CHAPTER 7

THE TYRANNY OF SCALES

ROBERT W. BATTERMAN

1. Introduction

In this essay I will focus on a problem in physics and applied mathematics. This is the problem of modeling across scales. Many systems, say a steel girder, manifest radically different, dominant behaviors at different length scales. At the scale of meters, we are interested in its bending properties, its buckling strength, etc. At the scale of nanometers or smaller, it is composed of many atoms, and features of interest include lattice properties, ionic bonding strengths, etc. To design advanced materials (such as certain kinds of steel), materials scientists must attempt to deal with physical phenomena across 10+ orders of magnitude in spatial scales. According to a recent (2006) NSF research report, this "tyranny of scales" renders conventional modeling and simulation methods useless as they are typically tied to particular scales (Oden 2006, p. 29). "Confounding matters further, the principal physics governing events often changes with scale, so the models themselves must change in structure as the ramifications of events pass from one scale to another" (Oden, pp. 29–30). Thus, even though we often have good models for material behaviors at small and large scales, it is often hard to relate these scale-based models to each other. Macroscale models represent the integrated effects of very subtle factors that are practically invisible at the smallest, atomic, scales. For this reason it has been notoriously difficult to model realistic materials with a simple bottom-up-from-the-atoms strategy.

I would especially like to thank Mark Wilson for many helpful discussions. Conversations with Penelope Maddy, Julia Bursten, and Nic Fillion are also much appreciated. This research was supported, in part, by a grant from the Social Sciences and Humanities Research Council of Canada.

The widespread failure of that strategy forced physicists interested in overall macro-behavior of materials toward completely top-down modeling strategies familiar from traditional continuum mechanics.[1]

A response to the problem of the "tyranny of scales" would attempt to exploit our rather rich knowledge of intermediate micro- (or meso-) scale behaviors in a manner that would allow us to bridge between these two dominant methodologies. Macroscopic scale behaviors often fall into large common classes of behaviors such as the class of isotropic elastic solids, characterized by two phenomenological parameters—so-called elastic moduli. Can we employ knowledge of lower scale behaviors to understand this universality—to determine the moduli and to group the systems into classes exhibiting similar behavior? This is related to engineering concerns as well: Can we employ our smaller scale knowledge to better design systems for optimal macroscopic performance characteristics?

The great hope that has motivated a lot of recent research into so-called "homogenization theory" arises from a conviction that a "between-scales" point of view, such as that developed by Kadanoff, Fisher, and Wilson in the renormalization group approach to critical phenomena in fluids and magnets, may very well be the proper methodological strategy with which to begin to overcome the tyranny of scales. A number of philosophers have recently commented on the renormalization group theory, but I believe their focus has overlooked what is truly novel about the methodological perspective that the theory employs.

Philosophical discussions of the applicability of mathematics to physics have not, in my opinion, paid sufficient attention to contemporary work on this problem of modeling across scales. In many instances, philosophers hold on to some sort of ultimate reductionist picture: whatever the fundamental theory is at the smallest, basic scale, it will be sufficient in principle to tell us about the behavior of the systems at all scales. Continuum modeling on this view represents an *idealization*—as Feynman has said, "a smoothed-out imitation of a really much more complicated microscopic world" (Feynman, Leighton, and Sands 1964, p. 12). Furthermore, the suggestion is that such models are in principle eliminable.

There is a puzzle however. Continuum model equations such as the Navier-Stokes equations of hydrodynamics or the equations for elastic solids work despite the fact that they completely (actually, almost completely—this is crucial to the discussion below) ignore small scale or atomistic details of various fluids. The recipe (I call it "Euler's continuum recipe") by which we construct continuum models is safe: if we follow it, we will most always be led to empirically adequate successful equations characterizing the behavior of systems at the macroscopic level. Why? What explains the safety of this recipe? Surely this requires an answer. Surely, the answer must have something to do with the physics of the modeled systems at smaller scales. If such an answer cannot be provided, we will be left with a kind of skepticism: without such an answer, we cannot expect anything like a unified conception of applied mathematics' use of continuum idealizations.[2] If an

[1] For related discussions, see Mark Wilson's forthcoming *Physics Avoidance and Other Essays.*
[2] See Maddy (2008) for a forceful expression of this skeptical worry.

answer is forthcoming, then we have to face the reductionist picture mentioned above. Will such an answer—an answer that explains the robustness and safety of employing continuum modeling—support the view that continuum models are mere conveniences, only pragmatically justified, given the powerful simplifications gained by replacing large but finite systems with infinite systems? As noted, many believe that a reductionist/eliminitivist picture is the correct one. I maintain that even if we can explain the safety and robustness of continuum modeling (how this can be done is the focus of this essay), the reductionist picture is mistaken.

It is a mistaken picture of how science works. My focus here is on a philosophical investigation that is true to the actual modeling practices of scientists. (I am not going to be addressing issues of what might be done in principle, if not in practice.) The fact of the matter is that scientists do not model the macroscale behaviors of materials using pure bottom-up techniques.[3] I suggest that much philosophical confusion about reduction, emergence, atomism, and antirealism follows from the absolute choice between bottom-up and top-down modeling that the tyranny of scales apparently forces upon us. As noted, recent work in homogenization theory is beginning to provide much more subtle descriptive and modeling strategies. This new work calls into question the stark dichotomy drawn by the "do it in a completely bottom-up fashion" folks and those who insist that top-down methods are to be preferred.

The next section discusses the proposal that the use of continuum idealizations present no particular justificatory worries at all. Recent philosophical literature has focused on the role of continuum limits in understanding various properties of phase transitions in physical systems such as fluids and magnets. Some authors, particularly Jeremy Butterfield (2011) and John Norton (2011), have expressed the view that there are no particularly pressing issues here: the use of infinite limits is perfectly straightforwardly justified by appeal to pragmatic considerations. I argue that this view misses an important difference in methodology between some uses of infinite limits and those used by renormalization group arguments and homogenization theory.

In section 3, I present an interesting historical example involving nineteenth century attempts to derive the proper equations governing the behavior of elastic solids and fluids. A controversy raged throughout that century concerning the merits of starting from bottom-up atomic description of various bodies in trying to arrive at empirically adequate continuum equations. It turns out that the bottom-up advocates lost the debate. Correct equations apparently could only be achieved by eschewing all talk of atomic or molecular structure, advocating instead a top-down approach supplemented, importantly, with experimentally determined data. In section 4, I formulate the tyranny of scales as the problem, just mentioned, of trying to understand the connection between recipes for modeling at atomic scales (Euler's discrete recipe) and Euler's continuum recipe appropriate for continuum models. Finally, I present a general discussion of work on homogenization that

[3] Those who think that the renormalization group provides a bottom-up explanation of the universality of critical phenomena, e.g. Norton (2011), are mistaken, as we shall see below.

provides at least the beginning of an answer to the safety question and to the problem of bridging scales between the atomic and the continuum. This research can be seen as allaying skeptical worries about a unified applied mathematical methodology regarding the use of continuum idealizations of a certain kind.

2. Steel Beams, Scales, Scientific Method

Let us consider the steel girder in a bit more detail. In many engineering applications steel displays linear elasticity. This is to say that it obeys Hooke's Law—its strain is linearly proportional to its stress. One phenomenological parameter related to its stress/strain (i.e., stiffness) properties is Young's modulus appearing in the equations of motion for solids, as well as in equilibrium and variational equations. At scales of 1 meter to 10^{-3} meters, say, the steel girder appears to be almost completely homogeneous: zooming in with a small microscope will reveal nothing that looks much different. In fact, there appears to be a kind of local scale invariance here.[4] So for behaviors that take place within this range of scales, the steel girder is well-modeled or represented by the Navier-Cauchy equations:

$$(\lambda + \mu)\nabla(\nabla \cdot \mathbf{u}) + \rho\nabla^2\mathbf{u} + \mathbf{f} = 0.^5 \tag{1}$$

The parameters λ and μ are the "Lamé" parameters and are related to Young's modulus.

Now jump from this large-scale picture of the steel to its smallest *atomic* scale. Here the steel, for typical engineering purposes, is an alloy that contains iron and carbon. At this scale, the steel exhibits highly ordered crystalline lattice structures. It looks nothing like the homogeneous girder at the macroscales that exhibits no crystalline structure. Somehow between the lowest scale of crystals on a lattice and the scale of meters or millimeters, the low-level ordered structures must disappear. But that suggests that properties of the steel at its most basic, atomic level cannot, by themselves, determine what is responsible for the properties of the steel at macroscales. I will discuss this in more detail below.

In fact, the story is remarkably complex. It involves appeal to various geometrical properties that appear at *microscales* intermediate between the atomic and the macro,[6] as well as a number of other factors such as martensitic transformations.[7] The symmetry breaking is effected by a combination of point defects, line defects,

[4] "Local" in the sense that the invariance holds for scales of several orders of magnitude but fails to hold if we zoom in even further, using x-ray diffraction techniques, for example.

[6] I call these intermediate scales "microscales" and the structures at these scales "microstructures" following the practice in the literature, but it may be best to think of them as "mesoscopic."

[7] These latter are transformations that take place under cooling when a relatively high symmetry lattice such as one with cubic symmetry loses symmetry to become tetragonal. Some properties of steel girders therefore depend crucially on dynamical changes that take place at scales in between the atomic and the macroscopic (Phillips 2001, p. 547–8).

slip dislocations, and higher dimensional wall defects that characterize interfacial surfaces. All of these contribute to the *homogenization* of the steel we see and manipulate at the macroscale. And, of course, in engineering contexts the macro features (bending properties, for example) are the most important—we do not want our buildings or bridges to collapse.

