Physics and chance

Philosophical issues in the foundations of statistical mechanics

LAWRENCE SKLAR
University of Michigan, Ann Arbor

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http: //www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http: //www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1993
First paperback edition 1995
Reprinted 1996, 1998

Typeset in Garamond

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing-in-Publication Data is available

ISBN 0-521-55881-6 paperback

Transferred to digital printing 2004

sudden change of the motion into one in which macroscopic *convection* of the fluid sets in. If one carefully controls the experiment, the convection takes the form of the creation of stable hexagonal convection cells that transmit heat from the hotter to the cooler plate by a stable, steady-state, flow of heated matter. More complex situations can be generated by using solutions of a variety of chemicals that can chemically combine with one another and dissociate chemically from one another. Here, by varying chemical concentrations, temperatures, and so on, one can generate cases of steady-state flow, of oscillatory behavior with repetitive order both in space and in time, or of bifurcation in which the system jumps into one or another of a variety of possible self-sustaining flows, and so on.

There is at least some speculative possibility that the existence of such "self-organizing" phenomena as those described may play an important role in biological phenomena (biological clocks as generated by oscillatory flows, spatial organization of an initially spatially homogeneous mass by random change into a then self-stabilizing spatially inhomogeneous organization, and so on).

II. Kinetic theory

1. Early kinetic theory

Just as the theory of heat as internal energy continued to be speculated on and espoused, even during the period in which the substantival-caloric theory dominated the scientific consensus, so throughout the caloric period there appeared numerous speculations about just what *kind* of internal motion constituted that energy that took the form of heat. Here, the particular theory of heat offered was plainly dependent upon one's conception of the micro-constitution of matter. Someone who held to a continuum account, taking matter as continuous even at the micro-level, might think of heat as a kind of oscillation or vibration of the matter. Even an advocate of discreteness – of the constitution of matter out of discrete atoms – would have a wide variety of choices, especially because the defenders of atomism were frequently enamored of complex models in which the atoms of matter were held in place relative to one another by surrounding clouds of aether or something of that sort.

As early as 1738, D. Bernoulli, in his *Hydrodynamics*, proposed the model of a gas as constituted of microscopic particles in rapid motion. Assuming their uniform velocity, he was able to derive the inverse relationship of pressure and volume at constant temperature. Furthermore, he reflected on the ability of increasing temperature to increase pressure at a constant volume (or density) of the gas, and indicated that the square of the velocity of the moving particles, all taken as having a common identical velocity, would be proportional to temperature. Yet the caloric theory remained dominant throughout the eighteenth century.

The unfortunate indifference of the scientific community to Bernoulli's work was compounded by the dismaying tragi-comedy of Herepath and Waterston in the nineteenth century. In 1820, W. Herepath submitted a paper to the Royal Society, again deriving the ideal gas laws from the model of independently moving particles of a fixed velocity. He identified heat with internal motion, but apparently took temperature as proportional to particle velocity instead of particle energy. He was able to offer qualitative accounts of numerous familiar phenomena by means of this model (such as change of state, diffusion, and the existence of sound waves). The Royal Society rejected the paper for publication, and although it appeared elsewhere, it had little influence. (J. Joule later read Herepath's work and in fact published a piece explaining and defending it in 1848, a piece that did succeed, to a degree, in stimulating interest in Herepath's work.)

J. Waterston published a number of his ideas in a similar vein in a book in 1843. The contents of the kinetic ideas were communicated to the Royal Society in 1845. The paper was judged "nothing but nonsense" by one referee, but it was read to the Society in 1846 (although not by Waterston, who was a civil servant in India), and an abstract was published in that year. Waterston gets the proportionality of temperature to square of velocity right, understands that in a gas that is a mixture of particles of different masses, the energy of each particle will still be the same, and even (although with mistakes) calculates on the model the ratio of specific heat at constant pressure to that at constant volume. The work was once again ignored by the scientific community.

Finally, in 1856, A. Krönig's paper stimulated interest in the kinetic theory, although the paper adds nothing to the previous work of Bernoulli, Herepath, Waterston, and Joule. Of major importance was the fact that Krönig's paper may have been the stimulus for the important papers of Clausius in 1857 and 1858. Clausius generalized from Krönig, who had idealized the motion of particles as all being along the axes of a box, by allowing any direction of motion for a particle. He also allowed, as Krönig and the others did not, for energy to be in the form of rotation of the molecular particles or in vibrational states of them, as well as in energy of translational motion. Even more important was his resolution of a puzzle with the kinetic theory. If one calculates the velocity to be expected of a particle, it is sufficiently high that one would expect particles released at one end of a room to be quickly found at the other. Yet the diffusion of one gas through another is much slower than one would expect on this basis. Clausius pointed out that the key to the solution was in molecular collisions, and introduced the notion of mean free path – the average distance a molecule could be expected to travel between one collision and another.

The growing receptiveness of the scientific community to the kinetic theory was founded in large part, of course, on both the convincing quantitative evidence of the interconvertibility of heat and overt mechanical energy with the conservation of their sum, and on the large body of evidence for the atomic constitution of matter coming from other areas of science (chemistry, electro-chemistry, and so on).

2. Maxwell

In 1860, J. Maxwell made several major contributions to kinetic theory. In this paper we find the first language of a sort that could be interpreted in a probabilistic or statistical vein. Here for the first time the nature of possible collisions between molecules is studied, and the notion of the probabilities of outcomes treated. (What such reference to probabilities might mean is something we will leave for Section II,5 of this chapter.) Although earlier theories generally operated on some assumption of uniformity with regard to the velocities of molecules, Maxwell for the first time takes up the question of just what kind of distribution of velocities of the molecules we ought to expect at equilibrium, and answers it by invoking assumptions of probabilistic nature.

Maxwell realizes that even if the speeds of all molecules were the same at one instant, this distribution would soon end, because in collision the molecules would "on average" not end up with identical speeds. He then asks what the distribution of speeds ought to be taken to be. The basic assumptions he needs to derive his result are that in collisions, all directions of impact are equally likely, and the additional posit that for any three directions at right angles to one another, the distribution law for the components of velocity will be identical. This is equivalent to the claim that the components in the y and z directions are "probabilistically independent" of the component in the x direction. From these assumptions he is able to show that "after a great number of collisions among a great number of identical particles," the "average number of particles whose velocities lie among given limits" will be given by the famous Maxwell law:

Number of molecules with velocities between v and $v + \mathrm{d}v$ = $Av^2 \exp(-v^2/b)\mathrm{d}v$

It is of historical interest that Maxwell may very well have been influenced by the recently published theory of errors of R. Adrian, K. Gauss, and A. Quetelet in being inspired to derive the law. Maxwell is also aware that his second assumption, needed to derive the law, is, as he puts it in an 1867 paper, "precarious," and that a more convincing derivation of the equilibrium velocity distribution would be welcome.

But the derivation of the equilibrium velocity distribution law is not Maxwell's only accomplishment in the 1860 paper. He also takes up the problem of transport. If numbers of molecules change their density from place to place, we will have transport of mass. But even if density stays constant, we can have transfer of energy from place to place by molecular collision, which is heat conduction, or transfer of momentum from place to place, which is viscosity. Making a number of "randomness" assumptions, Maxwell derives an expression for viscosity. His derivation contained flaws, however, and was later criticized by Clausius.

An improved theory of transport was presented by Maxwell in an 1866 paper. Here he offered a general theory of transport, a theory that once again relied upon "randomness" assumptions regarding the initial conditions of the interaction of molecules on one another. And he provided a detailed study of the result of any such molecular interaction. The resulting formula depends upon the nature of the potential governing molecular interaction, and on the relative velocities of the molecules, which, given non-equilibrium, have an unknown distribution. But for a particular choice of that potential – the so-called Maxwell potential, which is of inverse fifth power in the molecular separation – the relative velocities drop out and the resulting integrals are exactly solvable. Maxwell was able to show that the Maxwell distribution is one that will be stationary – that is, unchanging with time, and that this is so independently of the details of the force law among the molecules. A symmetry postulate on cycles of transfers of molecules from one velocity range to another allows him to argue that this distribution is the *unique* such stationary distribution. Here, then, we have a new rationale for the standard equilibrium distribution, less "precarious" than that offered in the 1860 paper.

