase transition) has infinite correlation length and so, as we have seen, it gets mapped to another point with infinite correlation length. This defines the critical trajectory which is the same as the line from p to p^* in figure 5. Thus it lies on the critical surface. Suppose that physical manifold, I, corresponds to some other metal, say, iron; and likewise that physical manifold, G, corresponds to some third material, gadolinium. Notice that the critical trajectories from the physical manifolds N, I, and G all flow to the same fixed point p^*. They must, therefore, all lie on the critical surface S_∞. Now, recall that the critical behavior—in particular, the set of critical exponents—is a function of the nature of the renormalization group flow near the relevant fixed point. Thus, nickel, iron, and gadolinium must all have their phase transitions characterized by the same critical exponents: *They all lie in the basin of attraction of the same fixed point.* This is the key to

---

[18]See the discussion in [Fisher(1983)], pp. 84–87.

[19]The 'physical line' of figures 4 and 5 is a line on this physical manifold with an endpoint at the physical critical point.
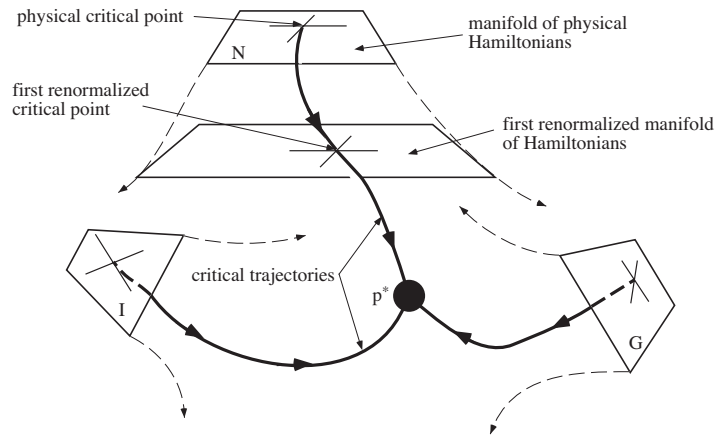
Figure 7: Universality of critical phenomena. Critical trajectories are bold lines. p^* is a common fixed point for the three distinct physical critical points on the manifolds N, I, and G. (See [Fisher(1983)].)

understanding the universality of critical phenomena.

Notice the absolutely essential role played by the divergence of the correlation length, ξ, in this explanatory story. It is this that opens up the possibility of a fixed point solution to the renormalization group equations. Without that divergence and the corresponding loss of characteristic scale, no calculation of the exponent β would be possible. The divergence of ξ is possible only in the thermodynamic limit and it is this that allows both for the simplified treatment of the different systems at criticality and more detailed predictions of the behavior of near critical systems.

On the assumption with which we began this section—the assumption that emergent protected states of matter exist—we see that such emergent phenomena can be understood provided there are means for eliminating unwanted or irrelevant degrees of freedom. On a finite theory without divergences, it is difficult, if not impossible, to see how this can be accomplished. While one may be able to tell detailed (microstructurally *dependent*) stories about why *individual* fluids/magnets behave the way they do at criticality, such stories simply *cannot account for the key property of the emergent protectorates*—namely, "their insensitivity to microscopics." [Lauglin and Pines(2007), p. 261]

Furthermore, the explanatory picture outlined briefly in this section goes

well-beyond any unexplicated appeal to higher organizing principles. Laughlin and Pines are correct to hold that from-first-principle derivations the existence of protectorates are not to be had. They are wrong to claim that the existence of such protectorates depend *only* upon higher organizing principles like spontaneous symmetry breaking. One can locate the necessary explanatory features in the mathematical singularities that appear in the thermodynamic limit. Such infinities/singularities play absolutely crucial roles in our theories of diverse states of matter and are not unreasonable at all.

Let me just briefly respond to a question that may arise at this point.[20] Perhaps it is true that in order to explain the universality of critical phenomena, we need to appeal to a mathematical singularity. But, non-critical thermodynamic systems also exhibit universal behavior from the point of view of statistical mechanics: We are virtually guaranteed (with probability approaching one) to get the right values for equilibrium quantities by calculating phase averages using the uniform (microcanonical) measure over states in phase space. This universality, however, doesn't appear to require any sort of singularity.