2.1 Reduction, Limits, Continuum Models

A simpler case than steel involves trying to connect the finite statistical mechanical theory of a fluid at the atomic scale to the thermodynamic continuum theory at macro scales.[8] The relationship between statistical mechanics and thermodynamics has received a lot of attention in the recent philosophical literature. Debates about intertheoretic reduction, its possibility, and its nature have all appealed to examples from thermodynamics and statistical mechanics. Many of these discussions, in the recent literature, have focused on the nature and potential emergence of phase transitions in the so-called thermodynamic limit[9] (Butterfield 2011a; Menon and Callender 2012; Belot 2005; Bangu 2009). What role does the thermodynamic limit play in connecting theories? What role does it play in understanding certain particular features of thermodynamic systems? It will be instructive to consider the role of this limit in a more general context than that typical of the literature. This is the context in which we consider the generic problem of upscaling from atomic to laboratory scales, as in the case of the steel girder discussed above. In doing this, I hope it will become clear that many of the recent philosophical discussions miss crucial features of the methodology of applying limits like the thermodynamic limit.

Before turning to the debates about the use of the thermodynamic limit and the justification of using infinite limits to understand the goings on in finite systems, I think it is worthwhile to step back to consider, briefly, some general issues about theory reduction. As mentioned above, many philosophers and physicists tacitly (and sometimes explicitly) maintain some sort of in principle reductionist point of view. I do not deny that maybe in some as yet to be articulated sense there may be an in principle bottom-up story to be told. However, appeals to this possibility ignore actual practices and furthermore are never even remotely filled out in any detail. Typically the claim is simply: "The fundamental theory (whatever it is, quantum mechanics, quantum field theory, etc.), because it is fundamental (whatever that ultimately means), *must* be able to explain/reduce everything."

Nagel's seminal work (1961) considered the reduction of thermodynamics to statistical mechanics to be a straightforward and paradigm case of intertheoretic

[8] Though simpler than the case of understanding how atomic aspects of steel affect its phenomenological properties, this is, itself, a difficult problem for which a Nobel prize was awarded.

[9] This is the limit in which the number of particles N in a system approaches infinity in such a way that the density remains constant—the volume has to go to infinity at the same time as the number of particles.

reduction. On his view, as is well known, one derives the thermodynamic laws from the laws of statistical mechanics employing so-called bridge laws connecting terms/predicates appearing in the reduced theory with those appearing in the reducing theory.[10] In several places I have argued that this Nagelian strategy and its variants fail for many cases of so-called reduction (Batterman 1995, 2002). I have argued that a limiting sense of reduction in which, say, statistical mechanics "reduces to" thermodynamics in an appropriate limit (if it does) provides a more fruitful conception of intertheoretic reduction than the Nagelian strategies where the relation seems to go the other way around: on the Nagelian strategies one has it that thermodynamics reduces to statistical mechanics, in the sense of deductive derivation. There are a number of reasons for thinking the nonNagelian, "limiting," sense of reduction is a superior sense of reduction. For one, there is the difficulty of finding the required definitional connections that the bridge laws are meant to embody.[11] But in addition, the kinds of connections established between theories by taking limits do not appear to be expressible as definitional extensions of one theory to another. In many cases, the limits involved are singular, and even when they are not, the use of mathematical limits invokes mathematics well beyond that expressible in the language of first order logic—a characteristic feature of Nagel's view of reduction and of its neoNagelian refinements.

Despite these arguments a number of authors have recently tried to argue that reduction should be understood in Nagelian terms; that is, as the definitional extension of one theory to another. Jeremy Butterfield and Nazim Bouatta, for example,

> …take reduction as a relation between theories (of the systems concerned). It is essentially deduction; though the deduction is usually aided by adding appropriate definitions linking two theories' vocabularies. This will be close to endorsing the traditional philosophical account [Nagel's], despite various objections levelled against it. The broad picture is that the claims of some worse or less detailed (often earlier) theory can be deduced within a better or more detailed (often later) theory, once we adjoin to the latter some appropriate definitions of the terms in the former. …So the picture is, with D standing for the definitions: $T_b \& D \Rightarrow T_t$. In logicians' jargon T_t is a *definitional extension* of T_b. (Butterfield and Bouatta 2011)

In the current context the more ***b***asic, ***b***etter theory (statistical mechanics) is T_b and the reduced, ***t***ainted theory (thermodynamics) is T_t.[12]

[10] See Batterman 2001 and 2006 for surveys of this and more sophisticated strategies.

[11] In the present example, it is hard indeed to see how to define or identify a nonstatistical quantity such as temperature or pressure in thermodynamics with a necessarily statistical quantity or set of quantities in the reducing statistical mechanics. (See Sklar 1993, Chapter 9.)

[12] I believe that the use of the evaluative terms "better," "worse," and "tainted" reflects an inherent prejudice against nonreductionist points of view. In particular, as one of the issues is whether a more detailed (atomic)

Butterfield and Bouatta obviously are not moved by the objections to the Nagelian scheme that I briefly mentioned above. I suggest though, as we delve a bit more deeply into the examples of phase transitions and of the steel girder, that we keep in mind the question of whether the continuum account of the bending behavior of the steel can be reduced to the theory of its atomic constituents in the sense that we can derive that continuum behavior from the "better," "more detailed," and "later" atomic theory. Even if we extend the logicians' sense of deduction (as definitional extension) beyond that of first order logic so as to include inferences involving mathematical limits, will such a deduction/reduction be possible?

So the real question, as both of these examples employ continuum limits, concerns why the use of such limits is justified. The debate about the justification of the use of infinite limits and, ultimately, about reduction concerns whether the appeal to limits can in the end be eliminated. It is a pressing debate, because no party thinks that at the most fundamental level, the steel girder is a continuum. And no party thinks that a tea kettle boiling on the stove contains an infinite number of molecules. What justifies our employing such infinite idealizations in describing and explaining the behaviors of those systems?

For Butterfield there is a "Straightforward Justification" for the use of infinite limits in physical modeling.

> This Justification consists of two obvious, very general, broadly instrumentalistic, reasons for using a model that adopts the limit $N = \infty$: mathematical convenience, and empirical adequacy (up to a required accuracy). So it also applies to *many* other models that are almost never cited in philosophical discussions of emergence and reduction. In particular, it applies to the many classical continuum models of fluids and solids, that are obtained by taking a limit of a classical atomistic model as the number of atoms N tends to infinity (in an appropriate way, e.g. keeping the mass density constant). (2011, p. 1080)

He continues by emphasizing two "themes" common to the use of many different infinite models:

> The first theme is abstraction from finitary effects. That is: the mathematical convenience and empirical adequacy of many such models arises, at least in part, by abstracting from such effects. Consider (a) how transient effects die out as time tends to infinity; and (b) how edge/boundary effects are absent in an infinitely large system.
>
> The second theme is that the mathematics of infinity is often much more convenient than the mathematics of the

theory is really better for explanatory, predictive, and modeling concerns, this way of speaking serves to block debate before it can get started.

> large finite. The paradigm example is of course the convenience of the calculus: it is usually much easier to manipulate a differentiable real function than some function on a large discrete subset of $\mathbb{R}$ that approximates it. (2011, p. 1081)

The advantages of these themes are, according to Butterfield, twofold. First, it may be easier to know or determine the limit's value than the actual value primarily because of the removal of boundary and edge effects. Second, in many examples of continuum modeling we have a function defined over the finite collection of atoms or lattice sites that oscillates or fluctuates and so can take on many values. In order to employ the calculus we often need to "have each *value* of the function defined as a limit (namely, of values of another function)" (pp. 1081–82). Butterfield seems to have in mind the standard use of *averaging* over a "representative elementary volume" (REV)[13] and then taking limits $N \to \infty$, volume going to zero, so as to identify a continuum value for a property on the macroscale. In fact, he cites continuum models of solids and fluids as paradigm examples:

> For example, consider the mass density varying along a rod, or within a fluid. For an atomistic model of the rod or fluid, that postulates N atoms per unit volume, the average mass-density might be written as a function of both position $\mathbf{x}$ within the rod or fluid, and the side-length L of the volume L^3 centred on $\mathbf{x}$, over which the mass density is computed: $f(N, \mathbf{x}, L)$. Now the point is that for fixed N, this function is liable to be intractably sensitive to $\mathbf{x}$ and L. But by taking a continuum limit $N \to \infty$, with $L \to 0$ (and atomic masses going to zero appropriately so that quantities like density do not "blow up"), we can define a continuous, maybe even differentiable, mass-density function $\rho(x)$ as a function of position—and then enjoy all the convenience of the calculus.
>
> So much by way of showing in general terms how the use of an infinite limit $N = \infty$ can be justified—but not mysterious! At this point, the general philosophical argument of this paper is complete! (p. 1082)

So for Butterfield most of the discussions concerning the role, and particularly the justification, of the use of the thermodynamic limit in the debates about phase transitions have generated a lot of hot air. The justification, on his view, for employing such limits in our modeling strategies is largely pragmatic—for the sake of convenience. In addition, there is, as he notes, the further concern that the use of such limits be empirically adequate—getting the phenomena right to within appropriate error bounds. Much of his discussion concerns showing that the use of such

[13] I have taken this terminology from Hornung (1997, p. 1).

limits can most always be shown to be empirically adequate in this sense (Butterfield 2011). Unfortunately, I think that sometimes things are more subtle than the straightforward justification admits. In fact, there are good reasons to think that the use of the thermodynamic limit in the context of the renormalization group (RG) explanation of critical phenomena—one of the cases he highlights—fails to be justified by his own criteria. It is a different methodology, one that does not allow for the sort of justificatory story just told. The straightforward story as described above cannot be told for the RG methodology for the simple reason that that story fails to be empirically adequate in those contexts.

One can begin to understand this by making a distinction between what might be called "ab initio" and "post facto" computational strategies. Butterfield's remarks about the mass density in a rod (say a steel girder) in one sense appear to endorse the ab initio strategy. Consider a model of the rod at the scale of atoms where the atoms lock together on a crystal lattice. The limit averaging strategy has us increase the size of the lattice until we have, in effect, a perfect crystal of infinite extent. This lets us ignore boundary effects as he notes. The limiting average density that we arrive at using this ab initio (atomic only) strategy will actually be grossly incorrect at higher scales. This is because, at higher micro (meso) scales real iron contains many structures such as dislocations, grain boundaries, and other metastabilities that form within its mass and that energetically allow local portions of the material at these higher scales to settle into stable modes with quite different average densities. See figure 7.1. These average densities will be quite different than the ab initio calculations from the perfect crystal. What special features hidden within the quantum chemistry of iron bond allow those structures to form? We really don't know. But until we gain some knowledge of how those structures emerge, we will not be able accurately to determine computationally the bulk features of steel girders in the way described. Values for Young's modulus and fracture strength that we may try determine on the basis of this ab initio reasoning will be radically at variance with the actual measured values for real steel.

