Rough Draft! Please don't quote.

Hot and Heavy Matters in the Foundations of Statistical Mechanics

Craig Callender

Department of Philosophy, UCSD, La Jolla, CA 92093, USA

1. INTRODUCTION

Is thermodynamics true of self-gravitating systems? Or better put, does equilibrium statistical mechanics, the theory describing the microscopic basis of thermal phenomena, apply when the dominant coupling in a system is via (classical) gravitation? This question is the subject of increasing interest in astrophysics, but it is rarely pursued from a foundational perspective.[1] From this standpoint, the issue is not only fascinating in its own right, but it is an important prism through which to view other foundational projects in statistical mechanics. By considering it, we increase our understanding of the conditions under which a thermodynamic description of the world can emerge.

To motivate the question, notice that the long-range nature of the gravitational force causes many prima facie problems for a statistical mechanical account of a system. Unlike the equally infinite Coulomb force, the gravitational force does not saturate. In systems wherein Coulomb forces dominate, the positive and negative charges effectively cancel out at certain scales, e.g., in plasmas on scales larger than Debye spheres. At such scales the net interaction energy effectively vanishes, and this allows us to treat such systems as if they consist of many statistically independent subsystems. Such an assumption is at the heart of statistical mechanics. As a result of this non-saturation and related features, self-gravitating systems face many difficulties.

These difficulties are often dubbed the "paradoxes" of gravitational thermodynamics. Here is a quick taste of the difficulties. First, classical thermodynamics and statistical mechanics apply to systems in thermal equilibrium, but it's certainly not obvious what the equilibrium state of a self-gravitating system is. If it's a "singular" point, as many think, can such a monstrosity function as equilibrium? Second, in statistical thermodynamics the basic parameters of a state are extensive, yet self-gravitating systems typically cannot localize their gravitational potential energy enough to be considered extensive. The system's features are shaped by the overall bulk properties of the universe, not just near neighbors. Third, statistical mechanics requires that one be able to calculate the partition function of the relevant ensemble, but in self-gravitating systems the partition function can diverge. Fourth, in classical thermodynamics the specific heat is always positive, but in the gravitational case it is said to be frequently negative. By removing energy from a self-gravitating system, you can heat it up!

---

[1] For an entry to this literature, see the introductions to the collections by Campa et al 2008 and Dauxois et al 2002, in addition to Heggie and Hut 2003, Saslaw 2000 and Padmanabhan 1990.

These puzzles and others attract attention in astrophysics. Or rather, an applied version of these worries appears in this field. The reason it appears there is that various stellar systems are well modeled by the classical *N*-body gravitational equation, e.g., globular star clusters (N=10^5-10^6), galaxies (N=10^2-10^4), and of course, planetary systems. Because solving the million-body problem is analytically intractable, physicists naturally turn to statistical mechanics as a way of attacking these systems. The question becomes (roughly): if we think of the stars as the "particles" in a box of gas, interacting primarily via gravitation, do they execute behavior consistent with statistical mechanical principles?

To the extent that physicists have dealt with these paradoxes, the community divides over the fate of gravitational statistical mechanics. The conventional wisdom is that these paradoxes make statistical mechanics inapplicable. In its place various recipes and techniques remain, but the foundations from Maxwell, Boltzmann and Gibbs are judged inapplicable. By contrast, others react against the prevailing opinion and try to show that statistical mechanics does apply.

On the "nay" side, here is the chemist Rowlinson expressing a common sentiment:

> "[Thermodynamics] is essentially a human science; it started with steam engines and went on to describe many physical and chemical systems whose size is of the order of a metre. Its laws are not truly a theory but a highly condensed and abstract summary of our experience of how such systems behave. We have, therefore, no right to expect them to apply to other quite different systems, whether extremely large or extremely small. They clearly are inapplicable to the solar system or to galaxies. Here gravity is the dominant force; there is no equilibrium, the energy is no longer proportional to the amount of material, and so there are no *extensive* functions. Clausius's famous remark that the energy of the universe is constant but its entropy is increasing to a maximum is derived from the behaviour of a closed adiabatic system of constant volume. The universe is neither closed in any classical sense, nor of constant volume. Clearly classical *thermodynamics* is not a useful branch of science in cosmology; we have extrapolated too far from its human-sized origins." (1993, 873)

Hut 1997 writes that there are "fundamental problems that prevent a full thermodynamic treatment of a self-gravitating system of point particles" and that "the traditional road to equilibrium thermodynamics is blocked" (41). Regarding application, one prominent textbook explicitly raises the contrast:

> The question we now have to address is, what determines the particular configuration to which a given stellar system settles. Two classes of explanation are in principle possible: (i) The configuration actually adopted is favored by some fundamental physical principle in the same way that the velocity distribution of an ideal gas always relaxes to the Maxwell-Boltzmann distribution. (ii) The present configuration of the galaxy is simply a reflection of the particular initial conditions that gave rise to the galaxy's formation, in the

> same way that the shape of a particular stone in a field is due to particular circumstances rather than any general physical principle. (Binney and Tremaine 1987, 266)

By "fundamental physical principle" Binney and Tremaine essentially mean statistical mechanics. Because of a result we'll consider later, they come down in favor of (ii). Many others working in this area also appeal to one "paradox" or other as evidence that the core principles of statistical mechanics break down in self-gravitating systems. The claim that self-gravitating systems have no equilibrium, in particular, is the norm rather than the exception.

Against this opinion, one reads that

> "…statistical mechanics of gravitating systems is a controversial subject. However, our modern understanding of statistical mechanics and thermodynamics does handle gravitational interactions rigorously with complete satisfaction" Kiessling 1999, 545

> "on the contrary …[the] usual tools and ideas of statistical mechanics do apply to such systems, both at equilibrium and out of equilibrium" (Bouchet and Barre 2006, 19)

> "not all was well with these three arguments [three of the above problems]". Saslaw 2000, 301

These authors and others believe that statistical mechanics works for self-gravitating systems. None feel that the problems mentioned spell the end for statistical mechanical application. However, not all the answers on the "yea" side recommend the same answers to our puzzles. The motley character of their answers shouldn't be surprising—given the motley character of the foundations of statistical mechanics itself. With so many competing programs, there are many starting points and non-overlapping "core" principles in play.

No one denies that many statistical mechanical techniques can be applied to stellar systems. But are these merely residual mathematical tools with the core physical principles left behind? Answering this question relies on an antecedent division of statistical mechanical core and non-core principles, and this is bound to be controversial. The dispute as I have framed it so far can degenerate into a quibble about semantics, but there is much of interest here besides whether a set of techniques gets dubbed "equilibrium statistical mechanics". Very important physical principles are challenged by self-gravitating systems, and it's of interest how to respond to these challenges—regardless of how we categorize these principles.

While our discussion will have ramifications for the application problem, our primary focus will be on the logical problem of compatibility. Suppose (contrary to fact) that no actual stellar systems are good examples of classical N-body gravitation problems, and hence, that no actual system tests our question. Still the question exists of whether self-gravitating systems obey statistical thermodynamics.

As we'll see, discussion of this question will embroil us in many other foundational questions of statistical mechanics: the meaning of equilibrium, the direction of time, our understanding of phase transitions, the (non)equivalence of Gibbsian ensembles, and the use of the thermodynamic limit. Put together, these issues promise to teach us an awful lot about the foundations of statistical mechanics and the emergence of thermodynamics.

In what follows I introduce the so-called "paradoxes" of gravitational statistical mechanics with an eye on foundational matters. Many papers treat some small subset of the problems mentioned. It will be useful to have all the so-called "paradoxes" in one place, in order to compare them and see how they relate to one another. The paper's primary goal is to advertise the problems for further study—and hence it's longer on questions than on answers. However, where possible in such a short space I do hope to make some progress on the answers.