My response to this is to note that the very same kind of argument (a renormalization group argument) provides the justification for the use of the microcanonical measure by establishing the universality of the "critical exponent" (1/2) appearing the gaussian probability distribution. Why are the distributions for values of quantities displayed by systems away from criticality best characterized by gaussian or normal probability distributions? (In other words, why the ubiquity of normal distributons?) The answer is given by an RG argument that essentially demonstrates that the gaussian is the fixed point characterizing a large universality class of probability distributions that all exhibit the same limiting–normal behavior. And, following Khinchin, one can appeal features of the normal distrbution to justify the calculation of equilibrium values for thermodynamic quantities as phase averages. For the details see, [Khinchin(1949), Sinai(1992), Batterman(1998), Batterman(2010)] and references therein. Thus, the same kind of argument explains certain universal behavior of non-critical systems. Note, though, that in such situations we are not interested in qualitative changes in the states of matter—there are no (thermodynamic) singularities to try to capture. Put another way, *the issue of emergence of new states of matter does*

_____

[20]Thanks to an anyonomous referee for raising it.

not arise in this context.

6 Conclusion

The logical positivists and empiricists typically held scientific theories to be axiomatized sets of sentences in some language. Explanation, following Hempel [Hempel(1965)] and theory reduction, following Nagel [Nagel(1961)] were taken to involve deductive relations between various sentences. Given an inconsistent set of sentences, it is possible in first order logic to derive *any* sentence or set of sentences whatsoever. As a result, the logical consistency of our scientific theories was taken to be a minimum requirement on their acceptance. After all, a theory that is inconsistent provides no information about the world at all.

Theories of physics such as quantum field theory, as we've seen involve considerably more mathematics than that provided by deductive logical relations among sets of sentences. And, often, as we also noted, predictions and explanations involve detailed mathematical calculations involving various limits and "naturally" arising series expansions, with many of the latter in various contexts, ultimately divergent. These divergences are the targets of renormalization in quantum field theory. Part of the debate about the merits of such divergences depends, I believe, upon the mistaken inference that divergences and infinities entail inconsistency. Ultimately, I believe, this is what lies behind the idea that singularities are information sinks—they are indications that the theory cannot be counted on to legitimately provide information about the world. To the contrary, I'm suggesting here that such *mathematical* singularities are often sources of information about the world. They neither entail inconsistency nor inadequacy.[21]

Theories that effectively decouple from one another at different scales are natural places to find divergences and divergent series expansions. And so, if this sort of decoupling is related to the existence of emergent protected states of matter, it is natural to expect that infinities and divergences to be deeply involved in the understanding of how those protectorates are possible. The question with which we began section 5 concerned how such protectorates, with their characteristic independence from higher energy scales—their insensitivity to microscopics—are possible. *A priori* the answer to this "how is

---

[21]This raises a number of interesting philosophical questions about the relationship between mathematics and physics. For a discussion see [Batterman(2009)].

it possible" question, has to be a story about why the detailed microscopics of the various macrostable phenomena (e.g., universal behavior at criticality) are *irrelevant*. That is, any understanding of the existence of such protectorates must provide an account of why the microdetails that genuinely distinguish one system in the protectorate from another (each fluid, for example, has a distinct critical temperature that does depend upon the nature of its constitutive molecules) are irrelevant for the universal behavior of interest. This is exactly what the renormalization group strategy of Kadanoff, Fisher, and Wilson provides. And it does so by explicitly exploiting the infinities that arise in the thermodynamic limit.

If the goal is to answer the "how is it possible" question about universal behavior, it seems highly unlikely that a scheme that depends upon a mathematical derivation from the finite, "true," theory of everything can fulfill that role. After all, such a derivational story must, of necessity start from the detailed microstructural constitution of the individual molecules in a *particular* fluid. But *why should that individual derivation* have any bearing on a completely different individual derivation for a different fluid with a potentially radically different microstructural constitution? The singularities that appear when considering the relationships between micro-theories at some high energy/short distance scale and those theories at smaller energies/larger lengths are absolutely necessary for an understanding of emergent phenomena in physics.

References

[Anderson(1972)] P. W. Anderson. More is different. *Science*, 177(4047): 393–396, 1972. ISSN 00368075. URL http://www.jstor.org/stable/1734697.

[Batterman(1998)] R. W. Batterman. Why equilibrium statistical mechanics works: Universality and the renormalization group. *Philosophy of Science*, 65:183–208, June 1998.