On the other hand, if we possessed a realistic model of steel at all length scales, then we could conceivably define a simple average over a representative volume (at a much higher scale than the atomic). But this post facto calculation would rely upon complete data about the system at all scales. No limits would be involved whatsoever. Perhaps some super genius may someday in principle propose an incredibly detailed model of iron bonding that would allow the calculation of the macro parameters like Young's modulus in a kind of ab initio mode imagined by Butterfield, but such a hypothetical project is certainly not the aim of the RG techniques that are under consideration here.

Such ab initio calculations provide wrong answers because they cannot "see" the energetically allowed local structural configurations that steel manifests at larger scales. On the other hand, if we are investigating materials that (for whatever reason) display nice scaling relationships across some range of scales (as steel does for scales 8–10 orders of magnitude above the atomic), then we will be able to employ RG type techniques to determine the various universality classes (characterized by

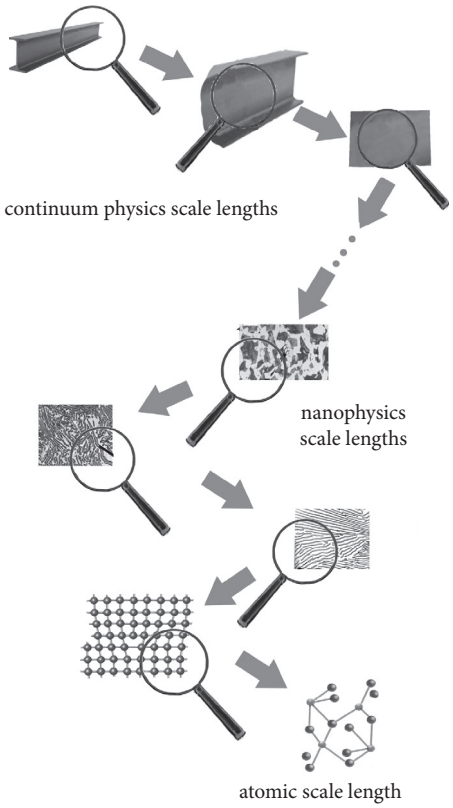


Figure 7.1 Microstructures of steel

the phenomenological parameters—Young's modulus, e.g.) into which they must fall. Thus the RG methodology, unlike the ab initio REV averaging strategy, provides a rationale for evading extreme bottom-up computations so as to gain an understanding of why steel, for example, only requires a few effective parameters to describe its behavior at macroscales.

While there surely are cases in which averaging is appropriate, and the straightforward justification may be plausible, there are other cases, as I have been arguing, in which it is not. In order to further elucidate this point, I will say a bit about what the RG argument aims to do. I will then give a very simple example of why one should, in many instances, expect the story involving averaging over a representative volume element (REV) to fail. In fact, the failure of this story is effectively the motivation behind Wilson's development of the distinct RG methodology. More generally, if our concern is to understand why continuum models such as the Navier-Cauchy equation are safe and robust, the straightforward justification will miss what is most crucial.

I have discussed the RG in several publications (Batterman 2002; 2005; 2011). Butterfield (2011) and Butterfield and Bouatta (2011) present concise descriptions as well. For the purposes here, as noted earlier, I am going to present some of the details with a different emphasis than these other discussions have provided.

In particular, I want to stress the role of the RG as part of a methodology for upscaling from a statistical theory to a hydrodynamic/continuum theory. In so doing, I follow a suggestion of David Nelson (2002, pp. 3–4) who builds on a paper of Ken Wilson (1974). The suggestion is that entire phases of matter (not just critical phenomena) are to be understood as determined by a "fixed point" reflecting the fact that "universal physical laws [are] insensitive to microscopic details" (2002, p. 3). Specifically, the idea is to understand how details of the atomic scale physics get encoded (typically) into a few phenomenological parameters that appear in the continuum equations governing the macroscopic behavior of the materials. In a sense, these phenomenological parameters (like viscosity for a fluid, and Young's modulus for a solid) characterize the appropriate "fixed point" that defines the class of material exhibiting universal behavior despite potentially great differences in microscale physics.

Let us consider a ferromagnet modeled as a set of classical spins σ_i on a lattice—the Ising model. In this model, neighboring spins tend to align in the same direction (either up or down: $\sigma_i = \pm 1$). Further, we might include the effect of an external magnetic field B. Then the Hamiltonian for the Ising model is given by

$$H[\{\sigma_i\}] = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j + \mu B \sum_i \sigma_i,$$

with the first sum over nearest neighbor pairs of spins, μ is a magnetic moment. A positive value for the coupling constant J reflects the fact that neighboring spins will tend to be aligned, both up or both down.

For ferromagnets we can define an order parameter—a function of the net magnetization for the system—whose derivative exhibits a discontinuity or jump at the so-called critical or Curie temperature, T_c. Above T_c, in zero magnetic field, the spins are not correlated due to thermal fluctuations and so the net magnetization is zero. As the system cools down to the Curie temperature, there is singularity in the magnetization (defined as a function of the free energy). (See figure 7.2.) The

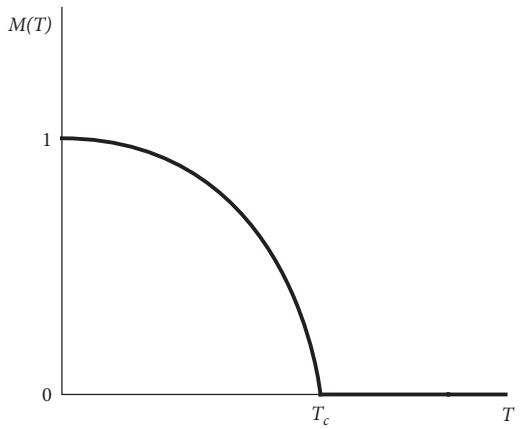


Figure 7.2 Spontaneous magnetization at T_c

magnetization exhibits power law behavior near that singularity characterized by the relation

$$M \propto |t|^{\beta},$$

where t is the reduced temperature $t = \frac{T-T_c}{T_c}$. It is a remarkable fact that physically quite distinct systems—magnets modeled by different Hamiltonians, and even fluids (whose order parameter is the difference between vapor and liquid densities in a container)—all exhibit the same power law scaling near their respective critical points: The number β is universal and characterizes the phenomenological behavior of a wide class of systems at and near criticality.[14]

The RG provides an explanation for this universal behavior; and in particular, it allows one theoretically to determine the value for the exponent β. For the 3-dimensional Ising model, that theoretical value is approximately .33. Experimentally determined values for a wide class of fluids and magnets are found in the range .31–.36. So-called "mean field" calculations predict a value of .5 for β (Wilson 1974, p. 120). A major success of the RG was its ability to correct mean field theory and yield results in close agreement with experiment. In a mean field theory, the order parameter M is defined to be the magnetic moment felt at a lattice site due to the average over all the spins on the lattice. This averaging ignores any large-scale fluctuations that might (and, in fact, are) present in systems near their critical points. The RG corrects this by showing how to incorporate fluctuations at all length scales, from the atomic to the macro, that play a role in determining the macroscopic behavior (specifically the power law dependence—$M \propto |t|^{\beta}$) of the systems near criticality. In fact, near criticality the lattice system will contain "bubbles" (regions of correlated spins—all up or all down) of all sizes from the atomic to the system size. As Kadanoff notes, "[f]rom this picture we conclude that critical phenomena are connected with fluctuations over all length scales between ξ [essentially the system size] and the microscopic distance between particles" (Kadanoff 1976, p. 12).

So away from criticality, below the critical temperature, say, the lattice systems will look pretty much *homogeneous*.[15] For a system with $T \ll T_c$ in figure 7.2 we would have relatively large correlated regions of spins pointing in the same direction. There might be only a few insignificantly small regions where spins are correlated in the opposite direction. This is what is responsible for there being a positive, nonzero, value for M at that temperature. Now suppose we were interested in describing a large system like this away from criticality using the continuum limit as understood by Butterfield above. We would choose a representative elementary volume of radius L around a point $\mathbf{x}$. The volume is small with respect to the system size ξ, but still large enough to contain many spins. Next we would average the quantity $M(N, \mathbf{x}, L)$ over that volume and take the limits $N \to \infty$, $L \to 0$ so as to obtain the proper

[14] See Kadanoff (2000) and Batterman (2002, 2005, 2011) for details.

[15] Systems above the critical temperature will also appear homogeneous as the spins will be uncorrelated, randomly pointing up and down.

continuum value and so that we would be able to model the actually finite collection of spins using convenient continuum mathematics.

But near the critical temperature (near T_c) the system will look heterogeneous—exhibiting a complicated mixture of two distinct phases as in figure 7.3. Now we face a problem. In fact, it is the problem that effectively undermined the mean field approach to critical phenomena. The averaging method employing a representative elementary volume element misses what is most important. For one thing, we will need to know how to weight the different phases as to their import for the macroscopic behavior of the system. In other words, were we to perform the REV averaging, all of the physics of the fluctuations responsible for the coexisting bubbles of up spins and bubbles of down spins would be ignored.

Here is a simple example to see why this methodology will often fail for heterogeneous systems (Torquato 2002, p. 11). Consider a composite material consisting of equal volumes of two materials, one of which is a good electrical conductor and one of which is not. A couple of possible configurations are shown in figure 7.4.