The paper then applies the fundamental results just obtained to a variety of transport problems: heat conduction, viscosity, diffusion of one gas into another, and so on. The new theory allows one to calculate from basic micro-quantities the values of the "transport coefficients," numbers introduced into the familiar macroscopic equations of viscous flow, heat conduction, and so on by experimental determination. This allows for a comparison of the results of the new theory with observational data, although the difficulties encountered in calculating exact values in the theory, both mathematical and due to the need to make dubious assumptions about micro-features, and the difficulties in exact experimental determination of the relevant constants, make the comparison less definitive than one would wish.

3. *Boltzmann*

In 1868, L. Boltzmann published the first of his seminal contributions to kinetic theory. In this piece he generalizes the equilibrium distribution for velocity found by Maxwell to the case where the gas is subjected to an external potential, such as the gravitational field, and justifies the distribution by arguments paralleling those of Maxwell's 1866 paper.

In the second section of this paper he presents an alternative derivation of the equilibrium distribution, which, ignoring collisions and kinetics, resorts to a method reminiscent of Maxwell's first derivation. By assuming that the "probability" that a molecule is to be found in a region of space and that momentum is proportional to the "size" of that region, the usual results of equilibrium can once again be reconstructed.

In a crucially important paper of 1872, Boltzmann takes up the problem of non-equilibrium, the approach to equilibrium, and the "explanation" of the irreversible behavior described by the thermodynamic Second Law. The core of Boltzmann's approach lies in the notion of the distribution function $f(x,t)$ that specifies the density of particles in a given energy range. That is, $f(x,t)$ is taken as the number of particles between some specified value of the energy x and $x + dx$. He seeks a differential equation that will specify, given the structure of this function at any time, its rate of change at that time.

The distribution function will change because the molecules collide and exchange energy with one another. So the equation should have a term telling us how collisions effect the distribution of energy. To derive this, some assumptions are made that essentially restrict the equation to a particular constitution of the gas and situations of it. For example, the original equation deals with a gas that is, initially, spatially homogeneous. One can generalize out of this situation by letting f be a function of position as well as of energy and time. If one does so, one will need to supplement the collision term on the right-hand side of the equation by a "streaming" term that takes account of the fact that even without collisions the gas will have its distribution in space changed by the motion of the molecules unimpeded aside from reflection at the container walls. The original Boltzmann equation also assumes that the gas is sufficiently dilute, so that only interactions of two particles at a time with one another need be considered. Three and more particle collisions/interactions need not be taken into account. In Section III,6,1 I will note attempts at generalizing beyond this constraint.

In order to know how the energy distribution will change with time, we need to know how many molecules of one velocity will meet how many molecules of some other specified velocity (and at what angles) in any unit of time. The fundamental assumption Boltzmann makes here is the famous *Stosszahlansatz*, or Postulate with Regard to Numbers of Collisions. One assumes the absence of any "correlation" among molecules of given velocities, or, in other words, that collisions will be "totally random." At any time, then, the number of collisions of molecules of velocity v_1 and v_2 that meet will depend only on the proportion of molecules in the gas that have the respective velocities, the density of the gas, and the proportion of volume swept out by one of the molecules. This – along with an additional postulate that any collision is matched by a time-reverse collision in which the output molecules of the first collision would, if their directions were reversed, meet and generate as output molecules that have the same speed and reverse direction of the input molecules of collisions of the first kind (a postulate that can be somewhat weakened to allow "cycles" of collisions) – gives Boltzmann his famous kinetic equation:

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int d^3v \int d\Omega\sigma(\Omega)|v_1 - v_2|(f_2'f_1' - f_2f_1)$$

Here the equation is written in terms of velocity, rather than in terms of energy, as it was expressed in Boltzmann's paper. What this equation describes is the fraction of molecules with velocity v_1, f_1 changing over time. A molecule of velocity v_1 might meet a molecule of velocity v_2 and be knocked into some new velocity. On the other hand, molecules of velocities v_1' and v_2' will be such that in some collisions of them there will be the output of a molecule of velocity v_1. The number f_2 gives the fraction of molecules of velocity v_2, and the numbers f_1', and f_2' give the respective fractions for molecules of velocities v_1' and v_2'. The term $\sigma(\Omega)$ is determined by the nature of the molecular collisions, and rest of the apparatus on the right-hand side is designed to take account of all the possible ways in which the collisions can occur (because of the fact that molecules can collide with their velocities at various angles from one another).

The crucial assumption is that the rate of collisions in which a molecule of velocity v_1 meets one of velocity v_2, $f(v_1,v_2)$, is proportional to the product of the fraction of molecules having the respective velocities, so that it can be written as $f(v_1)f(v_2)$. The two terms, $f_2'f_1'$ and $-f_2f_1$, on the right-hand side of the equation can then be shown to characterize the number of collisions that drive molecules into a particular velocity range from another velocity range, and the number of those that delete molecules from that range into the other, the difference being the net positive change of numbers of molecules in the given range.

Introducing the Maxwell–Boltzmann equilibrium velocity distribution function into the equation immediately produces the result that it is stationary in time, duplicating Maxwell's earlier rationalization of this distribution by means of his transfer equations.

But how can we know if this standard equilibrium distribution is the *only* stationary solution to the equation? Knowing this is essential to justifying the claim that the discovery of the kinetic equation finally provides the micro-mechanical explanation for the fundamental fact of thermodynamics: the existence of a unique equilibrium state that will be ceaselessly and monotonically approached from any non-equilibrium state. It is to justifying the claim that the Maxwell–Boltzmann distribution is the unique stationary solution of the kinetic equation that Boltzmann turns.

To carry out the proof, Boltzmann introduces a quantity he calls E. The notation later changes to H, the standard notation, so we will call it that. The definition of H is arrived at by writing $f(x,t)$ as a function of velocity:

$$H = \int d^3v f(v,t) \log f(v,t)$$

Intuitively, H is a measure of how "spread out" the distribution in velocities of the molecules is. The logarithmic aspect of it has the virtue that the total spread-outness of two independent samples proves to be the sum of their individual spreads. Boltzmann is able to show this as long as $f(v,t)$ obeys the kinetic equation,

$$dH/dt \leq 0$$

and that $dH/dt = 0$ only when the distribution function has its equilibrium form. Here, then, is the needed proof that the equilibrium distribution is the uniquely stationary solution to the kinetic equation.

4. *Objections to kinetic theory*

The atomistic-mechanistic account of thermal phenomena posited by the kinetic theory received a hostile reception from a segment of the scientific community whose two most prominent members were E. Mach and P. Duhem. Their objection to the theory was the result of two programmatic themes, distinct themes whose difference was not always clearly recognized by their exponents.

One theme was a general phenomenalistic-instrumentalistic approach to science. From this point of view, the purpose of science is the production of simple, compact, lawlike generalizations that summarize the fundamental regularities among items of observable experience. This view of theories was skeptical of the postulation of unobservable "hidden" entities in general, and so, not surprisingly, was skeptical of the postulation of molecules and their motion as the hidden ground of the familiar phenomenological laws of thermodynamics.

The other theme was a rejection of the demand, common especially among English Newtonians, that all phenomena ultimately receive their explanation within the framework of the *mechanical* picture of the world. Here the argument was that the discovery of optical, thermal, electric, and magnetic phenomena showed us that mechanics was the appropriate scientific treatment for only a portion of the world's phenomena. From this point of view, kinetic theory was a misguided attempt to assimilate the distinctive theory of heat to a universal mechanical model.