2. THE SYSTEM

The system we will consider, unless otherwise noted, is one composed of N particles, each of equal mass m, evolving in three-dimensional Euclidean space and interacting via a potential V. Such a system's Hamiltonian is

$$H = \sum_{i-1}^{N} \frac{\mathbf{p}_i^2}{2m} + \frac{c}{2}\sum_{i,j=1}^{N} V(\mathbf{q}_i - \mathbf{q}_j) \tag{1}$$

where c is a coupling parameter and $(\mathbf{q}_i, \mathbf{p}_i)$ are the canonical coordinates of the i-th particle. Long-range forces are typically associated with two body potentials $V(\mathrm{r}) = 1/r^s$, where s is less than or equal to the spatial dimensionality of the system. In the case of self-gravitation, s=1, $V(r)$=1/r, and c=-Gm^2, so (1) becomes

$$H = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} - \frac{Gm^2}{2}\sum_{i=1, i\neq j}^{N} \frac{1}{|q_i - q_j|} \tag{2}$$

With those substitutions (1) becomes Newton's famous N-body problem in gravitation theory. Although our emphasis will be on gravity, gravity is *not necessary* for many of the puzzles that follow. There are non-gravitational systems that display some of these behaviors and raise some of the same questions, e.g., so-called geophysical fluid dynamical models (s=0), some non-neutral plasmas (s=1), dipolar ferroelectrics and ferromagnetics (s=3), as well as some spin systems and toy models. Interestingly, many of the peculiar physical properties found in large stellar systems are also found in very tiny systems like nanoclusters. This is no accident. If we consider a system "small" if its spatial extension is smaller than the range of its dominant interaction, then a self-gravitating system is small too. In any case, the astrophysical applications of (1) make the gravitational case the most significant.

3. DIVERGENCES

Let's begin by arguing that one common problem for self-gravitating systems is in fact a pseudo-problem—or at least, if it is a problem, it's one common to short-range systems too. The problem is that self-gravitating systems have what might be called infrared and ultraviolent divergences.

The infrared divergence arises from the fact that the gravitational potential r^{-1} has infinite range. Since there is no shielding, strictly speaking the microcanonical density of states diverges. The density of states g(E) is given by

$$g(E) = \frac{1}{N!}\int \delta(E - H(q,p))dqdp$$

where E is the fixed energy, H the Hamiltonian, and N the number of systems. The above integral diverges if the range of spatial integrations is infinite. Indeed, the configuration space of an N-body self-gravitating system diverges badly:

$$\Gamma_q \sim \lim_{V\to\infty} V^{N-2}$$

(Cipriani and Pettini 2005). Since the Gibbs entropy $S(E)$ is essentially the logarithm of $g(E)$, this problem implies that the entropy diverges too. Since one gets a tremendous amount of thermodynamics from $S(E)$, one sees that this problem must be cured for any sensible thermodynamical behavior to emerge.

The ultraviolet divergence, by contrast, arises over short distance interactions. Here the problem is the local singularity of the Newtonian pair interaction potential. Two classical point particles can move arbitrarily close to one another. As they do so, they release infinite negative gravitational potential energy. Partition functions, which need to sum over all these states, thereby diverge. The entropy and other quantities, therefore, will not exist (see Rybicki 1971 and Kiessling 1989).

Let's begin by tackling the ultraviolet divergence. Technically, we can of course deal with this by replacing the gravitational potential with a softening or small cut-off potential. For instance, one can regularize the potential with the following replacement

$$\frac{1}{|r_i - r_j|} \rightarrow \frac{1}{\sqrt{(r_i - r_j)^2 + \eta^2}}$$

where η is a softening parameter that bounds from below the interaction potential. This softening will prevent arbitrarily large momenta from developing. The question is whether there is any justification for this replacement. Here I think the answer is yes. We can perform this maneuver with a clear conscience because we know that ultimately classical mechanics gives way to quantum mechanics, and quantum mechanics – via the

Pauli principle -- does indeed supply us with a mechanism preventing this kind of divergence.

The infrared divergence of g(E) is perhaps more of a problem. This occurs because the integral is taken over an infinite range of position space. However, as has been pointed out by others, there is nothing special here. The same divergence occurs for the non-gravitating ideal gas in a spatially infinite universe. (That issue has in fact been itself the subject of controversy among philosophers of science; see the four notes by Popper, ending in Popper 1958, and Hill and Grünbaum 1957.) Since we don't take this divergence to prohibit a proper thermodynamics obtaining, we shouldn't in the gravitational case either. The solution in the non-gravitational case is to put the gas "in a box" and the same can be done in the gravitational case. What we need to do is focus on the time and space scales. If a system is relatively isolated--if, say, its rate of evaporation is proportionally small—we can conceptually put a box around it. We can do so when the interactions with the "box" are much less dominant within the time scales of interest than the interactions among constituents. Eventually the interactions with the box, if it really were there, would be highly relevant to the physics. At this point the idealization breaks down. For some spatial and temporal scales and for some thermodynamic observables, however, the system-box interactions will be negligible. In principle there isn't a special problem with gravity. There remains a question of applicability: are there in fact self-gravitating systems and observables with spatial and temporal scales like this? The answer to that question seems to be yes, but that is beyond the scope of this essay.

The divergence problems, therefore, can be solved with plausible idealizations and approximations. The only stumbling block lay in mistakingly thinking such idealizations weren't needed all along in non-gravitational systems.

4. NO EQUILIBRIUM?

Since our inquiry is into equilibrium statistical mechanics, the natural next question to ask is whether the concept of equilibrium makes sense.[2] Many physicists voice skepticism about there being an equilibrium state for self-gravitating systems. In fact, it's common in many textbooks and articles simply to assert—with barely any justification—that self-gravitating systems lack an equilibrium point. This doubt is based primarily on the idea that gravity will cause increasing clumpiness in the system, to the point where the "final" state is an ill-defined singularity. The famous *gravitational catastrophe*, to be discussed below, reinforces this thought. Since I want to spend my time primarily on curiosities of non-extensivity, I can't give this question the attention it deserves here. It surely deserves a separate paper. Nonetheless, I want to argue that the skepticism about equilibrium is mostly unfounded. The gravitational catastrophe isn't actually so

[2] Callender 2009 is more-or-less an investigation of the related question whether Boltzmann's non-equilibrium statistical mechanics applies to self-gravitating systems.

catastrophic. Furthermore, there do exist exact models of self-gravitating systems for which even point-sized equilibrium states are well-defined. That will be enough for us here. Whether these models correspond to any real systems in the universe and whether in light of what one finds there one ought to alter one's sense of equilibrium is postponed to another time.

What is equilibrium, anyway? This is a foundational question raised by self-gravitating systems. Is it the state of maximum entropy? With self-gravitating systems in mind, some have challenged this identification (Landsberg 1984). If equilibrium is defined via maximum entropy, which entropy? Need equilibrium be defined via a partition function? Or should it be defined via the probability measure on state space? All of these questions and more come into play where self-gravitating systems are concerned. Let's begin by describing the reasons why many are unconvinced that self-gravitating systems possess a well-defined equilibrium state.

The first is related to the divergences mentioned above. If the entropy diverges and yet maximum entropy characterizes an equilibrium state, then we don't have equilibrium states. However, as we saw there, we can without embarrassment confine the system to a "box" and "soften" the potential. Doing so removes this reason for skepticism. And with Kiessling 1989, note that even if we don't cut off the close range effect of the potential, one could argue, as he does, that "the physical concept of a statistical mechanics equilibrium state is independent of concepts such as thermodynamic relations, from the beginning. Here, the basic quantity is the phase space probability measure μ, and that quantity has a meaning independently of whether Q [the configurational integral] exists or not" (205). A surprising amount of statistical mechanics can survive this difficulty.

A second reason for doubt stems from what is perhaps the most famous result in gravitational thermodynamics, the notorious *gravitational catastrophe*. Discovered by Antonsov 1962, dramatically elaborated upon by Lynden-Bell and Wood 1968, and found in many computer simulations, this phenomenon calls into question whether there is a largest most probable state toward which systems tend to evolve. If we take equilibrium to be defined as the state that maximizes Boltzmann entropy, then we have a worry.

Mathematically, one tries to find the density distribution $f(\mathbf{x},\mathbf{v})$, defined on 6-dimensional μ-space, that maximizes the Boltzmann entropy

$$S(f) = -\iint f \log f d^3\mathbf{x} d^3\mathbf{v}$$

subject to constraints on the total mass, energy, linear and angular momentum. This is a variational problem, and Antonov shows that it has no solution in the gravitational case. In fact, it has no solution even when we artificially confine the system to an arbitrarily shaped "ball" of physical space and allow radial asymmetries (Robert 1998). This result also comes with a physical gloss. The idea, in brief, is that if we begin with a system in a stably bound "core-halo" state, the central core as it develops with time will become increasingly concentrated without end. The core loses energy to the surrounding "halo" of particles, but because of its negative heat capacity, it will get hotter as it contracts. If the surrounding halo can't warm up quickly enough, then we get runaway instability. The system gets more and more inhomogeneous, both in configuration and velocity space, without limit. Hence its entropy increases without limit. For any state--since we're dealing with point particles--there is always a higher entropy state found by making the core denser and the halo more diffuse. There is no state of maximum Boltzmann entropy; therefore, if this is the criterion we use, there is no final equilibrium point.