Suppose that the dark, connected phase is the good conductor. If we were to proceed using the REV recipe, then, because the volume fractions are the same, we would grossly underestimate the bulk conductivity of the material in the left configuration and grossly underestimate its bulk insulating capacities in the right configuration. REV averaging treats only the volume fraction and completely misses microstructural details that are relevant to the bulk (macroscale) behavior of the material. In this simple example, the microstructural feature that is relevant is the topological connectedness of the one phase vs. the other—that is, the details about the boundaries between the two phases. Note that the fact that boundaries play an important role serves to undermine the first "theme" of the Straightforward Justification for the use of limits; namely, that taking the limits enable us to remove edge and boundary effects. To the contrary, these can and do play very important roles in determining the bulk behavior of the materials.

One might object that all one needs to do to save the REV methodology would be to properly weight the contribution of the different phases to the overall average. But this is not something that one can do a priori or through ab initio calculations

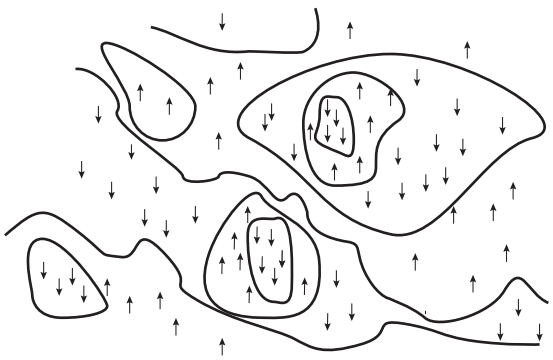


Figure 7.3 Bubbles within bubbles within bubbles …(after Kadanoff 1976, pp. 11–12)

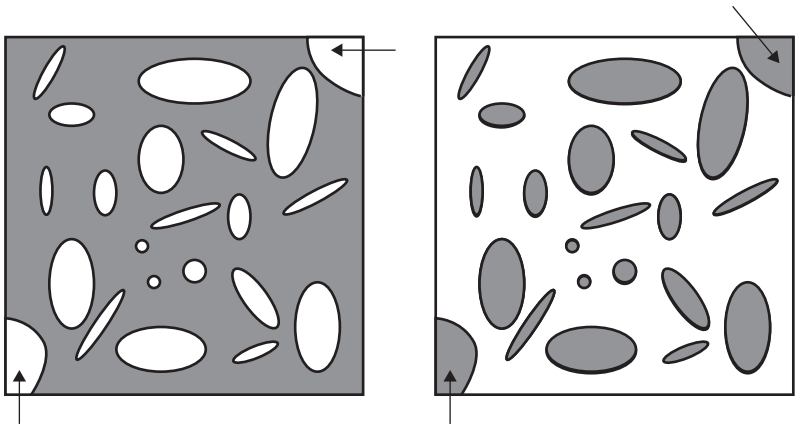


Figure 7.4 50–50 volume mixture

appealing to details and properties of the individual atoms at the atomic scale. Even worse, note the partial blobs at the corners marked by the arrows in figure 7.4. How large are the complete blobs of which they are a part? We do not know because the limited scale of the window (size L of the REV) does not allow us to "see" what is happening at large scales. It is entirely possible (and in the case of critical phenomena actually the case) that these partial blobs will be part of larger connected regions only visible at greater scale lengths. They may be dreaded *invaders from a higher scale*.[16] If such invaders are present, then we have another reason to be wary of limiting REV averaging methods—we will grossly fail to estimate the effective conductivity of the material at macroscales. On the other hand, if we have some nice scaling data about the behavior of material of the sort exploited by the RG, we may well gain enough of a handle on the material's overall behavior to place its conductivity in a firm universality class with other materials that scale in similar ways.

As noted above, in more complicated situations, such as the steel girder with which we began, microstructural features include mesoscale dislocations, defects of various kinds, and martensitic transformations. If we engaged in a purely bottom-up lattice view about steel, paying attention only to the structures for the pure crystal lattice, then we would get completely wrong estimates for its total energy, for its average density, and for its elastic properties. The relevant Hamiltonians require terms that simply do not appear at the smallest scales.[17]

The upshot, then, is that the straightforward justification for the use of infinite limits will miss exactly what is important for understanding what is going on for systems at and near criticality. There, they no longer appear homogeneous across a large range of scales. If we are to try to connect (and thereby extract) correct phenomenological macroscopic values for appropriate parameters (e.g., β) we need to consider

[16] Thanks to Mark Wilson for the colorful terminology!
[17] See (Phillips 2001).

structures that exist at scales greater than the fundamental/basic/atomic. Again, what does this say about the prospects for an overall reductionist understanding of the physics of systems viewed at macroscales?

The RG considers such intermediate scales by including in the calculations the effects of fluctuations or equivalently, the fact that bubbles within bubbles of different phases appear near criticality. *We need methods that tell us how to homogenize heterogeneous materials.* In other words, to extract a continuum phenomenology, we need a methodology that enables us to upscale models of materials that are heterogeneous at small scales to those that are homogeneous at macroscales, as is evidenced by the fact that only a very small number of phenomenological parameters are required to characterize their continuum level behaviors. It appears, then, that the straightforward justification of the use of continuum limits needs to be reconsidered or replaced in those contexts where the materials of interest exhibit heterogeneous microstructures.

In section 5 I will say a bit more about the nature and generality of this different methodology. In the next section, I present a historical discussion, one aim of which is to illustrate that this debate about modeling across scales is not, in the least bit, new. Furthermore, the discussion should give pause to those who think continuum models are ultimately unnecessary. This is the story of deriving appropriate continuum equations for the behavior of elastic solids and gave rise to a controversy that lasted for most of the nineteenth century.

3. Bridging across Scales: A Historical Controversy

Why are the Navier-Stokes equations named after Navier and Stokes? The answer is not as simple as "they both, independently, arrived at the same equation." In fact, there are differences between the equation Navier first came up with and that derived by Stokes. The differences relate to the assumptions that each employed in his derivation, but more importantly, these different assumptions actually led to different equations. Furthermore, the difference between the equations was symptomatic of a controversy that lasted for most of the nineteenth century (de Boer 2000, p. 86).

While the Navier-Stokes equation describes the behavior of a viscous fluid, the controversy has its roots in the derivation of equations for the behavior of an elastic solid. I intend to focus on the latter equations and only at the end make some remarks about the fluid equations.

The controversy concerned the number of material constants that were required to describe the behavior of elastic solids. According to Navier's equation, a single constant marked a material as isotropic elastic. According to Stokes and Green, two constants were required. For anisotropic elastic materials (where symmetries cannot be employed) the debate concerned whether the number of necessary constants

was 15 or 21. This dispute between, respectively, "rari-constancy" theorists and "multi-constancy" theorists depended upon whether one's approach to the elastic solid equations started from a hypothesis to the effect that solids are composed of interacting molecules or from the hypothesis that solids are continuous.

Navier's derivation began from the hypothesis that the deformed state of an elastic body was to be understood in terms of *forces acting between individual particles or molecules that make up the body*. Under this assumption, he derived equations containing only one material constant ϵ.

Navier's equations for an elastic solid are as follows (de Boer 2000, p. 80):

$$\epsilon\left(\triangle u + 2\frac{\partial\Theta}{\partial x}\right) + \rho X = \rho\frac{\partial^2 u}{\partial t^2}, \tag{2}$$

$$\epsilon\left(\triangle v + 2\frac{\partial\Theta}{\partial y}\right) + \rho Y = \rho\frac{\partial^2 v}{\partial t^2}, \tag{3}$$

$$\epsilon\left(\triangle w + 2\frac{\partial\Theta}{\partial z}\right) + \rho Z = \rho\frac{\partial^2 w}{\partial t^2}. \tag{4}$$

Here ϵ, Navier's material constant, reflects the molecular forces that are supposed to turn on when external forces are applied to the body. x, y, z are the coordinates representing the location of a *material point* in the body.[18] u, v, w are the displacement components in the directions x, y, z; X, Y, Z represent the external accelerations (forces) in the directions x, y, z; $\triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ is the Laplace operator; $\Theta = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}$ is the volume strain; and ρ is the material density.

Cauchy also derived an equation for isotropic elastic materials by starting from a molecular hypothesis similar to Navier's. However, his equation contains the correct number of material constants (two). It is instructive to write down Cauchy's equations and to discuss how, essentially, a mistaken, inconsistent derivational move on his part yielded a more accurate set of equations than Navier.

Cauchy's equations for an elastic solid are as follows (de Boer 2000, p. 81) (compare with equation (1)):

$$(R + A)\left(\triangle u + 2\frac{\partial\Theta}{\partial x}\right) + \rho X = \rho\frac{\partial^2 u}{\partial t^2}, \tag{5}$$

$$(R + A)\left(\triangle v + 2\frac{\partial\Theta}{\partial y}\right) + \rho Y = \rho\frac{\partial^2 v}{\partial t^2}, \tag{6}$$

$$(R + A)\left(\triangle w + 2\frac{\partial\Theta}{\partial z}\right) + \rho Z = \rho\frac{\partial^2 w}{\partial t^2}. \tag{7}$$

[18] Note that in continuum mechanics, generally, a material point or "material particle" is not an atom or molecule of the system; rather it is an imaginary region that is large enough to contain many atomic subscales (whether or not they really exist) and small enough relative to the scale of field variables characterizing the impressed forces. Of course, as noted, Navier's derivation did make reference to atoms.

R, A are the two material constants. Cauchy noted, explicitly, that when $A = 0$ his equations agree with Navier's when $R = \epsilon$[19] (de Boer 2000, p. 81). How did Cauchy arrive at a different equation than Navier, despite starting, essentially, from the same molecular assumptions about forces? He did so by assuming that, despite the fact that he is operating under the molecular hypothesis, he can, in his derivation replace certain summations by integrations. In effect, he actually employs a *continuum condition* contradictory to his fundamental starting assumption.[20]

George Green, in 1839, published a study that arrived at the correct equations—essentially (5)–(7)—by completely eschewing the molecular hypothesis. He treated the entire body as composed of "two indefinitely extended media, the surface of junction when in equilibrium being a plane of infinite extent."[21] He also assumed that the material was not crystalline and, hence, isotropic. Then using a principle of the conservation of energy/work he derived, using variational principles of Lagrangian mechanics, his multi-constant equation.