There was certainly confusion in the view that a phenomenalistic-instrumentalistic approach to theories required in any way the rejection of atomism, which is, after all, a theory that can be given a phenomenalistic-instrumentalistic philosophical reading if one is so inclined. Furthermore, from the standpoint of hindsight we can see that the anti-mechanistic stance was an impediment to scientific progress where the theory of heat was concerned. It is only fair to note, however, that the anti-mechanist rejection of any attempt to found electromagnetic theory upon some mechanical model of motion in the aether did indeed turn out to be the route justified by later scientific developments.

More important, from our point of view, than these philosophical-methodological objections to the kinetic theory were specific technical objections to the consistency of the theory's basic postulates with the mechanical theory of atomic motion that underlay the theory. The first difficulty for kinetic theory, a difficulty in particular for its account of the irreversibility of thermal phenomena, seems to have been initially noted by Maxwell himself in correspondence and by W. Thomson in publication in 1874. The problem came to Boltzmann's attention through a point made by J. Loschmidt in 1876–77 both in publication and in discussion with Boltzmann. This is the so-called *Umkehreinwand*, or Reversibility Objection.

Boltzmann's *H*-Theorem seems to say that a gas started in any non-equilibrium velocity distribution must monotonically move closer and closer to equilibrium. Once in equilibrium, a gas must stay there. But imagine the micro-state of a gas that has reached equilibrium from some non-equilibrium state, the gas energetically isolated from the surrounding environment during the entire process. The laws of mechanics guarantee to us that a gas whose micro-state consists of one just like the equilibrium gas – except that the direction of motion of each constituent molecule is reversed – will trace a path through micro-states that are each the "reverse" of those traced by the first gas in its motion toward equilibrium. But because H is indifferent to the direction of motion of the molecules and depends only upon the distribution of their speeds, this

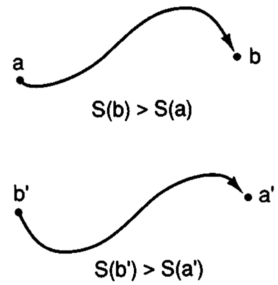

Figure 2-1. Loschmidt's reversibility argument. Let a system be started in micro-state a and evolve to micro-state b. Suppose, as is expected, the entropy of state b, S(b) is higher than that of state a, S(a). Then, given the time-reversal invariance of the underlying dynamical laws that govern the evolution of the system, there must be a micro-state b′, that evolves to a micro-state a′ and such that the entropy of b′, S(b′), equals that of b and the entropy of a′ equals that of a, S(a′), (as Boltzmann defines statistical entropy). So for each "thermodynamic" evolution in which entropy increases, there must be a corresponding "anti-thermodynamic" evolution possible in which entropy decreases.

means that the second gas will evolve, monotonically, *away* from its equilibrium state. Therefore, Boltzmann's *H*-theorem is incompatible with the laws of the underlying micro-mechanics. (See Figure 2-1.)

A second fundamental objection to Boltzmann's alleged demonstration of irreversibility only arose some time after Maxwell and Boltzmann had both offered their "reinterpretation" of the kinetic theory to overcome the Reversibility Objection. In 1889, H. Poincaré proved a fundamental theorem on the stability of motion that is governed by the laws of Newtonian mechanics. The theorem only applied to a system whose energy is constant and the motion of whose constituents is spatially bounded. But a system of molecules in a box that is energetically isolated from its environment fits Poincaré's conditions. Let the system be started at a given time in a particular mechanical state. Then, except for a "vanishingly small" number of initial states (we shall say what this means in Section 3,I,3), the system will eventually evolve in such a way as to return to states as close to the initial state as one specifies. Indeed, it will return to an arbitrary degree of closeness an unbounded number of times. (See Figure 2-2.)

In 1896, E. Zermelo applied the theorem to generate the *Wiederkehreinwand*, or Recurrence Objection, to Boltzmann's mechanically derived *H*-Theorem. The *H*-Theorem seems to say that a system started in non-equilibrium state must monotonically approach equilibrium. But, according to Poincaré's Recurrence Theorem, such a system, started in non-equilibrium, if it does get closer to equilibrium, must at

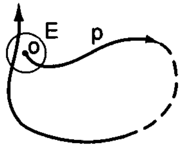

Figure 2-2. Poincaré recurrence. We work in phase-space where a single point represents the exact microscopic state of a system at a given time – say the position and velocity of every molecule in a gas. Poincaré shows for certain systems, such as a gas confined in a box and energetically isolated from the outside world, that if the system starts in a certain microscopic state o, then, except for a "vanishingly small" number of such initial states, when the system's evolution is followed out along a curve p, the system will be found, for any small region E of micro-states around the original micro-state o to return to a micro-state in that small region E. Thus, "almost all" such systems started in a given state will eventually return to a microscopic state "very close" to that initial state.

some point get back to a state mechanically as close to its initial state as one likes. But such a state would have a value of *H* as close to the initial value as one likes as well. Hence Boltzmann's demonstration of necessary monotonic approach to equilibrium is incompatible with the fundamental mechanical laws of molecular motion.

5. The probabilistic interpretation of the theory

The result of the criticisms launched against the theory, as well as of Maxwell's own critical examination of it, was the development by Maxwell, Boltzmann, and others of the probabilistic version of the theory. Was this a revision of the original theory or merely an explication of what Clausius, Maxwell, and Boltzmann had meant all along? It isn't clear that the question has any definitive answer. Suffice it to say that the discovery of the Reversibility and Recurrence Objections prompted the discoverers of the theory to present their results in an enlightening way that revealed more clearly what was going on than did the original presentation of the theory.

As we relate what Maxwell, Boltzmann, and others said, the reader will find himself quite often puzzled as to just how to understand what they meant. The language here becomes fraught with ambiguity and conceptual obscurity. But it is not my purpose here either to lay out all the possible things they might have meant, or to decide just which of the many understandings of their words we ought to attribute to them. Again, I doubt if there is any definitive answer to those questions. We shall be exploring a variety of possible meanings in detail in Chapters 5, 6, and 7.

Throughout this section it is important to keep in mind that what was primarily at stake here was the attempt to show that the apparent contradiction of the kinetic theory with underlying micro-mechanics could be avoided. That is not the same thing at all as showing that the theory is correct, nor of explaining why it is correct. We will see here, however, how some of the fundamental problems of rationalizing belief in the theory and of offering an account as to why it is correct received their early formulations.

Maxwell's probabilism. In a train of thought beginning around 1867, Maxwell contemplated the degree to which the irreversibility expressed by the Second Law is inviolable. From the new kinetic point of view, the flow of heat from hot to cold is only the mixing of molecules faster on the average with those slower on the average. Consider a Demon capable of seeing molecules individually approaching a hole in a partition and capable of opening and closing the hole with a door, his choice depending on the velocity of the approaching molecule. Such an imagined creature could sort the molecules into fast on the right and slow on the left, thereby sorting a gas originally at a common temperature on both sides into a compartment of hot gas and a compartment of cold gas. And doing this would not require overt mechanical work, or at least not the amount of this demanded by the usual Second Law considerations. From this and related arguments, Maxwell concludes that the Second Law has "only a statistical certainty."

Whether a Maxwell Demon could really exist, even in principle, became in later years a subject of much discussion. L. Brillouin and L. Szilard offered arguments designed to show that the Demon would generate more entropy in identifying the correct particles to pass through and the correct particles to block than would be reduced by the sorting process, thereby saving the Second Law from the Demon's subversion. Later, arguments were offered to show that Demon-like constructions could avoid that kind of entropic increase as the result of the Demon's process of knowledge accrual.