This result, as fascinating as it is, has little bearing on our topic. The demonstration is within astrophysics and awash in idealizations. These idealizations are perfectly natural ways of modeling some astrophysical phenomena[3], but they do limit the applicability of this result for our foundational question. First, the system is assumed to be in virial equilibrium, but in the long run this won't always be correct. The corrections needed for describing the system when outside virial equilibrium may significantly affect the catastrophe. Second, the demonstration assumes that the formation of binary star systems will be unimportant, but eventually this may not be negligible—even for realistic systems. Third, the Boltzmann entropy that is maximized is the one over the one-particle distribution function in 6-dimensional μ-space, not the volume corresponding to the macrostate in $6N$-dimensional Γ-space. The former is fine as an approximation of many stellar systems, for they don't have many close encounters or collisions. But at many stages of the above process additional physical mechanisms will become important. If we think of the logarithm of the $6N$-dimensional volume as the "real" entropy, then we've only shown that an approximation for some spatial and temporal scales doesn't have an equilibrium point. Fourth, and most importantly, catastrophic behavior occurs for only certain combinations of the mass, energy and radius of the system. Other combinations, such as for so-called Lane-Embden isothermal spheres, possess an asymptotic state that obeys Maxwell-Boltzmann statistics. So even if we ignored all the idealizations, the gravitational catastrophe shows only that <u>some</u> self-gravitating systems suffer from problems in defining equilibrium.[4]

A third problem for equilibrium stems from a combined intuitive repugnance coupled with skepticism about the well-definedness of the presumed equilibrium state. It is widely assumed (we'll see why momentarily) that the notion of equilibrium appropriate to gravitational thermodynamics is a state in which all N point particles are sitting on a single point. Equilibrium, it is often thought, instead should be an extended state, not a monstrosity like that. Kiessling, however, counsels that our ordinary notion of thermal equilibrium originates from the non-gravitational cases; therefore, it shouldn't bias us against non-orthodox states appearing in more unfamiliar systems:

> Let us therefore try to take a less biased standpoint and consider a generalized meaning of equilibrium thermodynamics simply as a tool that describes the average fate of physical systems possessing an overwhelmingly large number of degrees of freedom, without making too restrictive assumptions about the

---

[3] See Ipser 1974 on this topic and also equilibrium.

[4] Another similar reason for skepticism is that in mean field theory, where one uses the Vlasov equation to model stellar systems, there are <u>many</u> equilibrium (stationary) solutions. But this takes the equilibria found in the mean field limit to be the genuine article.

properties of the final state….(205)

With the intuitive concerns put aside – as I think they should be – the question becomes whether such a "singular" state is well-defined and possibly part of a functioning statistical mechanics. As it turns out, it is.

Before sketching how, let's step back and ask the prior question of why so many think equilibrium will be a state with all particles resting on a single point. In fact, let's step even further back and begin by noting that the choice of equilibrum state is really absolutely basic in statistical mechanics. One can't prove that a state is equilibrium, at least not in the sense of showing that most systems will end up in such and such a state. Any such proof will of course depend on the probability measure one adopts, but that is the very topic at issue. The best one can do is argue from physical considerations that such-and-such a state is expected, and then show mathematically that such a state can be characterized and that it plays the roles you want of it. Importantly, one then shows that this description characterizes the behavior one initially sought to describe; that is, one hopes there is a connection between the probability measure and the actual frequencies. If not, then back to the drawing board. "Finding" equilibrium can be a messy process.

In the mathematical physics literature it is assumed that the equilibrium state of a self-gravitating system is a single material point. The motivation for this is thinking about an energetically open system confined to a box (as in Section 3), but without a short-distance cut-off. It is expected that gravitational energy will gradually be lost to the outside, resulting in a more and more collapsed state. It is also expected that angular momentum built up from the formation of binaries will eventually be carried outside the box. The end result of this process is the most collapsed system possible, the state when all the particles sit on a single point. This state is identified as equilibrium. The system as described is most naturally represented with the canonical ensemble (N particles trapped in a box, energetically open). Since the momentum and configurational equilibrium densities factorize in the canonical ensemble, yielding a trivial Gaussian measure over the momentum sector (see Kiessling, 213), Kiessling and others focus on the configurational sector. The question posed is whether the canonical equilibrium measure can describe this collapsed state in $3N$-dimensional configuration space.

Kiessling's 1989 proof that it can for three-dimensional systems is a major result. There he examines modified gravitational potentials with smoothed out singularities. He shows that in the limit, as these potentials approach the exact Newtonian potential, the canonical ensemble converges on a superposition of Dirac distributions each describing a collapsed state. Which exact spot will suffer the collapse is equal to a probability density that depends on the external potential and a Boltzmann-like factor. Kiessling is then able to show that this work connects up with thermodynamics in the mean field limit (see Section 8).

Now, as imaginative and rigorous as Kiessling's work is, one can fault it along various lines. One might object to his assumption that the potential energy of the wall of the box be independent of the particles' momenta. Or one might have wanted to start with an isolated system and the microcanonical ensemble, especially given that foundationally one typically justifies the canonical from the microcanonical, not vice versa. Then perhaps a state of maximal dispersal in configuration space is most natural due to evaporation. We will not pursue these or similar questions, although I invite readers to do so. Obviously more work is needed. However, we have in this result a rigorous proof that equilibrium is well-defined by the canonical equilibrium measure even for what is in many ways the worst-case scenario, the situation where all particles sit on a single material point. That is enough for us to press forward.[5]

5. NON-EXTENSIVITY

Intuitively put, extensive quantities are those that depend upon the amount of material or size of the system, whereas intensive quantities are those that do not. The mass, internal energy, entropy, volume and various thermodynamic potentials (e.g., F,G,H) are examples of extensive variables. The density, temperature, and pressure are examples of intensive variables. Mathematically, the most common expression for extensivity is the definition that a function f of thermodynamic variables is extensive if it is homogeneous of degree one. If we consider a function of the internal energy U, volume V, and particle number N, homogeneity of degree one means that

$$f(aU, aV, aN) = af(U,V,N) \qquad (3)$$

for all positive numbers a.

In thermodynamics and statistical mechanics both the energy and entropy are assumed to be extensive. This neatly encodes the experimental facts. Consider a box of gas in equilibrium with a partition in the middle and consider the entropy, so that a=2 and f=S. Then (3) describes the experimental fact that if we double the U,V and N of the system then we also double the entropy. The same goes for the energy. Thermodynamically, we can take energy or entropy to be the dependent variable, U=U(S,V,N) or S=S(U,V,N); if one is not extensive then the other is not. Statistical mechanically, matters are a bit subtler.

Extensive functions are often conflated with additive ones. A function—using our example, entropy—is additive if $S(U_1+U_2, V_1+V_2, N_1+N_2)=S(U_1+V_1+N_1)+S(U_2+V_2+N_2)$. With minimal assumptions homogeneity of degree one does imply additivity. But the two concepts are of course distinct and can come apart in some systems; but for most realistic physical systems they stand or fall together. For this reason, except as noted,

[5] Interestingly, note that few in the astrophysics community care about whether there is a "final" equilibrium point, a global maximum entropy. Often using the one-particle distribution form of the Boltzmann entropy, the field is focused on so-called meta-stable states, states of local entropy maxima, as the relevant object of study (e.g., Chavanis 2005). Stellar systems typically evolve very slowly, so slowly that many variables can be assumed essentially fixed. And in many numerical simulations one finds relatively stable long-lasting states whose average values for thermodynamic parameters don't fluctuate wildly. Statistical mechanics is then applied to these meta-stable or quasi-equilibrium states.

we'll use the two terms interchangeably. See Dunning-Davies 1983 and Touchette 2002 for excellent discussions of extensivity and additivity.