Finally, following the discussion of Todhunter and Pearson (1960), we note that Stokes's work supported the multi-constancy theory in that he was able to generalize his equations for the behavior of *viscous fluids* to the case of elastic solids by making no distinction between a viscous fluid and a solid undergoing permanent—plastic—deformation. "He in fact draws no line between a plastic solid and a viscous fluid. The formulae for the equilibrium of an isotropic plastic solid would thus be bi-constant" (Todhunter and Pearson 1960, p. 500). This unification of continuum equations lends further support to the multi-constancy theory.

The historical debate represents just the tip of the iceberg of the complexity surrounding both theoretical and experimental work on the behavior of the supposedly simpler, isotropic, cases of elastic solids. Nevertheless, the multi-constancy theory wins the day for appropriate classes of structures. And, derivations that start from atomic assumptions fail to arrive at the correct theory. It seems that here may very well be a case where a continuum point of view is actually superior: bottom-up derivation from atomistic hypotheses about the nature of elastic solid bodies fails to yield correct equations governing the macroscopic behavior of those bodies. There are good reasons, already well understood by Green and Stokes, for eschewing such reductionist strategies.

This controversy is important for the current project for the following reason. Green and Stokes were moved by the apparent scaling or homogeneity observed in elastic solids and fluids. That is, as one zooms in with reasonable powerful microscopes one sees the steel to be the same at different magnifications; likewise for the fluid. Green and Stokes then extrapolated this scale invariance to hold at even larger magnifications—at even smaller scales. We now know (and likely they suspected) that this extrapolation is not valid beyond certain scale lengths—the atomistic nature of the materials will begin to show itself. Nevertheless, the

[19] I have fixed a typographical error in these equations.
[20] See Todhunter and Pearson (1960, pp. 224 and 235–27) for details. Note also how this limiting assumption yields *different and correct* results in comparison to the finite atomistic hypotheses.
[21] Cited in Todhunter and Pearson (1960, p. 495).

continuum modeling was dramatically successful in that it predicted the correct number and the correct character of the phenomenological constants.

De Boer reflects on the reasons for why this controversy lasted so long and was so heated:

> Why was so much time spent on molecular theory considerations, in particular, by the most outstanding mechanics specialists and mathematicians of the epoch? One of the reasons must have been the temptation of gaining the constitutive relation for isotropic and anisotropic elastic continua directly from pure mathematical studies and simple mechanical principles;[22] It was only later realized that Hooke's *generalized law* is an assumption, and that the foundation of the linear relation had to be supported by experiments. (2000, pp. 86–87)

The upshot of this discussion is reflected in de Boer's emphasis that the constitutive equations or special force laws (Hooke's law) are dependent, for their very form, on experimental results. So a simple dismissal of continuum theories as "in principle" eliminable, as reducible, and merely pragmatically justified, is mistaken. Of course, the phenomenological parameters, like Young's modulus (related to Navier's ϵ), must encode details about the actual atomistic structure of elastic solids. But it is naive, indeed, to think that one can, in any straightforward way derive or deduce from atomic facts what are the phenomenological parameters required for continuum model of a given material. It is probably even more naive to think that one will be able to derive or deduce from those atomic facts what are the actual values for those parameters for a given material.

This historical discussion and the intense nineteenth-century debate between the rari- and multi-constancy theorists apparently supports the view that there is some kind of fundamental incompatibility between small scale and continuum modeling practices. That is, it lends support to the stark choice one must apparently make between bottom-up and top-down modeling suggested by the tyranny of scales.

A modern, more nuanced, and better informed view challenges this consequence of the tyranny of scales and will be discussed in section 5. However, such a view will not, in my opinion, bring much comfort to those who believe the use of continuum models or idealizations is only pragmatically justified. A modern statement supporting this point of view can be found in (Phillips 2001):

> [M]any material properties depend upon more than just the identity of the particular atomic constituents that make up the material.…[M]icrostructural features such as point defects, dislocations, and grain boundaries can each alter the measured macroscopic "properties" of a material. (pp. 5–8)

[22] This is the temptation promised by an ultimate reductionist point of view.

It is important to reiterate that, contrary to typical philosophical usage, "microstructural features" here is not synonymous with "atomic features"! Defects, dislocations, etc. exist at higher scales.

In the next section I will further develop the stark dichotomy between bottom-up modeling and top-down modeling as a general philosophical problem arising between different recipes for applying mathematics to systems exhibiting different properties across a wide range of scales.

4. Euler's Recipes: Discrete and Continuum

4.1 Discrete

Applied mathematical modeling begins with an attempt to write down an equation governing the system exhibiting the phenomenon of interest. In many situations, this aim is accomplished by starting with a general dynamical principle such as Newton's second law: $\mathbf{F} = m\mathbf{a}$. Unfortunately, this general principle tells us absolutely nothing about the material or body being investigated and, by itself, provides no model of the behavior of the system. Further data are required and these are supplied by so-called "special force laws" or "constitutive equations."

A recipe, due to Leonhard Euler, for finding an appropriate model for a system of particles proceeds as follows (Wilson 1974):

1. Given the class of material (point particles, say), determine the kinds of special forces that act between them. Massive particles obey the constitutive gravitational force: $\mathbf{F}_G = G\frac{m_i m_j}{r_{ij}^2}$. Charged particles additionally will obey the Coulomb force law: $\mathbf{F}_E = k_e\frac{q_i q_j}{r_{ij}^2}$.
2. Choose Cartesian coordinates along which one decomposes the special forces.
3. Sum the forces acting on each particle along the appropriate axis.
4. Set the sum for each particle i equal to $m_i\frac{d^2x}{dt^2}$ to yield the total force on the particle.

This yields a differential equation that we then employ (= try to solve) to further understand the behavior of our point particle system. Only rarely (for very few particles or for special symmetries) will this equation succumb to analytical evaluation. In many instances, further simplification employing mathematical strategies of variable reduction, averaging, etc. enable us to gain information about the behavior of interest.

4.2 Continuum

As we saw in section 3, Cauchy had a role in the derivation of equations for elastic solids. We note again that he was lucky to have arrived at the correct equations, given

that he started with a bottom-up derivation in mind. Nevertheless, Cauchy was an important figure in the development of continuum mechanics: it turns out that at macroscales, forces within a continuum can be represented by a single second-rank tensor, despite all of the details that appear at the atomic level. This is known as the Cauchy stress tensor (Philips 2001, p. 39). The analog of Newton's second law, for continua is the principle of balance of linear momentum. It is a statement that "the time rate of change of the linear momentum is equal to the net force acting on [a] region Ω":[23]

$$\frac{D}{Dt}\int_{\Omega}\rho\mathbf{v}dV=\int_{\partial\Omega}\mathbf{t}dA+\int_{\Omega}\mathbf{f}dV. \tag{8}$$

Here $\partial\Omega$ is the boundary of the region Ω, $\mathbf{t}$ is the traction vector representing surface forces (squeezings, for instance), and $\mathbf{f}$ represents the body forces such as gravity. The left-hand side of equation (8) is the time rate of change of linear momentum. The material time derivative, D/Dt, is required because in addition to explicit time dependence of the field, we need to consider the fact that the material itself can move into a region where the field is different.

As with Euler's discrete recipe, equation (8) requires input from constitutive equations to apply to any real system.

> Whether our interest is in the description of injecting polymers into molds, the evolution of Jupiter's red spot, the development of texture in a crystal, or the formation of vortices in wakes, we must supplement the governing equations of continuum mechanics with some constitutive description. (Phillips 2001, p. 51)

For the case of a steel girder, considered in the regime for use in constructing bridges or buildings we need the input that it obeys something like Hooke's law—that its stress is linearly related to its strain. In modern terminology, we need to provide data about the Cauchy stress tensor. For isotropic linear elastic solids, symmetry considerations come into play and we end up with equation (1)—the Navier-Cauchy equation that characterizes the equilibrium states of such solids:

$$(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u})+\rho\nabla^2\mathbf{u}+\mathbf{f}=0.$$

The "Lamé" parameters (related to Young's modulus) express the empirical details about the material response to stress and strain.

4.3 Controversy

A question of pressing concern is why the continuum recipe should work at all. We have seen in the historical example that it does, and in fact, we have seen that were we

[23] See Phillips (2001, pp. 41–42).

simply to employ the discrete (point particle) recipe, we would not arrive at the correct results. In asking why the continuum recipe works on the macroscale, we are asking about the relationship between the dynamical models that track the behavior of individual atoms and molecules and equations like those of Navier, Stokes, Cauchy, and Green that are applicable at the scale of millimeters. Put slightly differently, we would like an account of why it is safe to use the Cauchy momentum equation in the sense that it yields correct equations with the appropriate (few) parameters for broadly different classes of systems—from elastic solids to viscous fluids.

From the point of view of Euler's continuum recipe, one derives the equations for elastic solids, or the Navier-Stokes equations, independently of any views about the molecular or atomic makeup of the medium. (In the nineteenth century the question of whether matter was atomistic had yet to be settled.)

To ask for an account of why it is safe to use the continuum recipe for constructing macroscale models is to ask for an account of the robustness of that methodology. The key physical fact is that the bulk behaviors of solids and fluids are almost completely insensitive to the actual nature of the physics at the smallest scale. The "almost" here is crucial. The atomic details that we do not know (and, hence, do not explicitly refer to) when we employ continuum recipe are encoded in the small number of phenomenological parameters that appear in the resulting equations—Young's modulus, the viscosity, etc. So the answer to the safety question will involve showing how to determine the "fixed points" characterizing broad classes of macroscopic materials—fixed points that are characterized by those phenomenological parameters. Recall the statement by Nelson cited above in section 2.1. In the context of critical phenomena and the determination of the critical exponent β, this upscaling or connection between the Euler's discrete and continuum recipes is accomplished by the renormalization group. In that context, the idea of a critical point and related singularities plays an important role. But Nelson's suggestion is that upscaling of this sort should be possible even for classes of systems without critical points. For example, we would like to understand *why Young's modulus is the appropriate phenomenological parameter for classifying solids as linear elastic, despite rather severe differences in the atomic structure of members of that class.* Finding answers to questions of this latter type is the purview of so-called "homogenization" theory, of which one can profitably think the RG to be a special case.