More recently, another attack had been launched on the very possibility of an "in principle" Maxwell Demon. In these works it is argued that after each molecule has been sorted, the Demon must reset itself. The idea is that the Demon, in order to carry out its sorting act, must first register in a memory the fact that it is one sort of particle or the other with which it is dealing. After dealing with this particle, the Demon must "erase" its memory in order to have a blank memory space available to record the status of the next particle encountered. R. Landauer and others have argued that this "erasure" process is one in which entropy is generated by the Demon and fed into its environment. It is this entropy generation, they argue, that more than compensates for the entropy reduction accomplished by the single act of sorting.

In his later work, Maxwell frequently claims that the irreversibility captured by the Second Law is only "statistically true" or "true on the average." At the same time he usually seems to speak as though the notions of randomness and irregularity he invokes to explain this are only due to limitations on our knowledge of the exact trajectories of the "in principle" perfectly deterministic, molecular motions. Later popular writings, however, do speak, if vaguely, in terms of some kind of underlying "objective" indeterminism.

Boltzmann's probabilism. Stimulated originally by his discussions with Loschmidt, Boltzmann began a process of rethinking of his and Maxwell's results on the nature of equilibrium and of his views on the nature of the process that drives systems to the equilibrium state. Various probabilistic and statistical notions were introduced without it being always completely clear what these notions meant. Ultimately, a radically new and curious picture of the irreversible approach of systems (and of "the world") toward equilibrium emerged in Boltzmann's writings.

One paper of 1877 replied specifically to Loschmidt's version of the Reversibility Objection. How can the *H*-Theorem be understood in light of the clear truth of the time reversibility of the underlying micromechanics?

First, Boltzmann admits, it must be clear that the evolution of a system from a given micro-state will depend upon the specific micro-state that serves to fix the initial conditions that must be introduced into the equations of dynamical evolution to determine the evolution of the system. Must we then, in order to derive the kinetic equation underlying the Second Law of Thermodynamics, posit the existence of specific, special initial conditions for all gases? Boltzmann argues that we can avoid this by taking the statistical viewpoint. It is certainly true that every individual micro-state has the same probability. But there are vastly more micro-states corresponding to the macroscopic conditions of the system being in (or very near) equilibrium than there are numbers of micro-states corresponding to non-equilibrium conditions of the system. If we choose initial conditions at random, then, given a specified time interval, there will be many more of the randomly chosen initial states that lead to a uniform, equilibrium, micro-state at the later time than there will be initial states that lead to a non-equilibrium state at the later time. It is worth noting that arguments in a similar vein had already appeared in a paper of Thomson's published in 1874.

In his 1877 paper, Boltzmann remarks that "one could even calculate, from the relative numbers of different state distributions, their probabilities, which might lead to an interesting method for the calculation of thermal equilibrium." He develops this idea in another paper also published in 1877.

Here the method familiar to generations of students of elementary kinetic theory is introduced. One divides the available energy up into small finite intervals. One imagines the molecules distributed with so-and-so many molecules in each energy range. A weighing factor is introduced that converts the problem, instead, into imagining the *momentum* divided up into the equal small ranges. One then considers all of the ways in which molecules can be placed in the momentum boxes, always keeping the number of molecules and the total energy constant. Now consider a state defined by a distribution, a specification of the number of molecules in each momentum box. For a large number of particles and boxes, one such distribution will be obtained by a vastly larger number of assignments of molecules to boxes than will any other such distribution. Call the probability of a distribution the number of ways it can be obtained by assignments of molecules to boxes. Then one distribution is the overwhelmingly most probable distribution. Let the number of boxes go to infinity and the size of the boxes go to zero and one discovers that the energy distribution among the molecules corresponding to this overwhelmingly probable distribution is the familiar Maxwell–Boltzmann equilibrium distribution. (See Figure 2-3.)

It is clear that Boltzmann's method for calculating the equilibrium distribution here is something of a return to Maxwell's first method and away from the approach that takes equilibrium to be specified as the unique stationary solution of the kinetic equation. As such it shares "precariousness" with Maxwell's original argument. But more has been learned by this time. It is clear to Boltzmann, for example, that one must put the molecules into equal momentum boxes, and not energy boxes as one might expect, in order to calculate the probability of a state. His awareness of this stems from considerations of collisions and dynamics that tell us that it is only the former method that will lead to stationary distributions and not the latter. And, as we shall see in the next section, Boltzmann is also aware of other considerations that associate probability with dynamics in a non-arbitrary way, considerations that only become fully developed shortly after the second 1877 paper appeared.

Combining the definition of H introduced in the paper on the kinetic equation, the calculated monotonic decrease of H implied by that equation, the role of entropy, S, in thermodynamics (suggesting that S in some sense is to be associated with $-H$), and the new notion of probability of a state, W, Boltzmann writes for the first time the equation that subsequently became familiar as $S = -K \log W$. The entropy of a macrostate is determined simply by the number of ways in which the macrostate can be obtained by arrangements of the constituent molecules of the system. As it stands, much needs to be done, however, to make this "definition" of entropy fully coherent.

One problem – by no means the only one – with this new way of viewing things is the use of "probability." Botzmann is not oblivious to

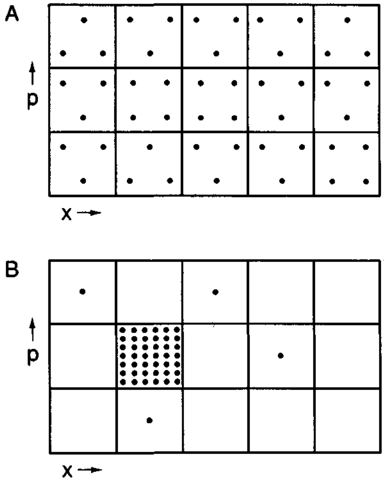

Figure 2-3. Boltzmann entropy. Box A (and Box B) represent all possible values of position x and momentum p a molecule of gas might have. This "molecular phase space" is divided up into small sub-boxes. In the theory, these sub-boxes are of small size relative to the size of the entire molecular phase space but large enough so that many molecules will generally be in each box for any reasonable micro-state of the gas. An arrangement of molecules into boxes like that of Fig. A can be obtained by many permutations of distinct molecules into boxes. But an arrangement like that of Fig. B can be obtained only by a much smaller number of permutations. So the Boltzmann entropy for arrangement A is much higher than that for arrangement B. The equilibrium momentum distribution (the Maxwell–Boltzmann distribution) corresponds to the arrangement that is obtainable by the maximum number of permutations, subject to the constraints that total number of molecules and total energy of molecules remain constant.

the ambiguities latent in using that term. As early as 1881 he distinguished between "probability" as he means it, taking that to be "the time during which the system possesses this condition on the average," and as he takes Maxwell to mean it as "the ratio of the number of [innumerable similarly constituted systems] which are in that condition to the total number of systems." This is an issue to which we will return again and again.

Over the years, Boltzmann's account of irreversibility continues to evolve, partly inspired by the need to respond to critical discussion of his and Maxwell's ideas, especially in England, and partly due to his own ruminations on the subject. To what extent can one stand by the new, statistical reading of the *H*-Theorem, now taken to be read as describing the "overwhelmingly probable" course of evolution of a system from "less probable" states to "more probable" states? Once again, we are using states here not to mean micro-states, all of which are taken as equally probable, but states defined by numbers of particles in a given momentum range.

Most disturbing to this view is a problem posed by E. Culverwell in 1890. Boltzmann's new statistical interpretation of the *H*-Theorem seems to tell us that we ought to consider transitions from micro-states corresponding to a non-equilibrium macro-condition to micro-states corresponding to a condition closer to equilibrium as more "probable" than transitions of the reverse kind. But if, as Boltzmann would have us believe, all micro-states have equal probability, this seems impossible. For given any pair of micro-states, S_1, S_2, such that S_1 evolves to S_2 after a certain time interval, there will be a pair S_2', S_1' – the states obtained by reversing the directions of motion in the respective original micro-states while keeping speeds and positions constant – such that S_2' is closer to equilibrium than S_1', and yet S_2' evolves to S_1' over the same time interval. So "anti-kinetic" transitions should be as probable as "kinetic" transitions.