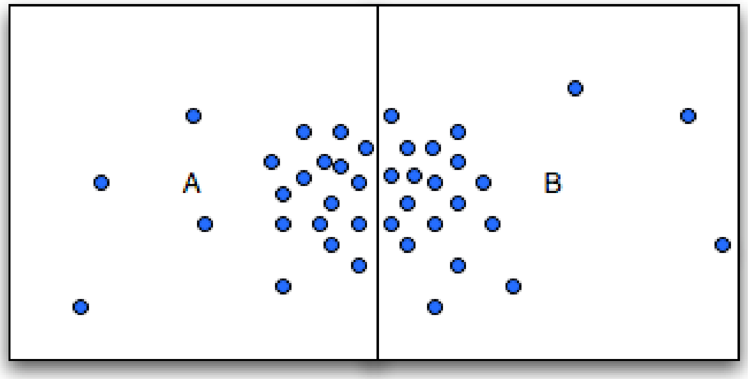
The problem we want to consider arises from the fact that extensivity (of the entropy and energy) seems to not even be approximately true for self-gravitating systems. Yet without extensivity holding, it's not clear how much of equilibrium thermodynamics or statistical mechanics survive.

Think about energy. The assumption that energy is extensive is plausible partly because in terrestrial cases we are usually dealing with short-range potentials. At a certain scale matter is electrically neutral and gravity is so weak as to be insignificant. If the potential is short-range and our subsystems aren't too small, then the subsystems will interact with one another only at or in the neighborhood of their boundaries. When we add up the energies of the subsystems, we ignore these interaction energies. The justification for this is that the interaction energies are proportional to the <u>surfaces</u> of the subsystems, whereas the subsystem energies are proportional to the <u>volumes</u> of the subsystems. So long as the subsystems are big enough, the subsystem energies will vastly trump the interaction energies as the number of subsystems increases because the former scale as $(\text{length})^3$ and the latter as $(\text{length})^2$. The assumption is thereby justified.

However, if gas molecules are replaced by stars—that is, short-range potentials replaced by long-range potentials—this reasoning doesn't work. The interaction energies may not be proportional to the subsystem surfaces. For short-range potentials, the dominant contribution to the energy comes from nearby particles; but for long-range potentials, the dominant contribution can originate from far away particles. To drive home the point, consider a sphere filled with a uniform distribution of particles. Now add a particle to the origin and consider its internal energy U:

$$U \propto \int_0^R 4\pi r^2 dr \rho r^{-3-\varepsilon} \propto \int_0^R dr r^{-1-\varepsilon}$$

One can verify that with $\varepsilon>0$ ("short-range potentials"), the significant contribution to the integral comes from near the particle's origin, whereas with $\varepsilon<0$ ("long-range potentials"), the contribution comes from far from the origin (Padmanabhan 1990). The interaction energies do not become negligible as the system grows.



As a result, think of what happens in the standard case of a chamber of gas divided into two equal boxes, A and B. If the particles are interacting via long-range forces, the particles in box A will feel the particles in box B as much or even more than the particles nearby. Let E_A represent the energy of box A and E_B represent the energy of box B. Because

$$E=E_A+E_B+f(E_A,E_B)$$

where $f(E_A,E_B)$ is the non-negligible interaction energy, it is easy to devise scenarios whereby $E_A = E_B = -a$, where $a>0$, yet where the energy of the combined system E vanishes. The energy might not be even approximately additive. Same goes with the entropy.[6]

Of course, the problem is really one of degree. Real finite systems aren't perfectly extensive. The Coulomb forces are just as long-range as the gravitational. The difference is that at certain scales the Coulomb forces are screened. Are they perfectly screened? No. Systems interacting primarily via the Coulomb force are therefore non-extensive too—strictly speaking. Nonetheless, there is a real physical difference between the gravitating and non-gravitating cases, for one has screening mechanisms the other doesn't. It's important not to let the distinction collapse merely because it's ultimately one of degree.

One way of collapsing the distinction appears in "rigorous" treatments of statistical mechanics. Here one tries to regain the exact formal properties of thermodynamics despite the fact that we know it holds only approximately. Elsewhere I have dubbed this practice "taking thermodynamics too seriously" (see Callender 2000) and pointed out that that it can often lead us astray. In this case it is simply a matter of semantics. In approaches to statistical mechanics like Ruelle 1968's, it's commonly said that <u>no</u> finite system is <u>ever</u> extensive. What is meant is that in approaches wherein one defines the

[6] There are ways of re-defining the entropy of energy so that one or the other but not both are non-additive; see Johal, R. 2006.

entropy and energy via the partition function, one is unable to reproduce the distinction between extensive and intensive for any finite system. No matter how large the system, if it's finite, surface effects contribute to the partition function. Entropy and energy are claimed to be truly extensive only in the thermodynamic limit, where N,V go to infinity while the ratio N/V is held constant. In this field a variable f is extensive if the thermodynamic limit of f is infinity while the thermodynamic limit of f/V is constant. Strictly speaking, only in infinite systems can entropy and energy be truly extensive. Hence all real systems, be they gravitational or not, are strictly non-extensive. Of course, it is open to retort that we only deal with finite systems (field theories to one side). Clearly the idealization of strict additivity, without even microscopic surface interactions among subsystems, is a strong idealization.

But no matter. It's still a fact that self-gravitating systems, unlike Coulomb systems, have no built-in tendency causing them to be even approximately extensive, in the sense defined earlier. This physical difference, although invisible in certain contexts, is no less real for being a matter of degree. Indeed, the cost of the energy and entropy being non-extensive in this sense is quite profound.

6. EXTENSIVITY IN THERMODYNAMICS

Let's concentrate initially on the costs incurred by classical phenomenological thermodynamics. The trouble with doing so is that there is no one canonical way of developing the theory. Some authors take as axioms what others derive, and vice versa. So given any isolated assumption that is falsified by non-extensivity, one can always find another treatment where that assumption is not basic but derived. That said, non-extensivity wreaks havoc with traditional thermodynamics no matter the route by which it reverberates through the subject. In most treatments, extensivity is axiomatic. Rief 1965 takes it as axiomatic and Landsberg 1961 suggests that the claim that all thermodynamic variables are either extensive or intensive might be raised to the level of the "Fourth Law" of thermodynamics.

For example, the now-classic textbook of Callen 1960 begins with a handful of postulates. Postulates I, II and III either explicitly (in III) or implicitly (I and II) assume the additivity/extensivity of the system. Postulate III states that "the entropy of a closed system is additive over the constituent subsystems" (25), so additivity is treated as an axiom. Postulate II requires that entropy is a function of the extensive parameters of the system in equilibrium. In that case, if energy is not extensive then entropy is not a function of energy. Worse than that, the fundamental equation of state from which scores of results are derived would no longer be $S(U,V,N)$ but $S(V,N)$. Even Postulate I, the existence of equilibrium states, tacitly requires extensivity of the entropy and/or energy, for Callen declares an equilibrium state as being completely characterizable by extensive parameters. Of course, there are extensive parameters in addition to energy and entropy, but a formalism based around, say, volume and mole number, would scarcely be recognizable as thermodynamics. With the first three basic postulates impacted by the loss of extensivity, scores of important implications are lost, e.g.:

- One needs additivity to derive that subsystems mutually in equilibrium have equal temperatures (Callen, 37).

- If subsystems mutually in equilibrium don't have to have equal temperatures, then the transitivity of equilibrium also fails, i.e., the "Zeroth Law" of thermodynamics (that if two thermodynamic systems are each in equilibrium with a third, they are also in equilibrium with each other) fails.

- When a system is in equilibrium, its large subsystems also will be in equilibrium. Clearly this doesn't follow if the system is non-extensive.

- The Gibbs-Duhem equality, in which one shows that three intensive quantities are interdependent, is one of the most important equations of classical thermodynamics. Yet it requires additivity in its derivation (in order to use Euler's theorem) and so it's not generally true for non-extensive systems.

- In many axiomatic treatments of thermodynamics, extensivity is a requirement for the equilibrium Second Law (see, e.g., Lieb and Yngvason 1998).

The reader can no doubt think of more trouble caused by non-extensivity. He or she is invited to go through a thermodynamics text and circle all the derivations lost by removing the assumption of extensivity. I'll assume, however, that the above list suffices as an indication of the damage.

7. EXTENSIVITY IN STATISTICAL MECHANICS

Moving to statistical mechanics, matters are in many ways more interesting. However, in the end, here too the approximate extensivity of the energy is often taken as axiomatic.