In the next section, I will spend a bit more time on the RG explanation of the universality of critical behavior, filling in some gaps in the discussion in section 2.1. And, I will try to say something about general methodology of upscaling through the use of homogenization limits.

5. A Modern Resolution

To begin, consider a problem for a corporation that owns a lot of casinos. The CEO of the corporation needs to report to the board of trustees (or whomever) on the

expected profits for the corporation. How is she to do it? Assuming (contrary to fact) that casino gaming is fair, she would present to the board a Gaussian or normal probability distribution showing the probabilities of various profits and losses, with standard deviations that would allow for statistical predictions as to expected profits and losses. She may also seek information as to how to manipulate the mean and variance so as to guarantee the likelihood of greater profits for less risk, etc. The Gaussian distribution is a function characterized by two parameters—the mean μ and the variance σ^2. Where will the CEO get the values for the mean and variance? Most likely by *empirically investigating* the actual means and variances displayed over the past year by the various casinos in the corporation. Consider figure 7.5. Should the CEO look to the individual gambles or even to collections of individual gambles of different types in particular casinos? A bottom-up reductionist would say that all of the details about the corporation as a whole are to be found by considering these details. But, in fact, (i) she should not focus too much on spatiotemporal local features of a single casino: suppose someone hits the jackpot on a slot machine. Likely, many people will run to that part of the casino, diminishing profits from the roulette wheels and blackjack tables, and skewing the prediction of the actual mean and variance she is after. Nor (ii) would it be wise to focus too much on groups of casinos say in a particular geographic area (such as Las Vegas) over casinos owned in another area (such as Atlantic City). After all, different tax structures in these different states and municipalities play an important role as well. Such intermediate structures and environmental considerations are crucial—consider again the bubbles within bubbles structures that characterize the heterogeneities at lower scales in the case of the universality of critical phenomena. The CEO needs to look at large groups of collections of casinos where there is evident scaling and self-similarity. Apparent scaling behavior and self-similarity at large scales is an indication of homogeneity. Thus, as with our steel girder, empirical data (at large scales) is required to determine the values of the relevant parameters.

Now why should she think that these two parameters—properties of collections of casinos offering different and varied kinds of games (roulette, poker, blackjack, slots, etc.)—are the correct ones with which to make the presentation to the board? Equivalently, *why she should employ a* Gaussian *probability distribution (it is uniquely defined by the mean and variance) in the first place, as opposed to some other probability distribution*? The answer is effectively provided by an RG argument analogous to that which allows us to determine the functional form of the order parameter M near criticality—that it scales as $|t|^{\beta}$ near criticality. It is an argument that leads us to expect behavior in accord with the central limit theorem. There are deep similarities between the arguments for why the functional form with exponent β is universal and why Gaussian or central limiting behavior is so ubiquitous. In the former case, the RG demonstrates that various systems all flow to a single fixed point in an abstract space of Hamiltonians or coupling constants. That fixed point determines the universality class that is characterized by the scaling exponent β. Similarly, the Gaussian probability distribution is a fixed point for a wide class of probability distributions under a similar renormalization group transformation. (For details see Batterman

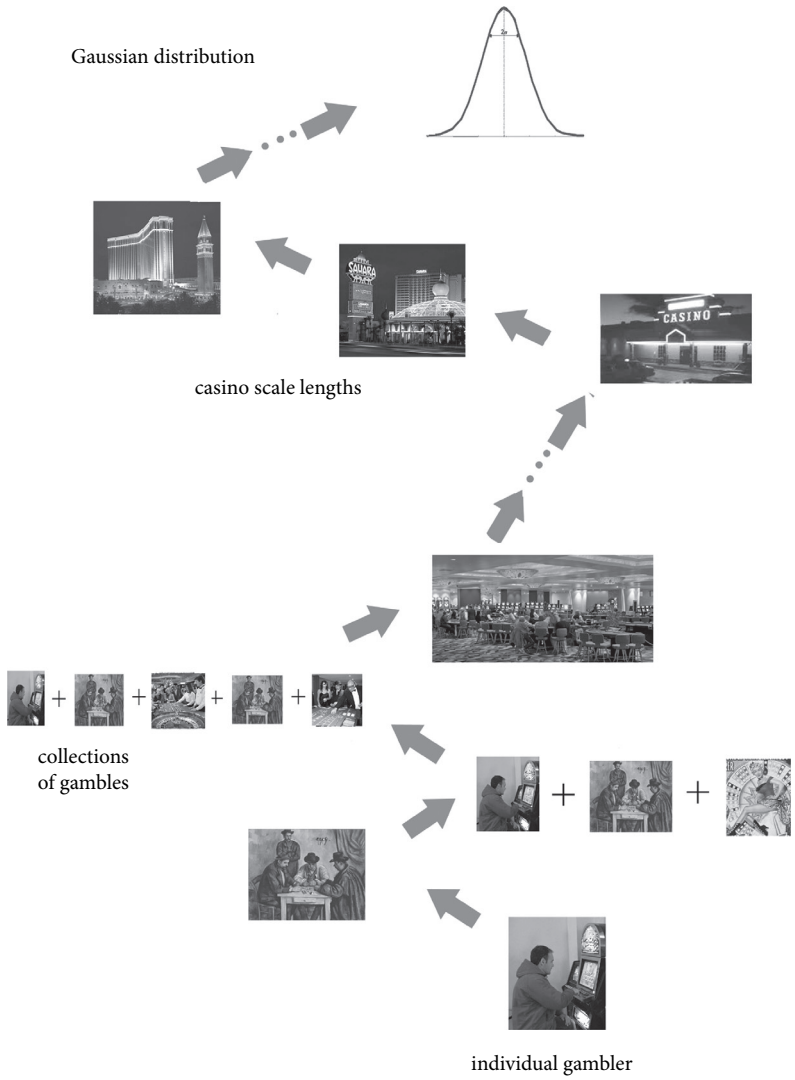


Figure 7.5 Gambles within gambles within gambles …

2010 and Sinai 1992.) Thus, the answer to why the mean μ and the variance σ^2 are the relevant parameters depends upon an RG, limiting argument. Generalizing, one should expect related argument strategies to tell us why the two elastic "constants" (related to Young's modulus) are the correct parameters with which to characterize the universality class of elastic solids. The appeal to something like central limiting behavior is characteristic of homogenization theory and distinguishes this line of argumentation from that employing REV averaging techniques.

In fact, the difference between averaging and homogenization is related to the difference between the law of large numbers and the central limit theorem: averaging or first order perturbation theory "can often be thought of as a form (or consequence) of the law of large numbers." Homogenization or second order

perturbation theory "can often be thought of as a form (or consequence) of the central limit theorem" (Pavliotis and Stuart 2008, pp. 6–7).

Here is a brief discussion that serves to motivate these connections. Consider a sum function of independent and identically distributed random variables, Y_i: $S(n) = \sum_{i=1}^{n} Y_i$. The sample average $\overline{S(n)} = 1/n \sum_{i=1}^{n} Y_i$ converges to the mean or expected value μ. The strong law of large numbers asserts that

$$Pr\left(|\overline{S(n)} - \mu| > \epsilon\right) = 0.$$

As such it tells us about the first moment of the random variable ($\overline{S(n)}$—the average. The central limit theorem by contrast tells us about the second moment of the normalized sum ($\overline{S(n)}$; that is it tells us about the behavior of fluctuations about the average μ. It says that for $n \to \infty$ the probability distribution of $\sqrt{n}(\overline{S(n)} - \mu)$ converges to the normal or Gaussian distribution $\mathcal{N}(0, \sigma^2)$, with mean 0 and variance $\frac{\sigma^2}{n}$ where σ is the standard deviation of the Y_i's.[24]

Thus again we see that in the probabilistic scenario, as in the case of critical phenomena, we must to pay attention to the fact that collections of gambles (bubbles) contribute to the behavior of the system at the macroscale. Once again, we need to pay attention to fluctuations about some average behavior, and not just the average behavior itself.

Furthermore, a similar picture is possible regarding the upscaling of our modeling of the behavior of the steel girder with which we started. Compare the two cases, figure 7.6, noting that here too only a small number of phenomenological parameters are needed to model the continuum/macroscale behavior. (E is Young's modulus and I is the area moment of inertia of a cross-section of the girder.)

The general problem of justifying the use of Euler's continuum recipe to determine the macroscopic equation models involves connecting a statistical/discrete theory in terms of atoms or lattice sites to a hydrodynamic or continuum theory. Much effort has been spent on this problem by applied mathematicians and materials scientists. And, as I mentioned above, the RG argument that effectively determines the continuum behavior of systems near criticality is a relatively simple example of this general homogenization program.

In hydrodynamics, for example Navier-Stokes theory, there appear density functions, $\rho(\mathbf{x})$, that are defined over a continuous variable $\mathbf{x}$. These functions exhibit no atomic structure at all. On the other hand, for a statistical theory, such as the Ising model of a ferromagnet, we have seen that one defines an order parameter (a magnetic density function) $M(\mathbf{x})$ that is the average magnetization in a volume surrounding $\mathbf{x}$ that contains many lattice sites or atoms. The radius of the volume, L, is intermediate between the lattice constant (or atomic spacing) and the correlation

[24] Proofs of the central limit theorem that involve moment generating functions $M(t)$ for the component random variables Y_i make explicit that there is an asymptotic expansion in a small parameter t, where truncation of the series at first order gives the mean, and truncation of the series at second order gives the fluctuation term. Hence the connection between these limit theorems and first and second order perturbation theory. In fact, two limits are involved: the limit as the small parameter $t \to 0$ and the limit $n \to \infty$.