Some of the discussants went on to examine how irreversibility was introduced into the kinetic equation in the first place. Others suggested that the true account of irreversibility would require some kind of "friction" in the system, either in form of energy of motion of the molecules interchanging with the aether or with the external environment.

In a letter of 1895, Boltzmann gave his view of the matter. This required, once again, a reinterpretation of the meaning of his kinetic equation and of its *H*-Theorem, and the introduction of several new hypotheses as well. These latter hypotheses, as the reader will discern, are of an unexpected kind, and, perhaps, unique in their nature in the history of science.

In this new picture, Boltzmann gives up the idea that the value of *H* will *ever* monotonically decrease from an initial micro-state. But it is still true that a system, in an initial improbable micro-state, will probably be found at any later time in a micro-state closer to equilibrium. As Culverwell himself points out, commenting on Boltzmann's letter, the trick is to realize that if we examine the system over a vast amount of time, the system will nearly always be in a close-to-equilibrium state. Although it is true that there will be as many excursions from a close-to-equilibrium micro-state to a state further from equilibrium as there will be of the reverse kind of transition, it is still true that *given* a micro-state far from equilibrium as the starting point, at a later time we ought to expect the system to be in a micro-state closer to equilibrium. (See Figure 2-4.)

But the theory of evolution under collision is now *time-symmetric*! For it will also be true that given a micro-state far from equilibrium at one time, we ought, on probabilistic grounds, to expect it to have been closer to equilibrium at any given past time. The theory is also paradoxical in

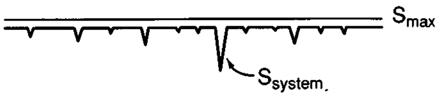

Figure 2-4. Time-symmetric Boltzmann picture. In this picture of the world, it is proposed that an isolated system "over infinite time" spend nearly all the time in states whose entropy S_{system} is close to the maximum value S_{max} – that is, in the equilibrium state. There are random fluctuations of the system away from equilibrium. The greater the fluctuation of a system from equilibrium, the less frequently it occurs. The picture is symmetrical in time. If we find a system far from equilibrium, we ought to expect that in the future it will be closer to equilibrium. But we ought also to infer that in the past it was also closer to equilibrium.

that it tells us that equilibrium is overwhelmingly probable. Isn't this a curious conclusion to come to in a world that we find to be grossly distant from equilibrium?

Boltzmann takes up the latter paradox in his 1895 letter. He attributes to his assistant, Dr. Schuetz, the idea that the universe is an eternal system that is, overall, in equilibrium. "Small" regions of it, for "short" intervals of time will, improbably, be found in a state far from equilibrium. Perhaps the region of the cosmos observationally available to us is just such a rare fluctuation from the overwhelmingly probable equilibrium state that pervades the universe as a whole.

It is in 1896 that Zermelo's application of the Poincaré Recurrence Theorem is now invoked to cast doubt on the kinetic explanation of irreversibility and Boltzmann responds to two short papers of Zermelo's with two short pieces of his own. Boltzmann's 1896 paper points out that the picture adopted in the 1895 letter of a system "almost always" near equilibrium but fluctuating arbitrarily far from it, each kind of fluctuation being the rarer the further it takes the system from equilibrium, is perfectly consistent with the Poincaré Recurrence Theorem.

The 1897 paper repeats the picture of the 1896 paper, but adds the cosmological hypothesis of Dr. Schuetz to it. In this paper, Boltzmann makes two other important suggestions. If the universe is mostly in equilibrium, why do we find ourselves in a rare far-from-equilibrium portion? The suggestion made is the first appearance of a now familiar "transcendental" argument: Non-equilibrium is essential for the existence of a sentient creature. Therefore a sentient creature could not find itself existing in an equilibrium region, probable as this may be, for in such regions no sentience can exist.

Even more important is Boltzmann's answer to the following obvious question: If the picture presented is now time-symmetrical, with every piece of the path that represents the history of the system and that slopes toward equilibrium matched by a piece sloping away from equilibrium to non-equilibrium, then why do we find ourselves in a portion of the universe in which systems approach equilibrium from past to future? Wouldn't a world in which systems move from equilibrium to non-equilibrium support sentience equally well? Here the response is one already suggested by a phenomenological opponent of kinetic theory, Mach, in 1889. What we *mean* by the future direction of time is the direction of time in which our local region of the world is headed toward equilibrium. There could very well be regions of the universe in which entropic increase was counter-directed, so that one region had its entropy increase in the direction of time in which the other region was moving away from equilibrium. The inhabitants of those two regions would each call the direction of time in which the entropy of their local region was increasing the "future" direction of time! The combination of cosmological speculation, transcendental deduction, and definitional dissolution in these short remarks has been credited by many as one of the most ingenious proposals in the history of science, and disparaged by others as the last, patently desperate, ad hoc attempt to save an obviously failed theory. We shall explore the issue in detail in Chapters 8 and 9, even if we will not settle on which conclusion is correct.

6. *The origins of the ensemble approach and of ergodic theory*

There is another thread that runs through the work of Maxwell and Boltzmann that we ought to follow up. As early as 1871, Boltzmann describes a mechanical system, a particle driven by a potential of the form $\frac{1}{2}(ax^2 + by^2)$, where a/b is irrational, where the path of the point in phase space that represents the motion of the particle will "go through the entire surface" of the phase space to which the particle is confined by its constant total energy. Here, by phase space we mean that abstract multiple dimensional space each point of which specifies the total position-momentum state of the system at any time. Boltzmann suggests that the motion of the point representing a system of interacting molecules in a gas, especially if the gas is acted upon by forces from the outside, will display this same behavior:

> The great irregularity of thermal motion, and the multiplicity of forces that act on the body from the outside, make it probable that the atoms themselves, by virtue of the motion that we call heat, pass through all possible positions and velocities consistent with the equation of kinetic energy. (See Figure 2-5.)

Given the truth of this claim, one can then derive such equilibrium features as the equipartition of energy over all available degrees of freedom in a simple way. Identify the equilibrium value of a quantity with

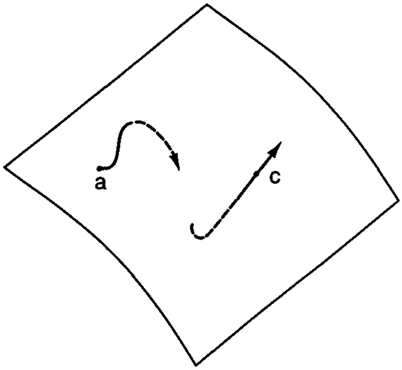

Figure 2-5. The Ergodic Hypothesis. Let a system be started in any microstate, represented by point a in the phase space. Let b represent any other microstate possible for the system. The Ergodic Hypothesis posits that at some future time or other, the system started in state a will eventually pass through state b as well. But the posit is in fact provable false.

its average over an infinite period of time, in accordance with Boltzmann's general approach of thinking of the "probability" of the system being in a given phase as the proportion of time the system spends in that phase over vast periods of time. If the system does indeed pass through *every* phase in its evolution, then it is easy to calculate such infinite time averages by simply averaging over the value of the quantity in question for each phase point, weighting regions of phase points according to a measure that can easily be derived. (We will see the details Chapter 5.) Here we see Boltzmann's attempt to eliminate the seeming arbitrariness of the probabilistic hypotheses used earlier to derive equilibrium features.