For example, Landau and Lifshitz 1969 take the statistical independence of subsystems and the associated additivity as perhaps their most basic assumption, for through a chain of reasoning they use it to ground the "most important conclusion…[that the] values of the additive integrals of motion … completely determine the statistical properties of a closed system" (11). With this, they write, it is possible to replace the "staggeringly large data" of the mechanical approach with a simple statistical distribution, i.e., the microcanonical ensemble, the basis of equilibrium statistical mechanics (11). It is not an exaggeration to say that extensivity of energy, in their treatment, is what makes statistical mechanics possible.

To take another example, consider Ruelle's 1969 elegant and influential treatment of statistical mechanics. In his hands, "normal" thermodynamic behavior is possible *only when* two conditions hold, stability and temperedness. We won't spend time describing these conditions here, but suffice to say, in self-gravitating systems stability is violated.[7] Interestingly, if we make our potential repulsive rather than attractive at large distances, then the temperedness condition is violated (Campa, 11). So one way or the other, long-range forces spell trouble for thermodynamic behavior as Ruelle understands it.

As in the thermodynamic case, one finds extensivity assumed throughout many treatments of statistical mechanics. Not surprisingly, since statistical mechanics grounds thermodynamics, it's assumed in statistical mechanics in many of the same places it is in thermodynamics. However, let's note a few additional features of interest:

- Let subsystem 1(2) correspond to the element of phase space $d\mathbf{p}^{(1)}d\mathbf{q}^{(1)}$ ($d\mathbf{p}^{(2)}d\mathbf{q}^{(2)}$) and the combined subsystem correspond to the element $d\mathbf{p}^{(12)}d\mathbf{q}^{(12)}$. Statistical independence means simply that the probability ρ the joint subsystem lies in $d\mathbf{p}^{(12)}d\mathbf{q}^{(12)}$ is equal to the direct product of the individual probabilities of lying in the individual elements of phase space, i.e., $\rho^{(12)}d\mathbf{p}^{(12)}d\mathbf{q}^{(12)} = \rho^{(1)}d\mathbf{p}^{(1)}d\mathbf{q}^{(1)} \cdot \rho^{(2)}d\mathbf{p}^{(2)}d\mathbf{q}^{(2)}$. Thus, in terms of the Gibbs entropy, $S = -S\rho^N[\log\rho^N]$, one can prove that S is additive iff ρ is factorizable among its component systems (Touchette 2002). If ρ is not factorizable, then the entropy won't be additive and all of the features hanging on entropy being additive won't follow. What we have seen is that in self-gravitating systems the Gibbs entropy may not be even approximately factorizable. It does, however, still exist – whereas it may be argued that non-extensive entropy in thermodynamics is a contradiction in terms, for on many accounts thermodynamic entropy is definitionally extensive.

- Statistical mechanics is based on the use of three probability distributions, or ensembles: the microcanonical, the canonical, and the grand canonical. The most commonly used ensemble is the canonical. This ensemble is used for a system in contact with a heat bath at a certain temperature. The system can exchange heat but not matter with the heat bath, so the macroscopic parameters are V,T, and N, and E is not fixed. Though used less often, the microcanonical ensemble is often regarded as more basic. This ensemble represents a closed system with E,V and N fixed. The distribution is the one that is uniform over the energy hypersurface corresponding to the system of interest, and one hopes to justify this distribution in some way. The justification of the canonical ensemble is then considered parasitic upon the justification of the microcanonical ensemble. One justifies or simply assumes the uniform measure for a closed system. Then one justifies the canonical ensemble by assuming the system of interest is a small subsystem of a much larger system itself described by the microcanonical ensemble. With the added assumption of additivity of energies, one can then derive the canonical ensemble as an appropriate probability distribution for the subsystem. Without this additivity, one expects strange non-canonical behavior to emerge from non-extensive systems in contact with heat reservoirs.

- Non-extensive systems have no conventional thermodynamic limit. The thermodynamic limit is one where we normalize the extensive variables by the

[7] Campa 2008 contains a good treatment of the reason why this is, and Katz 2003 is good on stability.

volume *V* or particle number *N*, e.g., *E*/*N*, thereby making the ratios intensive; then we let $N,V \rightarrow \infty$ while keeping these intensive ratios finite. For a rigorous discussion of this problem see Padmanabdum 1990 and references therein. Intuitively, this fact is not surprising. The existence of the thermodynamic limit depends on making the contribution of surface effects go to zero as $N,V \rightarrow \infty$. But with a non-additive system, we saw that the surface effects aren't going to get smaller as *N* and *V* increase. If we have a self-gravitating system in a box, as we increase the size of the box, the total kinetic energy will scale with the mass, which in turn scales with the size of the box: $E_{kin} \propto M \propto R^3$. However, the potential energy E_{pot} will grow much faster: $E_{pot} \propto M^{5/3}$. This makes the specific gravitational potential energy of the system, the total potential energy divided by N, increase indefinitely as N increases (see Heggie and Hut 2003). Another way the lack of a thermodynamic limit is unsurprising is that the connection between it and extensity is sometimes especially direct. Beck and Schloegl 1993, like many others, state that a system is said to possess a thermodynamic limit only when its total entropy and energy are extensive quantities (in the rigorous sense described earlier). And adding one more link to the chain, Ruelle proves that the thermodynamic limit exists only for interactions that are thermodynamically stable, whereas non-extensive systems aren't stable. Ironically, the thermodynamic limit doesn't apply to very large systems unless the force relevant to such systems, i.e., gravity, is ignored.

- There can be negative specific heats in self-gravitating systems. The specific heat capacity *C* is a measure of the amount of heat energy required to raise the temperature of a material one degree Kelvin. In classical thermodynamics, C is always positive. For such a system, adding energy makes the temperature increase and removing energy makes the temperature decrease. Because it is so closely tied to the stability of a thermodynamic system, some texts treat positive C as a basic principle of thermodynamics. In the self-gravitating systems literature, however, it is a commonplace that there are systems with negative heat capacity. Removing energy from stellar systems can make them hotter. There is a lot of controversy here and much I could say, but I must leave this topic for another paper. Suffice to say, negative specific heat is quite counter-intuitive; furthermore, it is connected to the onset of phase transitions and ensemble inequivalence, two fascinating topics. See Lynden-Bell 1999 and Touchette et al 2004.

8. EXTENSIVITY REGAINED?

The physics community's reaction to non-extensivity divides into three broad camps. One group seeks to extend/modify equilibrium statistical mechanics to cover cases of non-extensive energy and entropy. Another uses the problems mentioned as a reason to ditch equilibrium statistical mechanics in favor of non-equilibrium techniques. But probably the majority of physicists pursue a conservative approach. Instead of modifying or abandoning equilibrium statistical mechanics, they continue to use it but seek approximations, limiting regimes, and so on, wherein extensivity is regained. In this section I'll focus on this last response. Since the other two groups pursue agendas outside the scope of this paper, I'll ignore them apart from a little discussion of the first group in Section 9.

Divide our N-body system up into two cells, 1 and 2. Due to the gravitational correlation energy U_{int} between cells, we've seen that the internal energy isn't extensive:

$$U = U_1 + U_2 + U_{\text{int}}.$$

Yet we can still ask whether there are systems for whom

$$U_{\text{int}} << U_1 + U_2.$$

If so then extensivity is a good approximation and ordinary equilibrium statistical mechanical techniques are used again. Let's investigate two ways of regaining extensivity.

First, here is a quick and loose argument one sometimes sees (e.g. Chavanis 2005) to show that entropy *can* be extensive for self-gravitating systems. Define the entropy via the standard

$$\frac{1}{T} = \frac{\partial S}{\partial E}$$

where E is the total energy and T the temperature. Then make two assumptions: (a) that the system is virialized, and (b) that the temperature is proportional to the mean kinetic energy. The first assumption holds if the system is stably bound. Then the virial theorem states that

$$\langle K \rangle_\tau = \frac{1}{2} \langle V \rangle_\tau$$

where τ is the time over which the averages are taken. The second assumption is that

$$T \propto \frac{\langle K \rangle}{k_B (3N - 6)}$$

where k_B is Boltzmann's constant. As one can quickly inspect, it then follows that both E and T are proportional to N; in fact they have the same N-dependence. Hence the entropy S, defined in terms of these notions, must also have similar N-dependence. S is therefore extensive.