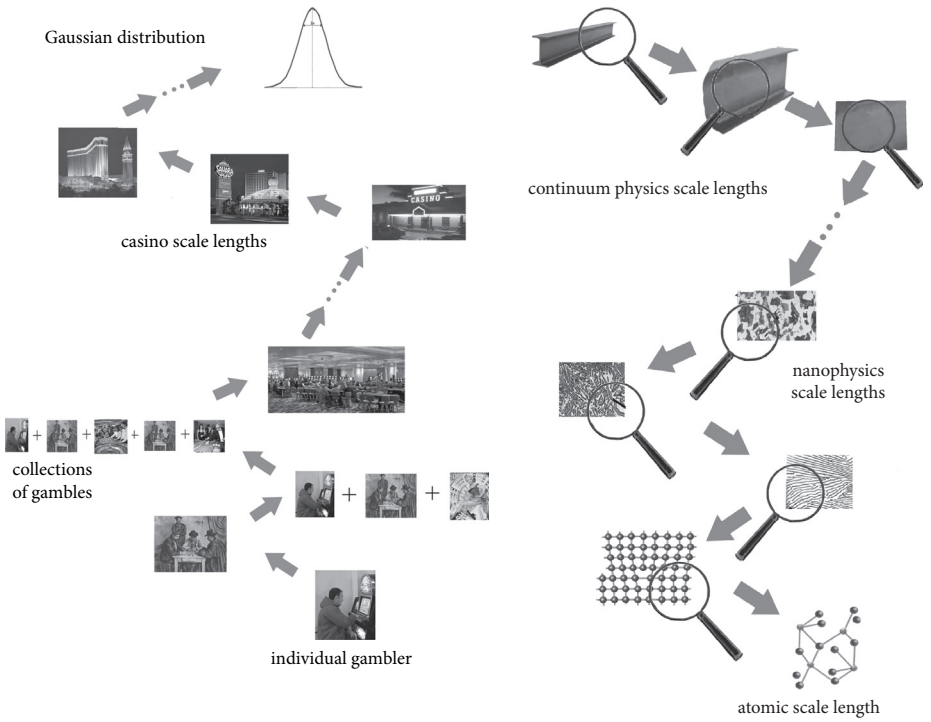


Figure 7.6 Gaussian and steel—few (macro) parameters: $[\mu, \sigma^2]$; $[E, I]$

length ξ: $(a \ll L \ll \xi)$. As noted in section 2.1 this makes the order parameter depend upon the length L (Wilson 1974, p. 123).

A crucial difference between the hydrodynamic (thermodynamic) theory and the statistical theory is that the free energy in the former is determined using the single magnetization function $M(\mathbf{x})$. In statistical mechanics, on the other hand, the free energy is "a weighted average over all possible forms of the magnetization $M(\mathbf{x})$." (Wilson 1974, p. 123) This latter set of functions is parameterized by the volume radius L. On the statistical theory due originally to Landau, the free energy defined as a function of $M(\mathbf{x})$ takes the following form:

$$F = \int \left([\nabla M(\mathbf{x})]^2 + RM^2(\mathbf{x}) + UM^4(\mathbf{x}) - B(M(\mathbf{x})) \right) d^3(\mathbf{x}), \tag{9}$$

where R and U are (temperature dependent) constants and B is a (possibly absent) external magnetic field. (Wilson 1974, p. 122) This (mean field) theory predicts the wrong value, 1/2, for β–the critical exponent. The problem, as diagnosed by Wilson, is that while the Landau theory can accommodate fluctuations for lengths $\lambda < L$ in its definition of M as an average, it cannot accommodate fluctuations of lengths L or greater.

> A sure sign of trouble in the Landau theory would be the dependence of the constants R and U on L. That is, suppose one sets up a procedure for calculating R and U

> which involves statistically averaging over fluctuations with wavelengths $\lambda < L$. If one finds R and U depending on L, this is proof that long-wavelength fluctuations are important and Landau's theory must be modified. (p. 123)

The RG account enables one to exploit this L-dependence and eventually derive differential equations (RG) for R and U as functions of L that allow for the calculation of the exponent β in agreement with experiment. The key is to calculate and compare the free energy for different averaging sizes L and $L+\delta L$. One can proceed as follows[25]: Divide $M(x)$ in the volume element into two parts:

$$M(\mathbf{x}) = M_H(\mathbf{x}) + mM_{fl}(\mathbf{x}). \tag{10}$$

M_H is a hydrodynamic part with wavelengths of order ξ and M_{fl} is a fluctuating part with wavelength between L and $L+\delta L$. The former will be effectively constant over the volume.

By performing a single integral over m—the scale factor in (10)—we get an iterative expression for the free energy for the averaging size $L+\delta L$, $F_{L+\delta L}$, in terms of the free energy for the averaging size L:

$$e^{-F_{L+\delta L}} = \int_{-\infty}^{\infty} e^{-F_L}\, dm. \tag{11}$$

In effect, one finds a step by step way to include all the fluctuations—all the physics—that play a role near criticality. One moves from a statistical theory defined over finite N and dependent on L to a hydrodynamic theory of the continuum behavior at criticality. "Including all of the physics" means that the geometric structure of the bubbles within bubbles picture gets preserved and exploited as one upscales from the finite discrete atomistic account to the continuum model at the scale of ξ—the size of the system. That is exactly the structure that is wiped out by the standard REV averaging, and it is for that reason that Landau's mean field theory failed.

5.1 Homogenization

Continuum modeling is concerned with the effective properties of materials that, in many instances, are microstructurally heterogeneous. These microstructures, as noted, are not always to be identified with atomic or lowest scale "fundamental" properties of materials. Simple REV averaging techniques often assume something like that, but in general the effective, phenomenological properties of materials are not simple mixtures of volume fractions of different composite phases or materials. Many times the microstructural features are geometric or topological including (in addition to volume fractions) "surface areas of interfaces, orientations, sizes, shapes, spatial distributions of the phase domains; connectivity of the phases; etc."

[25] Details in Wilson (1974, pp. 125–27).

(Torquato 2002, p. 12). In trying to bridge the scales between the atomic domain and that of the macroscale, one needs to connect rapidly varying local functions of the different phases to differential equations characterizing the system at much larger scales. Homogenization theory accomplishes this by taking limits in which the local length (small length scale) of the heterogeneities approaches zero in a way that preserves (and incorporates) the topological and geometric features of the microstructures.

Most simply, and abstractly, homogenization theory considers systems at two scales: ξ, a macroscopic scale characterizing the system size, and a microscopic scale, a, associated with the microscale heterogeneities. There may also be applied external fields that operate at yet a third scale Λ. If the microscale, a, is comparable with either ξ or Λ, then the modeler is stuck trying to solve equations at that smallest scale. However, as is often the case, if $a \ll \Lambda \ll \xi$, then one can introduce a parameter

$$\epsilon = \frac{a}{\xi}$$

that is associated with the fluctuations at the microscale of the heterogeneities—the local properties (Torquato 2002, pp. 305–6). In effect, then one looks at a family of functions u_ϵ and searches for a limit $u = \lim_{\epsilon \to 0} u_\epsilon$ that tells us what the effective properties of the material will be at the macroscale.

Figure 7.7 illustrates this. The left box shows the two scales a and ξ with two phases of the material K_1 and K_2. The homogenization limit enables one to treat the heterogeneous system at scale a as a homogeneous system at scale ξ with an effective material property represented by K_e. For an elastic solid like the steel girder, K_e would be the effective stiffness tensor and is related experimentally to Young's modulus. For a conductor, K_e would be the effective conductivity tensor that is related experimentally to the parameter σ—the specific conductance—appearing

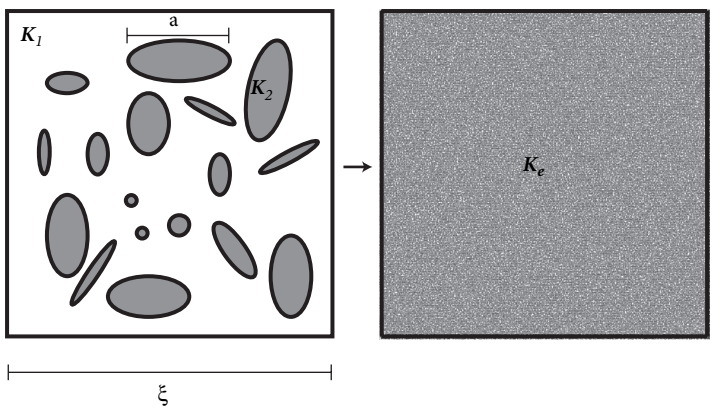


Figure 7.7 Homogenization limit (after Torquato 2002, pp. 2, 305–6)

in Ohm's law:

$$\mathbf{J} = \sigma \mathbf{E},$$

where **J** is the current density at a given location **x** in the material and **E** is the electric field at **x**. At the risk of being overly repetitive, *note that in these and other cases, it is unlikely that the effective material property K_e will be a simple average.*

Let me end this brief discussion of homogenization by highlighting what I take to be a very important concept for the general problem of upscaling. This is the concept of an order parameter and related functions. The notion of an order parameter was introduced in our discussion of continuous phase transitions in thermodynamics, and the statistical mechanical explanations of certain of their features. In effect, the order parameter is a microstructure (mesoscopic scale) dependent function introduced to codify the phenomenologically observed transition between different states of matter. As we have seen, the magnetization M represented in figure 7.2 is introduced to reflect the fact that at the Curie temperature the systems goes from an unordered phase, above T_c to an ordered phase, below T_c. In this context, the divergences and nonanalyticities at the critical point play an essential role in determining the fixed point that characterizes the class of systems exhibiting the same scaling behavior: $M \propto |t|^{\beta}$. But, again following Nelson's suggestion, entire classes of systems such as the class of linear elastic solids are also characterized by "fixed points" represented by a relatively few phenomenological parameters like Young's modulus.