Maxwell, in a very important paper of 1879, introduces a new method of calculating equilibrium properties that, he argues, will give a more general derivation of an important result earlier obtained by Boltzmann for special cases. The methods of Boltzmann's earlier paper allowed one to show, in the case of molecules that interact only upon collision, that in equilibrium the equipartition property holds. This means that the total kinetic energy available to the molecules of the gas will be distributed in equal amounts among the "degrees of freedom" of motion available to the molecules. Thus, in the case of simple point molecules all of whose energy of motion is translational, the x, y, and z components of velocity of all the molecules will represent, at equilibrium, the same proportion of the total kinetic energy available to the molecules. Maxwell proposes a method of calculating features of equilibrium from which the equipartition result can be obtained and that is independent of any assumption about the details of interaction of the molecules. It will apply even if they exert force effects upon one another at long range due to potential interaction.

Suppose we imagine an infinite number of systems, each compatible with the macroscopic constraints imposed on a given system, but having every possible micro-state compatible with those macroscopic constraints. We can characterize a collection of possible micro-states at a time by a region in the phase-space of points, each point corresponding to a given possible micro-state. If we place a probability distribution over those points at a time, we can then speak of "the probability that a member of the collection of systems has its micro-state in a specified region at time t." With such a probability distribution, we can calculate the average value, over the collection, of any quantity that is a function of the phase value (such as the kinetic energy for a specified degree of freedom). The dynamic equations will result in each member of this collection or ensemble having its micro-state evolve, corresponding to a path among the points in phase-space. In general, this dynamic evolution will result in the probability that is assigned to a region of phase points changing with time, as systems have their phases move into and out of that region.

There is one distribution for a given time, however, that is such that the probability assigned to a region of phase points at a time will not vary as time goes on, because the initial probability assigned at the initial time to any collection of points that eventually evolves into the given collection will be the same as the probability assigned the given collection at the initial time. So an average value of a phase-function computed with this probability distribution will remain constant in time. If we identify equilibrium values with average values over the phase points, then for this special probability assignment, the averages, hence the attributed equilibrium values, will remain constant. This is as we would wish, because equilibrium quantities are constant in time. This special probability assignment is such that the average value of the energy per degree of freedom is the same for each degree of freedom, so that our identification results in derivation of the equipartition theorem for equilibrium that is dependent only upon the fundamental dynamical laws, our choice of probability distribution, and our identification of equilibrium values with averages over the ensemble. But the result is independent of any particular force law for the interaction of the molecules. (See Figure 2-6.)

Maxwell points out very clearly that although he can show that the distribution he describes is one that will lead to constant probabilities being assigned to a specified region of the phase points, he cannot show, from dynamics alone, that it is the *only* such stationary distribution. One additional assumption easily allows that to be shown. The assumption is that "a system if left to itself . . . will, sooner or later, pass

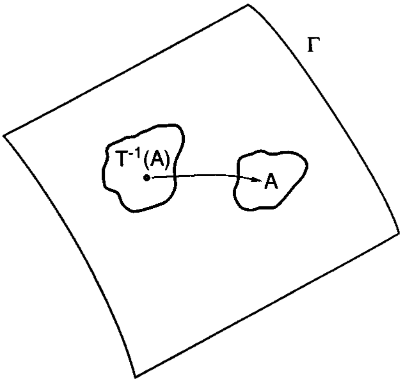

Figure 2-6. Invariant probability distribution. The space Γ represents all possible total micro-states of the system, each represented by a point of the space. A probability distribution is assigned over the space. In a time Δt, the systems whose points were originally in T^{-1} (A) have evolved in such a way that their phase points have moved to region A. Only if, for each "measurable" A, the total probability assigned to T^{-1} (A) is equal to that of A will a specified region of the phase space have a constant probability assigned to it as the systems evolve.

through every phase consistent with the energy." (How all this works we will see in detail in Chapter 5.)

Furthermore, Maxwell asserts, the encounters of the system of articles with the walls of the box will "introduce a disturbance into the motion of the system, so that it will pass from one undisturbed path to another." He continues:

It is difficult in a case of such extreme complexity to arrive at a thoroughly satisfactory conclusion, but we may with considerable confidence assert that except for particular forms of the surface of the fixed obstacle, the system will sooner or later, after a sufficient number of encounters pass through every phase consistent with the equation of energy.

Here we have introduced the "ensemble" approach to statistical mechanics, considering infinite collections of systems all compatible with the macroscopic constraints but having their micro-states take on every possible value. And we have the identification of equilibrium quantities with averages over this ensemble relative to a specified probability assigned to any collection of micro-states at a given time. We also have another one of the beginnings of the ergodic theory, that attempt to rationalize, on the basis of the constitution of the system and its dynamical laws, the legitimacy of the choice of one particular probability distribution over the phases as the right one to use in calculating average values.

In the 1884 paper, Boltzmann takes up the problem of the calculation of equilibrium values (a paper in which the term "ergodic" appears for the first time). Here he studies the differences between systems that will stay confined to closed orbits in the available region of phase space, and those that, like the hypothesized behavior of the swarm of molecules in a gas, will be such that they will wander all over the energetically available phase space, "passing through all phase points compatible with a given energy." In 1887 he utilizes the Maxwellian concept of an ensemble of macroscopically similar systems whose micro-states at a time take on every realizable possibility and the Maxwellian notion of a "stationary" probability distribution over such a micro-state.

The justifiable roles of (1) collections of macroscopically similar systems whose micro-states take on every realizable value, of (2) probability distributions over such collections, of stationary such probability distributions, of (3) the identification of equilibrium values with averages of quantities that are functions of the micro-state according to such probability measures, and of (4) the postulates that rationalize such a model by means of some hypothesis about the "wandering of a system through the whole of phase space allowed by energy" become, as we shall see, a set of issues that continue to plague the foundations of the theory.

III. Gibbs' statistical mechanics

1. Gibbs' ensemble approach

J. Gibbs, in 1901, presented, in a single book of extraordinary compactness and elegance, an approach to the problems we have been discussing that although taking off from the ensemble ideas of Boltzmann and Maxwell, presents them in a rather different way and generalizes them in an ingenious fashion.

Gibbs emphasizes the value of the methods of calculation of equilibrium quantities from stationary probability distributions reviewed in the last section. He looks favorably on the ability of this approach to derive the thermodynamic relations from the fundamental dynamical laws without making dubious assumptions about the details of the inter-molecular forces. He is skeptical that at the time, enough is known about the detailed constitution of gases out of molecules to rely on hypotheses about this constitution, and his skepticism is increased by results, well known at the time, that seem in fact to refute either the kinetic theory or the standard molecular models. In particular, the equipartition theorem for energy gives the wrong results even for the simple case of diatomic molecules. Degrees of freedom that ought to have their fair share of the energy at equilibrium seem to be totally ignored in the sharing out of energy that actually goes on. This is a problem not resolved until the underlying classical dynamics of the molecules is replaced by quantum mechanics.

In Gibbs' method we consider an infinite number of systems all having their micro-states described by a set of generalized positions and momenta. As an example, a monatomic gas with n molecules can be described by $6n$ such coordinates, 3 position and 3 momentum coordinates each for each point molecule. In a space whose dimensionality is the number of these positions and momenta taken together, a point represents a single possible total micro-state of one of the systems. Given such a micro-state at one time, the future evolution of the system having that micro-state is given by a path from this representative point, and the future evolution of the ensemble of systems can be viewed as a flow of points from those representing the systems at one time to those dynamically determined to represent the systems at a later time.

Suppose we assign a fraction of all the systems at a given time to a particular collection of phase points at that time. Or, assign to each region of phase points the probability that a system has its phase in that region at the specified time. In general, the probability assigned to a region will change as time goes on, as systems have their phase points enter and leave the region in different numbers. But some assignments of probability will leave the probability assigned a region of phases constant in time under the dynamic evolution. What are these assignments?

Suppose we measure the size of a region of phase points in the most natural way possible, as the "product" of its position and momentum sizes, technically the integral over the region of the product of the differentials. Consider a region at time t_0 of a certain size, measured in this way. Let the time go to t_1 and see where the flow takes each system whose phase point at t_0 was a boundary point of this region. Consider the new region at t_1 bounded by the points that represent the new phases of the "boundary systems." It is easy to prove that this new region is equal in "size" to the old.