This argument is problematic at every step. The virial theorem is a theorem of mechanics often used in astrophysics. It says that *if* the average time derivative of a function G known as the scalar virial is zero, then in cases like those of present interest the average potential energy is proportional to two times the average kinetic energy. In astrophysics the antecedent is often ignored if the system is stably bound for a long time. The reasoning is that if the scalar virial is bound between $G_{\min}$ and $G_{\max}$ for a long time, as happens in orbits, then its average is zero. For the time scales of interest, this assumption is usually perfectly innocuous. However, a priori there is no reason to think these time scales are equal to the thermodynamic times scales. Worse, that the system is in what is sometimes called "dynamical equilibrium", i.e., in a stable orbit, doesn't imply that it is in thermal equilibrium. The virial theorem holds regardless of whether the system is in thermal equilibrium. That the system is stably bound for a time period τ doesn't in the least imply that we can assign it a thermodynamic temperature. Furthermore, even if we do assume the system is in thermal equilibrium, it doesn't follow that its temperature is proportional to kinetic energy. There are many systems—and in fact some gravitating systems—where this is not so.

The quick argument needs more justification before it can be employed. Nonetheless, it does at least show that if these assumptions can be consistently employed, then the energy and entropy can be extensive even if the system is self-gravitational.

The more common method of regaining extensivity is by going to the mean field limit. First, one rescales equation (1) by what is known as the "Kac" prescription (Kac, Uhlenbeck and Hemmer 1963). That is, one chooses the spatial and temporal parameters so that the coupling parameter c=-Gm^2 becomes c=±1/N. Hence (1) becomes

$$H = \sum_{i=1}^{N} \frac{\mathbf{p}_i^{\,2}}{2m} + \frac{1}{2N} \sum_{i,j=1}^{N} V(\mathbf{q}_i - \mathbf{q}_j)$$

Now, keeping all parameters fixed, and in particular the volume, one takes the $N\to\infty$ limit. This differs from the usual thermodynamic limit, for in that limit the volume V also goes to infinity. The resulting limit has a number of nice features, for example, from it one can obtain the collisionless Boltzmann equation (also called the Vlasov equation), which is quite useful in astrophysics. For our purposes, however, what is important is that the energy H is proportional to N at this limit, and hence, the energy is extensive. In the literature it is commonly asserted that the $N,V\to\infty$ limit is appropriate for short range systems but the $N\to\infty$ limit is what is appropriate for long range systems. (In a similar development, de Vega and Sanchez 2002 argue that in the so-called "dilute limit", where $N,V\to\infty$ while $N/V^{1/3}$ stays fixed, the internal energy, free energy and entropy become extensive; however, this claim has been disputed (Laliena 2003).) We'll return to these limits in the next section.

Finally, I should mention two further regimes in which non-extensivity can be regained, but which I won't describe in detail here. Both are by William Saslaw. Saslaw 2003 is focused mostly on infinite N systems and the canonical ensemble. But he does mention some of the problems we discuss here and offer justifications for his methods. Although the first justification strictly departs from our Newtonian N-body situation, it is too good to ignore: Saslaw shows with a neat argument that the expansion of the universe effectively cancels out the long-range part of the gravitational force. This limits the bulk effects of gravity to certain lengthscales. So there is a gravitational kind of screening after all. And at these lengthscales the energy is approximately additive. The second justification focuses on features of the so-called correlation length used in astrophysics, e.g., its steepness, and shows that in some cases it too will leave the interaction energy negligible for certain spatial and temporal scales.

9. FOUNDATIONAL QUESTIONS

Gravitational thermodynamics stimulates many foundational questions in statistical mechanics. We have already broached some of them – e.g., what is equilibrium, whether systems are strictly extensive, the status of divergences, the problem of negative specific heat – and no doubt each can be discussed in far greater detail. Although there is much one could discuss, let me finish by concentrating on three philosophical points little discussed in the literature.

What Justifies the Thermodynamic Limit--and Which Limit?

The lack of a well-defined conventional thermodynamic limit is prima facie disturbing. Many approaches in the foundations of statistical mechanics crucially rely upon this limit. It is also said to provide the resolution of various "paradoxes" in statistical mechanics (see Styler 2004). However, I want to focus on a question raised by the result mentioned in the previous section, where we saw that extensivity can be regained by using an unconventional thermodynamic limit. In fact, it's not uncommon for physicists to use modified thermodynamic limits to achieve their ends. What should we make of this practice? Is one limit the "right" physical limit and the other the "wrong" one, given the goals? What makes a limiting procedure correct? Is a pragmatic justification, i.e., that the limit succeeds in getting the desired results, enough? Or do we require a more physically based justification?

Readers of this journal will be familiar with the funny business that can accompany the unrestrained use of limits. My favorite example (told to me by John Norton) is that in the limit where the blades are infinitely long a helicopter can generate lift even when its rotor is still. And more seriously, one can object to limits that alter the lawlike structure of a theory, such as limits that let fundamental constants go to infinity or zero, e.g., notoriously, the limit wherein Planck's constant goes to zero. With so much opportunity for mischief, one naturally wants to be on guard against reverse engineering: assuming, for instance, that extensivity holds and then taking the limits needed to get there. These issues definitely arise in the present case. In Lalienda's critique of deVega and Sota's "dilute limit", the author charges deVega and Sota with imposing "ad hoc" conditions to obtain a well-defined "dilute" thermodynamic limit. This criticism presupposes that some restrictions are legitimate and others not. By contrast, while not exactly denying this, in a side remark Barre and Bouchet 2006 approach the topic by demanding a limit that gives physics largely independent of how large *N* is. That is, they want to find the physics of effectively intensive variables. But in the context of the foundational question of whether one can regain extensive versus intensive variables, this motivation would seem to beg the question. It says: take whatever limit is needed to recover the 'intensive versus extensive' split. Obviously a criterion of some kind is needed to determine what limits (with what restrictions) justify what physical claims. I can't articulate one here, but I can gesture in the direction I think such a criterion can be found.[8]

Thermodynamics is a coarse-grained, approximate, high-level theory. The theory has many formal features, e.g., extensivity, that don't hold exactly at the microlevel. At the fundamental level it's not true that the entropy of a system is extensive. If you double the parameters E,V, and N, you don't exactly double the entropy. Surface effects exist and can contribute, albeit not much for large terrestrial systems. We should accept this lack of perfect extensivity at the microlevel without embarrassment. That said, our microlevel theory had better recover the phenomena. The phenomena are approximately extensive when N is large. How do we know our microphysical theory implies this?

One quick way to reassure oneself–though it is <u>neither necessary nor sufficient</u> for the job--is to show that this feature holds in the thermodynamic limit. The thermodynamic limit is a kind of expedient for this justificatory enterprise. Sure enough, in the thermodynamic limit, one can show that for many systems doubling E,V and N does exactly double the entropy. This doesn't tell us everything we need to know. It doesn't *guarantee* that for large finite N extensivity holds or even approximately holds (if the limit is not smooth). All the usual "short-run" versus "long-run" problems can surface. If one found out that for N=10^{23} the system's entropy was nowhere near doubled for doubled E,V, and N, then one would really worry. In this sense the thermodynamic limit isn't sufficient to assuage our worries. Neither is it necessary. If one had a guarantee that for all reasonable finite N extensivity approximately held, one wouldn't need the thermodynamic limit; after all, that regime is imaginary and the system doesn't have to be perfectly extensive. However, often we're not in a position to know that large finite N approximates thermal properties, so we make due with what we can, i.e., a proof in the thermodynamic limit. With this information one bets that the theory is on the right track.

Returning to the question of whether there is a "correct" thermodynamic limit, in one sense the answer is yes and in another no. The latter seems correct if we're asking whether one form of the thermodynamic limit, say, with V going to infinity, is privileged over other forms. Of course not. There is nothing in the above rationale that distinguishes for all types of systems the right limiting procedure for all ends. Yet the former seems right in the sense that there is a standard. The typical microphysics of the system at the thermodynamic scale provides the standard. When we use the conventional thermodynamic limit we usually tell a story about how the surface effects will be

[8] Liu 2001 is the only paper I know in the philosophical literature broaching this question. Also, it would be interesting to compare the conventional thermodynamic limit with the mean field limit with respect to "solving" Styler 2004's six paradoxes.

swamped by the volume effects, as in Section 2. This story is a sketch of a justification for looking at the physics when these "inessential" effects are driven down. In the absence of such a rationale, the choice of restrictions and variables to send to infinity can seem ad hoc.