It is useful to introduce an order-like parameter in this more general context of upscaling where criticality is not really an issue. For example, consider the left image in figure 7.7. In upscaling to get to the right image, one can begin by defining indicator or characteristic functions for the different phases as a function of spatial location (Torquato 2002, pp. 24–5). For instance, if the shaded phase occupies a region U_s in the space, then an indicator function of that phase is given by

$$\chi^s(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{x} \in U_s \\ 0, & \text{otherwise.} \end{cases}$$

One can also introduce indicator functions for the interfaces or boundaries between the two phases.[26] Much information can then be determined by investigating n-point probability functions expressing the probabilities that n locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are to be found in regions occupied by the shaded phase.[27]

$$S_n^s(\mathbf{x}_1, \ldots, \mathbf{x}_n) = Pr\left\{\chi^s(\mathbf{x}_1) = 1, \ldots, \chi^s(\mathbf{x}_n = 1)\right\}.$$

In this way many features, other than simple volume fraction, that exist at microscales can be represented and employed in determining the homogenization limit for complex heterogeneous systems. The introduction of such field variables, correlation functions, etc., allow us to characterize the heterogeneous structures

[26] These will be generalized distribution functions.
[27] See Torquato (2002) for a detailed development of this approach.

above the atomic scales. In some cases, such as the bubbles within bubbles structure of the different phases at a continuum phase transition, much of this additional apparatus will not be necessary. (Though, of course, it is essential to take into consideration that structure in that particular case.) But for many more involved upscaling problems such as steel, the additional mathematical apparatus will be critical in determining the appropriate effective phenomenological theory at the continuum level. As we have seen these microstructures are critical for an understanding of how the phenomenological parameters at the continuum scale emerge.

The main lesson to take from this all-too-brief discussion is that physics at these micro/meso-scopic scales need to be considered. Bottom-up modeling of systems that exist across a large range of scales is not sufficient to yield observed properties of those systems at higher scales. Neither is complete top-down modeling. After all, we know that the parameters appearing in continuum models must depend upon details at lower scale levels. The interplay between the two strategies—a kind of mutual adjustment in which lower scale physics informs upper scale models and upper scale physics corrects lower scale models—is complex, fascinating, and unavoidable.

6. Conclusion

The solution to the tyranny of scales problem has been presented as one of seeing if it is possible to exploit microstructural scale information (intermediate between atomic scales and macroscopic scales) to bridge between two dominant and apparently incompatible modeling strategies. These are the traditional bottom-up strategies associated with a broadly reductionist account of science and pure top-down strategies that held sway in the nineteenth century and motivated the likes of Mach, Duhem, Maxwell, and others. Despite great progress in understanding the physics of atomic and subatomic particles, the persistence of continuum modeling has led to heated debates in philosophy about emergence, reduction, realism, etc. We have canvassed several different attitudes to the apparent ineliminability of continuum level modeling in physics. On the one hand, there is the view of Butterfield and others, that the use of continuum limits represents nothing more than a preference for the mathematical convenience of the infinite. Another possible view, coming out of the tyranny of scales, suggests a kind of skepticism: we need both atomic scale models and continuum scale models that essentially employ infinite idealizations. However, a unified account of applied mathematics that incorporates both the literally correct atomic models and the essentially idealized continuum models seems to be beyond our reach.[28]

[28] See Maddy (2008) for a good discussion of this point of view among other interesting topics about the applicability of mathematics.

I claim that neither of these attitudes is ultimately acceptable. Butterfield et al. are wrong to believe that continuum models are simply mathematical conveniences posing no real philosophical concerns. This position fails to respect some rather deep differences between kinds of continuum modeling. In particular, the strategies employed in the renormalization group and in homogenization theory differ significantly from those employed in standard representative elementary volume (REV) averaging scenarios. The significance of Wilson's renormalization group advance was exactly to point out why such REV methods fail and how that failure can be overcome. The answer, as we have seen, is to pay attention to "between" scale structures as in the case of the bubbles within bubbles picture of what happens at phase transitions. Incorporating such structures—features that cannot be understood as averages over atomic level structures—is exactly the strategy behind upscaling attempts that connect Euler-type discrete modeling recipes to Euler-type continuum recipes. Homogenization lets us give an answer to why the use of the continuum recipe is safe and robust. It provides a satisfactory justification for the use of such continuum models, but not one that is "straightforward" or pragmatically motivated. As such, homogenization provides the beginning of an account of applied mathematics that unifies the radically different scale-dependent modeling strategies.

I have also tried here to focus attention on a rather large subfield of applied mathematics that should be of interest to philosophers working on specific issues of modeling, simulation, numerical methods, and idealizations. In addition, understanding the nature of materials in terms of homogenization strategies can inform certain questions about the nature of physical properties and issues about realism. For instance, we have seen that many materials at macroscales are characterized by a few phenomenological parameters such as the elastic constants. Understanding the nature of materials requires understanding why these constants and not others are appropriate, as well as understanding from where the constants arise. One important lesson is that many of these material defining parameters are not simply dependent upon the nature of the atoms that compose the material. There is a crucial link between structure at intermediate scales and observed properties at the macroscale.

It may do to end with an nice statement (partially cited earlier) from Rob Phillips's excellent book *Crystals, Defects, and Microstructures* (2001) expressing this point of view.

> Despite the power of the idea of a material parameter, it must be greeted with caution. For many features of materials, certain "properties" are not *intrinsic*. For example, both the yield strength and fracture toughness of a material depend upon its internal constitution. That is, the measured material response can depend upon microstructural features such as the grain size, the porosity, etc. Depending upon the extent to which the material has been subjected to prior working and annealing, these properties can vary considerably. Even a seemingly

> elementary property such as the density can depend significantly upon that material's life history. The significance of the types of observations given above is the realization than many material properties depend upon more than just the identity of the particular atomic constituents that make up that material.... [M]icrostructural features such as point defects, dislocations, and grain boundaries can each alter the measured macroscopic "properties" of a material. (pp. 5–8)

Philosophers who insist that bottom-up explanations of the macroscopic properties of materials are desirable to the exclusion of top-down modeling considerations are, I think being naive, similar to those who maintain that top-down continuum type modeling strategies are superior. The tyranny of scales appears to force us to choose between these strategies. However, new work on understanding the problem of upscaling or modeling across scales suggests that both types of strategies are required. Our top-down considerations will inform the construction of models at lower scales. And our bottom-up attempts will likewise induce changes and improvements in the construction of higher scale models. Mesoscopic structures cannot be ignored and, in fact, provide the bridges that allow us to model across scales.

add space after comma

References

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

Robert. W. Batterman. Theories between theories: Asymptotic limiting intertheoretic relations. *Synthese*, 103:171–201, 1995.

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

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

Robert W. Batterman. Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Modern Physics*, 36:225–244, 2005a.

Robert W. Batterman. Response to Belot's "Whose devil? Which details?". *Philosophy of Science*, 72(1):154–163, 2005b.

Robert W. Batterman. *Encyclopedia of Philosophy*, chapter Reduction. Macmillan Reference, Detroit, 2nd edition, 2006.

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.

Robert W. Batterman. Reduction and renormalization. In Gerhard Ernst and Andreas Hüttemann, editors, *Time, Chance, and Reduction: Philosophical Aspects of Statistical Mechanics*. Cambridge University Press, Cambridge, 2010.

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

Gordon Belot. Whose devil? which details? *Philosophy of Science*, 71(1):128–153, 2005.
Jeremy Butterfield. Less is different: Emergence and reduction reconciled. *Foundations of Physics*, 41(6):1065–1135, 2011a.
Jeremy Butterfield and Nazim Bouatta. Emergence and reduction combined in phase transitions. http://philsci-archive.pitt.edu/id/eprint/8554, 2011b.
Reint de Boer. *Theory of Porous Media: Highlights in Historical Development and Current State*. Springer, Berlin, 2000.
R. Feynman, R. Leighton, and M. Sands. *The Feynman Lectures on Physics*, volume 2. Addison-Wesley, Reading, Massachusetts, 1964.
Ulrich Hornung, editor. *Homogenization and Porous Media*, volume 6 of *Interdisciplinary Applied Mathematics*. Springer, New York, 1997.
Leo P. Kadanoff. Scaling, universality, and operator algebras. In C. Domb and M. S. Green, editors, *Phase Transitions and Critical Phenomena*, volume 5A. Academic Press, San Diego, 1976.
Leo P. Kadanoff. *Statistical Physics: Statics, Dynamics, and Renormalization*. World Scientific, Singapore, 2000.
Penelope Maddy. How applied mathematics became pure. *The Review of Symbolic Logic*, 1(1):16–41, 2008.
Tarun Menon and Craig Callender. Turn and face the strange . . . Ch-Ch-Changes: Philosophical questions raised by phase transitions, This Volume 2012.
Ernest Nagel. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Harcourt, Brace, & World, New York, 1961.
David R. Nelson. *Defects and Geometry in Condensed Matter Physics*. Cambridge University Press, Cambridge, 2002.
John Norton. Approximation and idealization: Why the difference matters. http://philsci-archive.pitt.edu/8622/, 2011.
J. T. Oden (Chair). Simulation based engineering science—an NSF Blue Ribbon Report. www.nsf.gov/pubs/reports/sbes_final_report.pdf, 2006.
Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale Methods: Averaging and Homogenization*. Texts in Applied Mathematics. Springer, New York, 2008.
Rob Phillips. *Crystals, Defects, and Microstructures: Modeling across Scales*. Cambridge University Press, Cambridge, 2001.
Ya G. Sinai. *Probability Theory: An Introductory Course*. Springer-Verlag, Berlin, 1992. Trans.: D. Haughton.
Lawrence Sklar. *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge University Press, Cambridge, 1993.
Isaac Todhunter and Karl Pearson (Ed.). *A History of the Theory of Elasticity and of the Strength of Materials from Galilei to Lord Kelvin*, volume 1: Galilei to Saint-Venant 1639–1850. Dover, New York, 1960.
Salvatore Torquato. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. Springer, New York, 2002.
Kenneth G. Wilson. Critical phenomena in 3.99 dimensions. *Physica*, 73:119–128, 1974.
Mark Wilson. Mechanics, classical. In *Routledge Encyclopedia of Philosophy*. Routledge, London, 1998.