Suppose we calculated the probability that a system is in a region at a time by using a function of the phases, $P(q,p)$, a probability density, such that the probability the system is in region A at time t is

$$\int \ldots \int_A P(q,p)\, dq_1 \ldots dqndp_1 \ldots dpn$$

What must $P(q,p)$ be like, so that the probability assigned to region A is invariant under dynamic evolution? We consider first the case where the systems have any possible value of their internal energy. The requirement on $P(q,p)$ that satisfies the demand for "statistical equilibrium" – that is, for unchanging probability values to be attributed to regions of the phase space – is simply that P should be a function of the q's and p's that stays constant along the path that evolves from any given phase point. If we deal with systems of constant energy, then any function P that is constant for any set of systems of the same energy will do the trick.

Gibbs suggests one particular such P function as being particularly noteworthy:

$$P = \exp(\psi - \varepsilon/\theta)$$

where θ and ψ are constants and ε is the energy of the system. When the ensemble of systems is so distributed in probability, Gibbs calls it canonically distributed. When first presented, this distribution is contrasted with others that are formally unsatisfactory (the "sum" of the probabilities diverges), but is otherwise presented as a posit. The justification for this special choice of form comes later.

Now consider, instead of a collection of systems each of which may possess any specific energy, a collection all of whose members share a common energy. The phase point for each of these systems at any time is confined to a sub-space of the original phase space of one dimension less than that of the full phase space. Call this sub-space, by analogy with surfaces in three-dimensional space, the energy surface. The evolution of each system in this new ensemble is now represented by a path from a point on this surface that is confined to the surface.

Given a distribution of such points on the energy surface, how can they be distributed in such a way that the probability assigned to any region on the surface at a time is always equal to the probability assigned to that region at any other time? Once again the answer is to assign probability in such a way that any region ever transformed by dynamical evolution into another region is assigned the same probability as that latter region at the initial time. Such a probability assignment had already been noted by Boltzmann and Maxwell. It amounts to assigning equal probabilities to the regions of equal volume between the energy surface and a nearby energy surface, and then assigning probabilities to areas on the surface in the ratios of the probabilities of the "pill box" regions between nearby surfaces they bound. Gibbs calls such a probability distribution on an energy surface the micro-canonical ensemble.

There is a third Gibbsian ensemble – the grand canonical ensemble – appropriate to the treatment of systems whose numbers of molecules of a given kind are not constant, but we shall confine our attention to ensembles of the first two kinds.

2. The thermodynamic analogies

From the features of the canonical and micro-canonical ensembles we will derive various equilibrium relations. This will require some association of quantities calculable from the ensemble with the familiar thermodynamic quantities such as volume (or other extensive magnitudes), pressure (or other conjugate "forces"), temperature, and entropy. Gibbs is quite cautious in offering any kind of physical rationale for the associations he makes. He talks of "thermodynamic analogies" throughout his work, maintaining only that there is a clear formal analogy between the quantities he is able to derive from ensembles and the thermodynamic variables. He avoids, as much as he can, both the Maxwell–Boltzmann attempt to connect the macro-features with specific constitution of the system out of its micro-parts, and avoids as well their attempt to somehow rationalize or explain why the ensemble quantities serve to calculate the thermodynamic quantities as well as they do.

He begins with the canonical ensemble. Let two ensembles, each canonically distributed, be compared to an ensemble that represents a system generated by the small energetic interaction of the two systems represented by the original ensembles. The resulting distribution will be stationary only if the θ's of the two original ensembles are equal, giving an analogy of θ to temperature, because systems when connected energetically stay in their initial equilibrium only if their temperatures are equal.

The next analogy is derived by imagining the energy of the system, which also functions in the specification of the canonical distribution, to be determined by an adjustable external parameter. If we imagine every system in the ensemble to have the same value of such an energy fixing parameter, and ask how the canonical distribution changes for a small change in the value of that parameter, always assuming the distribution to remain canonical, we get a relation that looks like this:

$$d\varepsilon = -\theta d\overline{\eta} - \overline{A}_1 da_1 \ldots - \overline{A}_n da_{\mathrm{n}}$$

where $\eta = \psi - \varepsilon/\theta$, the a_i's are the adjustable parameters, and the A_1's are given by $A = -d\varepsilon/da_{\mathrm{i}}$. A bar over a quantity indicates that we are taking its average value over the ensemble. If we compare this with the thermodynamic,

$$d\varepsilon = TdS - A_1 da_1 - \ldots - A_n da_{\mathrm{n}}$$

we get the analogy of θ as analogous to temperature, and $-\overline{\eta}$ as analogous to entropy (where the $A_{\mathrm{i}}da_{\mathrm{i}}$ terms are such terms as PdV, and so on).

Next, Gibbs takes up the problem of how systems will be distributed in their energies if they are represented by a canonical ensemble. The main conclusion is that if we are dealing with systems with a very large number of degrees of freedom, such as a system of gas molecules, then we will find the greatest probability of a system having a given energy centered closely around a mean energetic value.

This leads to the comparison of the canonical with the micro-canonical ensemble, and eventually to a picture of the physical situations that the two ensembles can plausibly be taken to represent. The conclusion Gibbs comes to is that a canonical ensemble, with its constant, θ, analogous to temperature, best represents a collection of identically structured systems all in perfect energetic contact with an infinite heat bath at a constant temperature. Here one would expect to find the systems closely clustered (if they have many degrees of freedom) about a central value of their energy, but with systems existing at all possible energies distributed around this central value in the manner described by the canonical distribution.

The micro-canonical ensemble, on the other hand, seems to be the appropriate representative of a collection of systems each perfectly isolated energetically from the external world. In the former case, the thermodynamic result of each system at a constant energy when in contact with the heat bath is replaced by the idea of the systems having varying energies closely centered around their thermodynamically predicted value. In the micro-canonical case the thermodynamic idea of an energetically isolated system as in a perfect, unchanging equilibrium state is replaced by that of systems whose components fluctuate from their equilibrium values, but that are, with overwhelming probability, near equilibrium and at equilibrium "on the average" over time. An examination of the fluctuation among components in an ensemble whose members are micro-canonically distributed shows that these fluctuations will be describable by a canonical distribution governed by the equilibrium temperature of the isolated system.

One can find thermodynamic analogies for the micro-canonical ensemble as one can for the canonical ensemble. The quantity "analogous" to entropy turns out to be log V, where V is the size of the phase space region to which the points representing possible micro-states of the system are confined by the macro specification of the system. Analogous to temperature is $d\varepsilon/d\log V$, although, as Gibbs is careful to point out, there are difficulties in this analogy when one comes to treating the joining together of systems initially isolated from one another.

Gibbs is also careful to point out that there are frequently a number of distinct quantities that converge to the same value when we let the number of the degrees of freedom of the system "go to infinity." When we are dealing with systems with a finite number of degrees of freedom, one of these quantities may be most "analogous" to a thermodynamic quantity in one context, whereas another one of them might, in its functional behavior, be more analogous to that thermodynamic feature in some other context.

Gibbs also points out that the values calculated for a quantity using the canonical ensemble and its thermodynamic analogy, and those calculated from the micro-canonical ensemble with its appropriate analogy, will also coincide in the limit of vast numbers of degrees of freedom. Because it is usually easier to analytically evaluate quantities in the canonical ensemble framework, this suggests this framework as the appropriate one in which to actually carry out calculations, even though Gibbs himself declares that the canonical ensemble "may appear a less natural and simple conception than that which we have called a micro-canonical ensemble of systems."

In his discussion of the thermodynamic analogies for the microcanonical ensemble, Gibbs also points out clearly the distinction between average values calculated by means of an ensemble and most probable values calculated by means of the same ensemble. Once again for finite systems these may well differ from one another, even if the usual convergence of values in the limit of systems of vast numbers of degrees of freedom seems to obliterate the distinction when it comes to calculating the values of thermodynamic quantities by means of the analogies.