Ensemble Nonequivalence: Will the True Ensemble Please Stand?

In mainstream Gibbsian statistical mechanics, the choice of probability distribution or ensemble is widely deemed a matter of convenience. One can choose the microcanonical, canonical, or grand canonical ensemble. They are supposed to be merely different windows onto the same physics, each corresponding merely to the fixing of different thermodynamic parameters. The physics isn't supposed to hang essentially on our choices on how we model the system.

What result supplies us with this confidence? It is widely claimed that the three ensembles are provably equivalent in the thermodynamic limit. Textbooks will gesture at facts like the following. Focusing on the canonical and the microcanonical ensembles, when N goes to infinity, the energies of most configurations in the canonical ensemble are equal to the average energy; hence in the limit one gets systems all of the same energy, i.e., the microcanonical ensemble. Reasoning like this, and the knowledge that there are relevant theorems proven, leads to the common belief that the three ensembles are equivalent in the conventional thermodynamic limit.

However, there is no completely general proof of equivalence of ensembles, so far as I am aware. Rigorous treatments (e.g., Ruelle) are more careful, claiming only that the three ensembles are equivalent in the thermodynamic for some or most physical systems. Some textbooks may even acknowledge that the three ensembles are inequivalent for systems undergoing phase transitions, a fact Gibbs 1902 pointed out more than a century ago. The so-called "problem of equivalence" has been a foundational problem for a long time, even before we turn to long-range systems. Galavotti 1999 writes that the problem of equivalence "is a fundamental problem because it would be a serious setback for the whole theory if there were different orthodic ensembles predicting different thermodynamics for the same system" (69). Dunning-Davies warns, "if one ensemble leads to one set of conclusions but a second leads to a different set, it is the whole theoretical discussion that should come under review." And Huang 1987 writes, "From a physical point of view, a microcanonical ensemble must be equivalent to a canonical ensemble, otherwise we would seriously doubt the utility of either" (148). With self-gravitating systems, it seems these worries are all the more pressing.

The reason it is more pressing, of course, is that without a well-defined conventional limit, even where we otherwise can prove equivalence of ensembles we cannot if they represent self-gravitating systems. (The unconventional limit described above doesn't reinstate ensemble equivalence.) In self-gravitating systems, the three ensembles can describe very different physics. In fact, in toy self-gravitating systems, one can demonstrate strikingly divergent physical behaviors for self-gravitating systems between the microcanonical and canonical ensembles. The former, for instance, allows negative specific heats whereas the latter cannot—and in gravitational thermodynamics this is regarded a virtue. Be that as it may, the situation has now changed: it's not just that we don't know (i.e., haven't yet proved) that the three ensembles are equivalent for some systems; it's that now we know that they're not equivalent for some systems, even ones not undergoing phase transitions.

What this nonequivalence means for the entire Gibbsian program has yet to be considered from a foundational perspective, but it would seem to call for serious study. When the ensembles are inequivalent, a number of questions face us. Do we distinguish one ensemble and view it as correct to the exclusion of the others? In fact, most of the gravitational thermodynamics literature does do this—but Gross 2001 most vocally. They take the real physics as being described by the microcanonical ensemble. And since one needs additivity to derive the canonical from microcanonical, one cannot derive the canonical from the microcanonical even if they aren't equivalent in the thermodynamic limit. Thus Combes and Roberts 2007 write that as a result of non-extensivity "the canonical formalism of Gibbs is no longer justified". Others, by contrast, stick with the canonical description as basic. It is, after all, the easiest to use, and so there is real incentive to standing by it.

Others consider moving to the use of new ensembles. Costeniuc et al 2006 define a new 'generalized canonical ensemble' by modifiying the Lebesque measure from which the canonical ensemble is crafted. They then show that this 'generalized canonical ensemble' is equivalent to the microcanonical measure in the conventional thermodynamic limit. They also consider the little-discussed Gaussian ensemble, which is the ensemble said in the first instance to be appropriate to a system in contact with a finite heat reservoir. This ensemble is also equivalent to the microcanoncal ensemble.

The phenomenon of ensemble nonequivalence and its connection to features such as non-concavity of the entropy, phase transitions, and negative specific heat have been the subject of an awful lot of research lately. The mathematical understanding of these properties and their relationships has advanced considerably over the recent years (for an introduction, see Touchette, Ellis and Turkington 2004). However, to my knowledge, there has not been a single paper in the philosophical foundations appraising this problem for the Gibbsian program nor recommending a course of action.

Tsallis Statistics

In response to the problems described here, there is a research program organized around non-extensive statistics. In 1988 Tsallis developed a generalization of the Boltzmann and Gibbs entropies, namely, the Tsallis entropy. The Tsallis entropy reduces to the Boltzmann and Gibbs entropies when the system is extensive, but is different otherwise. Since then a whole school devoted to non-extensive statistics has arisen based around this new entropy formula. The entropy is

$$S_q(p) = \frac{1}{q-1}\left(1 - \int p^q(x)dx\right)$$

where p the "probability" distribution (scare quotes because it's not clear they're really probabilities) and q is a positive real number interpreted as a kind of index of extensivity. When $q \rightarrow 1$ the Tsallis entropy reduces to the ordinary Gibbs entropy. The motivation behind the program is to show that the Tsallis entropy works well in situations where the Boltzmann and Gibbs entropies allegedly break down. Normal statistical mechanics is supposed to correspond only to the case where a system is approximately extensive. Long-range force systems like self-gravitating systems are supposed to be one prominent example where Tsallis statistics are needed.

The debate in the literature over the necessity of Tsallis statistics is sometimes very heated. Proponents of Tsallis statistics write papers displaying systems for which they claim traditional mechanical techniques come up short; defenders of statistical mechanics then respond with criticism of the new techniques and try to show how the Boltzmann-Gibbs framework suffices to account for the physics.

However, before one even engages in this back-and-forth, one might inquire from a foundational perspective whether the Tsallis entropy really makes sense. The paper by Nauenberg 2003 is quite critical of Tsallis entropy, but there is room for more debate. Those interested in the foundations of statistical mechanics may wish to weigh in on this debate.

11. CONCLUSION

As advertised, this discussion has raised more questions than it's answered. We are left with puzzles about the thermodynamic limit, ensemble non-equivalence, negative specific heats, the status of Tsallis statistics, the meaning of equilibrium, and more. That said, some answers have been sketched, and we have (I hope) learned something about how a thermodynamic description of the world arises. To a large extent I think what we have learned vindicates the way Landau and Lipfshitz begin their magisterial treatment of statistical mechanics. They view equilibrium statistical mechanics, and hence thermodynamics, as a description that emerges when certain systems on particular spatial and temporal scales become approximately extensive.

It bears mentioning that many possible topics have been left out. Of special note are the following. Self-gravitating systems are often said to be non-ergodic (see, e.g., Mukamel, Ruffo and Schreiber 2005). Since ergodicity is often said to be the dynamical foundation upon which thermodynamics depends, it's worth an investigation to see if self-gravitating systems enjoy the kind of dynamics (ergodic or not) that could sustain a thermodynamic description. Also, most of this paper has been written from the perspective of conventional Gibbsian statistical mechanics. However, when foundational questions arise, many authors prefer to think in terms of Boltzmann's picture (see, e.g., Lebowitz 1993). How would someone taking as fundamental the Boltzmann entropy – essentially the logarithm of the volume of $6N$-dimensiomnal phase space corresponding to the macrostate of interest – understand the puzzles surrounding self-gravitating systems?

Let me finish by returning to the original question: does thermodynamics hold of self-gravitating systems? The answer is unexpectedly unclear. The answer hangs on our evaluation of the means of regaining extensivity surveyed in Section 8. For instance, we saw that extensivity could be regained in the mean field limit. In this limit some but not all thermodynamic relationships will hold. So even if we find this limit to be unproblematic, our answer depends on a prior choice of what constitutes thermodynamics. Putting this semantics issue aside, should we really say that self-gravitating systems obey thermodynamics even if they only partially do in the mean field limit? We have begun we a very clear question and proceeded to muddy it up. Sometimes – and I hope this is one of those times – this is a sign of progress.