[Batterman(2001)] Robert W. Batterman. *Intertheory Relations in Physics*. The Stanford Encyclopedia of Philosophy, http://plato.stanford.edu/entries/physics-interrelate/, 2001.

[Batterman(2002)] Robert W. Batterman. *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford Studies in Philosophy of Science. Oxford University Press, 2002.

[Batterman(2006)] Robert W. Batterman. *Encyclopedia of Philosophy*, chapter Reduction. Macmillan Reference, 2nd edition, 2006.

[Batterman(2009)] Robert W. Batterman. On the explanatory role of mathematics in empirical science. *The British Journal for the Philosophy of Science*, doi = 10.1093/bjps/axp018:1–25, 2009. URL http://bjps.oxfordjournals.org/cgi/content/abstract/axp018v1.

[Batterman(2000)] Robert W. Batterman. Multiple realizability and universality. *The British Journal for the Philosophy of Science*, 51:115–145, 2000.

[Batterman(2010)] Robert W. Batterman. Reduction and renormalization. In Gehard Ernst and Andreas Hüttemann, editors, *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics*, pages 159–179. Ca, 2010.

[Berry(2002)] Michael V. Berry. Singular limits. *Physics Today*, pages 10–11, May 2002.

[Berry(1994)] Michael V. Berry. Asymptotics, singularities and the reduction of theories. In Dag Prawitz, Brian Skyrms, and Dag Westerståhl, editors, *Logic, Methodology and Philosopohy of Science, IX: Proceedings of*

the Ninth International Congress of Logic, Methodology and Philosophy of Science, Uppsala, Sweden, August 7–14 1991, Studies in Logic and Foundations of Mathematics, pages 597–607, Amsterdam, 1994. Elsevier Science B. V.

[Brush(1983)] Stephen G. Brush. *Statistical Physics and the Atomic Theory of Matter, From Boyle and Newton to Landau and Onsager*. Princeton Series in Physics. Princeton University Press, New Jersey, 1983.

[Fisher(1983)] Michael E. Fisher. Scaling, universality and renormalization group theory. In F.J.W. Hahne, editor, *Critical Phenomena*, volume 186 of *Lecture Notes in Physics*, Berlin, 1983. Summer School held at the University of Stellenbosch, South Africa; January 18–29, 1982, Springer-Verlag.

[Hempel(1965)] Carl G. Hempel. Aspects of scientific explanation. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, pages 331–496. The Free Press, 1965.

[Jackiw(1999)] Roman Jackiw. The unreasonable effectiveness of quantum field theory. In Tian Yu Cao, editor, *Conceptual Foundations of Quantum Field Theory*, pages 148–159, Cambridge, 1999. Cambridge University Press.

[Kadanoff(2000)] Leo P. Kadanoff. *Statistical Physics: Statics, Dynamics, and Renormalization*. World Scientific, Singapore, 2000.

[Khinchin(1949)] A. I. Khinchin. *Mathematical Foundations of Statistical Mechanics*. Dover, New Yor, 1949.

[Lauglin and Pines(2007)] Robert B. Lauglin and David Pines. The theory of everything. In Mark A. Bedau and Paul Humphreys, editors, *Emergence: Contemporary Readings in Philosophy and Science*, pages 259–268. The MIT Press, 2007.

[Nagel(1961)] Ernest Nagel. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace, & World, 1961.

[Nelson(2002)] David R. Nelson. *Defects and Geometry in Condensed Matter Physics*. Cambridge University Press, Cambridge, 2002.

[Pfeuty and Toulouse(1977)] P. Pfeuty and G. Toulouse. *Introduction to the Renormalization Group and to Critical Phenomena.* John Wiley and Sons, London, 1977. Translation: G. Barton.

[Sinai(1992)] Ya. G. Sinai. *Probability Theory: An Introductory Course.* Springer-Verlag, Berlin, 1992. Trans.: D. Haughton.

[Sklar(1993)] Lawrence Sklar. *Physics and Chance: Philosophical Issues in the Foundations of Statstical Mechanics.* Cambridge University Press, Cambridge, 1993.

[Wilson and Kogut(1974)] Kenneth G. Wilson and J. Kogut. The renormalization group and the ϵ expansion. *Physics Reports*, 12(2):75–199, 1974.

[Wilson(2009)] Mark Wilson. Determinism and the mystery of the missing physics. *The British Journal for the Philosophy of Science*, 60:173–193, 2009.