REFERENCES

Antonov, V. A. 1962. "Most Probable Phase Distribution in Spherical Star Systems and Conditions for Its Existence" *Vest. Leningrad University* 7, 135. Translated in *Dynamics of Star Clusters*, eds. Goodman, J. Hut, P. Dordrecht: D. Reidel.

Beck, C. and Schloegl, F. 1993. *Thermodynamics of Chaotic Systems: An Introduction*, Cambridge, UK: Cambridge University Press.

Binney, J.J. and Tremaine, S. 1987, *Galactic Dynamics,* Princeton University Press, Princeton, New Jersey.

Bouchet, F. and Barre, J. 2006. "Statistical Mechanics and Long Range Interactions", *Comptes Rendus Physique* 7, 414-421.

Callen, H.B. 1960. *Thermodynamics*. NY: John Wiley & Sons.

Callender, C. 2001. "Taking Thermodynamics Too Seriously" *Studies in the History and Philosophy of Modern Physics* 32, 539-53.

Callender, C. 2009. "The Past Hypothesis Meets Gravity" forthcoming in Gerhard Ernst and Andreas Hüttemann (eds.): *Time, Chance and Reduction. Philosophical Aspects of Statistical Mechanics*. Cambridge: Cambridge University Press.

Campa, A., Giansanti, A. Morigi, G., and F.S. Labini. (Eds.) 2008. *Dynamics and Thermodynamics of Systems with Long Range Interactions: Theory and Experiments* AIP Conference Proceedings, Vol. 970.

Campa, A. 2008. "The Study of the Equilibrium and of the Dynamical Properties of Long-Range Interacting Systems" in Campa, A. et al, (Eds.) 2008.

Chavanis, P.-H. 2005. "On the Lifetime of Metastable States in Self-gravitating Systems" Astronomy and Astrophysics 432, 117-.

Cipriani P. and Pettini, M. 2003. "Strong Chaos in N-body Problem and Microcanonical Thermodynamics of Collisionless Self Gravitating Systems". *Astrophysics and Space Science* 283, 347-368.

Combes, F., and Robert, R. 2007. "Comment on the Thematic Issue 'Statistical mechanics of Non-extensive Systems' [C. R. Physique 7 (3-4) (2006)]", *Comptes Rendus Physique* 8,

Costeniuc, M., Ellis, R.S., Touchette, H. and B. Turkington. 2006. "Generalized Canonical Ensembles and Ensemble Equivalence" *Physical Review* E 73, 026105.

Dauxois, T., Ruffo, S., Arimondo, E., and M. Wilkens (Eds.), 2002. *Dynamics and Thermodynamics of Systems with Long-Range Interactions*. Lecture Notes in Physics 602, Springer.

Dunning-Davies, J. 1983. "On the Meaning of Extensivity, *Physics Letters* A 94, Issue 8, 346-348.

Gibbs, J.W. 1902. [1960] *Elementary Principles in Statistical Mechanics.* Yale University Press, Yale, C.T., 1902. Reprinted by Dover, New York, 1960

Gross, D. 2001. *Microcanonical Thermodynamics: Phase Transitions in Small Systems*. Lecture Notes in Physics 66 (World Scientific, Singapore)

Grünbaum, A. 1973. *Philosophical Problems of Space and Time*. Second edition. *Boston Studies in the Philosophy of Science*, vol. XII. Dordrecht: D. Reidel.

Katz, J. 2003. "Thermodynamics and Self-Gravitating Systems", *Foundations of Physics* 33, 223-269.

Heggie, D. and Hut, P. 2003. *The Gravitational Million-Body Problem*. Cambridge: Cambridge University Press.

Hill, E.L. and Grünbaum, A. 1957. "Irreversible Process in Physical Theory", *Nature* CLXXIX, 1296.

Huang, K. 1987. *Statistical Mechanics*. New York: Wiley.

Ipser, J.R. 1974. "On Using Entropy Arguments to Study the Evolution and Secular Stability of Spherical Stellar-dynamical Systems". *Astrophys. Journal* **193**, 463-470.

Johal, R. 2006, "Mutual Thermal Equilibrium in Long-Range Ising Model using Nonadditive Entropy" *Physica A* 365, 155-161.

Kac, M., Uhlenbeck, G.E. and Hemmer, P.C. 1963. "On the van der Waals Theory of the Vapor-Liquid Equilibrium. I. Discussion of a One-Dimensional Model". Journal of Mathematical Physics 4, 216-228.

Kiessling, M. K.-H. 1989. "On the Equilibrium Statistical Mechanics of Isothermal Classical Self-Gravitating Matter" Journal of Statistical Physics 55 (1/2), 203-257.

Kiessling, M. K.-H. 1999. "Statistical Mechanics of Gravitational Condensation and the Formation of Galaxies". *Galaxy Dynamics*, D.Merritt, M. Valluri, and J. A. Sellwood (eds.). ASP Conference Series 182, 545-546. San Francisco: ASP.

Laliena, V. 2003. "On the Thermodynamical Limit of Self-gravitating Systems", *Nuclear Physics* B668, 403-411.

Landau, L.D., and Lifshitz, E.M. 1969. *Statistical Physics*, 2nd edition. Oxford: Pergamon.

Landsberg, P.T. 1961. *Thermodynamics with Quantum Statistical Illustrations*, NY: Interscience.

Landsberg, P.T., 1984. "Is Equilibrium Always an Entropy Maximum?", *Journal of Statistical Physics* 35, 159-169.

Lebowitz, J. L. 1993. "Macroscopic Laws, Microscopic Dynamics, Time's Arrow and Boltzmann's Entropy" *Physica A* 194, 1-27.

Liu, C. 2001. "Infinite Systems in SM Explanations: Thermodynamic Limit, Renormalization (Semi-) Groups, and Irreversibility" *Philosophy of Science* 68, S325-S344.

Lynden-Bell, D. and Wood, R. 1968. "The Gravo-Thermal Catastrophe in Isothermal Spheres and the Onset of Red Giant Structure for Stellar Systems" *Mon. Not. R. Astr. Soc*. 138, 495-525.

Lynden-Bell, D. 1999. "Negative Specific Heat in Astronomy, Physics and Chemistry" *Physica* A 263, 293-304.

Mukamel, D., Ruffo, S., and N. Schreiber. 2005. "Breaking of Ergodicity and Long Relaxation Times in Systems with Long-Range Interactions". *Physical Review Letters* 95, 240604.

Nauenberg, M. 2003. "Critique of *q*-entropy for Thermal Statistics", *Physical Review E* **67**, 036114.

Padmanabhan, T. 1990. "Statistical Mechanics of Gravitating Systems" *Physical Reports* 188, 285.

Popper, K. 1958. *Nature* CLXXXI, 402.

Reif, F. 1965. *Statistical and Thermal Physics*. New York: McGraw-Hill.

Robert, R. 1998. "On the Gravitational Collapse of Stellar Systems" *Classical and*

Quantum Gravity 15, 3827–3840.

Rowlinson, J. S. 1993. "Thermodynamics of Inhomogeneous Systems". *Pure & Applied Chemistry* **65**, 873-882.

Ruelle, D. 1969. *Statistical Mechanics: Rigorous Results*. Reading, MA: Addison-Wesley.

Rybicki, G.B. 1971. "Exact Statistical Mechanics of a One-Dimensional Self-Gravitating System" *Astrophysics and Space Science* 14, 56-72.

Saslaw. W. 2000. *The Distribution of the Galaxies*. Cambridge: Cambridge University Press.

Sota, Y., Iguchi, O., Morikawa, M., Tatekawa, T. and K. Maeda, 2001. *Physical Review* E 64, 056133.

Styer, D. 2004. "What Good is the Thermodynamic Limit?" *American Journal of Physics* 74, 25-29.

Touchette, H., Ellis, R.S., and B. Turkington, 2004. "An Introduction to the Thermodynamic and Macrostate Levels of Nonequivalent Ensembles" *Physica* A 340, 138-146. cond-mat/0404655

Touchette, H. 2002. "When is a Quantity Additive, and When is it Extensive?" *Physica A* 305, 84, cond-mat/0201134

Tsallis, C. 1988. "Possible Generalization of Boltzmann-Gibbs Statistics" *Journal of Statistical Physics* 52, 479-487.