3. The theory of non-equilibrium ensembles

The Gibbsian program supplies us with a method for calculating equilibrium values of any thermodynamic quantity. Given the constraints imposed on a system, set up the appropriate canonical or micro-canonical ensemble of systems subject to those constraints. Calculate from this ensemble the appropriate analogue to the thermodynamic quantity and identify the value obtained as the thermodynamic value for the system in equilibrium.

But how are we, within this ensemble view, to understand the inevitable approach to equilibrium of systems initially not in equilibrium? In Chapter XIII of his book, Gibbs shows, for a number of standard cases such as that of two systems originally isolated and at equilibrium that are then allowed to energetically interact, that if we assign to the systems the appropriate ensembles, the features of equilibrium so described will match the thermodynamic prediction. For example, in the case cited, after the interaction the component of the new combined system that was originally at the higher temperature will have transferred heat to the component of the new combined system that was originally at the lower temperature.

But what right do we have to assume that the modified system will be appropriately described by a canonical or micro-canonical ensemble? The problem here is this: We have already described by means of an ensemble the initial state of the system when it begins its transition to a new equilibrium by, say, having the constraints to which it originally was subject changed, making its old equilibrium condition no longer an equilibrium condition. But each system of the original ensemble is determined in its dynamical evolution by the underlying dynamical laws during the period after the constraints have been changed. So it is not fair to simply choose a new ensemble description of the system at the later time. Instead, the appropriate ensemble description that evolves by dynamics from the ensemble description of the system at the earlier time must be the ensemble description of the system at the later time. Can we in any way justify describing the system at the later time by means of the appropriate canonical or micro-canonical ensemble, rationalizing this by an argument showing that such a description is the appropriate dynamically evolved ensemble from the original ensemble?

Gibbs takes up this problem in two chapters of the book. Chapter XI is devoted to showing that the canonical ensemble is, in a certain sense, the ensemble that "maximizes entropy." Chapter XII takes up the dynamical evolution of an ensemble, as the member systems of the ensemble follow the evolution dictated to them by dynamical laws.

We have already seen how, for the canonical ensemble, Gibbs offers, by means of the thermodynamic analogies derived from considerations of the variation of the ensemble with varying external constraints, $\overline{\eta}$ or $-\overline{\log P}$, which is $-\int P \log P dp dq$ as the ensemble analogue to the thermodynamic entropy. In Chapter XI, Gibbs proves a number of "extremal" theorems about this ensemble analogue to entropy. The major conclusions are that (1) if an ensemble of systems is canonically distributed in phase, then the average index of probability, $\overline{\eta}$, is less than in any other distribution of the ensemble having the same average energy; and (2) a uniform distribution of a given number of systems of fixed energy within given limits of phase gives a smaller average index of probability of phase than any other distribution. Together, these two results give us, in a certain sense, the result that the canonical and micro-canonical distributions are, for their respective cases of fixed average energy or fixed energy, the ensembles that maximize ensemble entropy.

Next, Gibbs takes up the problem of the dynamic evolution of ensembles. Suppose, for example, we have a micro-canonical ensemble determined by some fixed values of constraints. We change the constraints. Can we expect the micro-canonical ensemble to evolve, in a manner determined by the dynamic evolution of each system in the ensemble in accordance with the dynamical laws, into the micro-canonical ensemble relative to the newly changed constraints?

Now the ensembles we want to be the endpoint of evolution are the unique ensembles that are such that two constraints are met: (1) the number of systems between any values of functions of the phases that are constants of motion are constant; and (2) the value of $\bar{\eta}$, the mean of the index of probability, is minimized. Can we show, then, that the ensemble whose constraints have been changed will have its $\bar{\eta}$ value evolve with time toward the minimum value? That would show that the ensemble is evolving to the unique, stationary ensemble consistent with the constants of motion that are the standard equilibrium ensemble. That is, can we use the "excess" of the $\bar{\eta}$ value of the ensemble at a time over the $\bar{\eta}$ value of the equilibrium ensemble as an indicator of how far the ensemble is from equilibrium, and then prove the monotonic decrease of this value for the evolving ensemble?

Alas, a simple application of the same results that were required to show the canonical and microcanonical ensembles stationary in the first place shows that any ensemble, no matter how it evolves, will always have a constant value of $\bar{\eta}$.

But, says Gibbs, one ensemble may in some sense approach another even if some function of the first does not approach that function of the other. He offers an explanation by analogy. Imagine two mutually insoluble fluids, *A* and *B*, in initially different but connected regions of a container. Stir the fluids. Because one never dissolves in the other, the volume occupied by fluid *A* is always constant and equal to its original volume. And the density of *A* at any point is always 1 or 0, just as it was when the mixing began. But take any small, fixed, spatial region of the container and look at the portion of that region occupied by the *A* fluid. Initially this will, except for a small number of regions on the boundary of the fluids, be the fraction 1 or 0. But as the mixing continues, we will expect, after a sufficiently long time, that the portion of any region that is occupied by the *A* fluid will be the same fraction of that region as the portion of original fluid that was *A*. (See Figure 2-7.)

We can expect an analogous result with ensembles. We divide the phase space into small but non-zero regions. When the constraints are changed, the ensemble initially fills a box entirely or the box is entirely unoccupied. As dynamic evolution goes on, the original compact ensemble "fibrillates" into a complex "strung out" ensemble ranging over the entire region of now, newly accessible phase space. Eventually we expect the portion of each box occupied by the ensemble to be the same, and equal to the original portion of the expanded phase space occupied by the ensemble when it started. In this sense the ensemble has become "uniform" over the expanded region of phase space, and, is therefore "like the appropriate equilibrium ensemble, which is genuinely, uniformly so spread, and not just spread uniformly in this new coarse-grained" sense. If we let time go to infinity and then let the boxes go to zero in

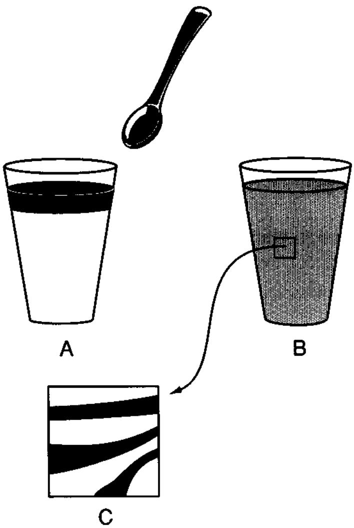

Figure 2-7. Mixing of insoluble fluids. A glass is filled 90% with clear water and 10% with a black ink as in A, the water and ink being insoluble in one another. The glass is stirred. The apparent result is uniform light gray inky water, as in B. But a magnified look at a small region of the fluid in B, as shown in C, reveals that each point is either totally in clear water or totally in black ink. The portion of any small volume filled by black ink is, however, 10%, if the ink is throroughly mixed into the water.

the limit, we will have a mathematical representation of this "spreading to uniformity" of the initial ensemble. We would miss this spreading if we first let the boxes go to zero size, and then let time go to infinity, because that would give us, once again, the result of constancy of the measure of "spread-outness" of the initial ensemble. (See Figure 2-8.)

With this new coarse-grained notion of spread-outness of an ensemble, we can introduce a new measure of the deviation of an ensemble from the equilibrium ensemble. Take the density of the ensemble in box i at a time as P_i, and define the "coarse-grained entropy" as $-\Sigma_i P_i \log P_i$. Whereas $\int P \log P d\Gamma$ is provably invariant, we can expect $\Sigma_i P_i \log P_i$ to decrease to its minimal value as time goes on. Gibb does point out that there will be "exceptional" ensembles that will not approach uniformity in even this coarse-grained sense. He doesn't offer any proof, however, that such "exceptional" ensembles will really be rare "in the world." Nor is there any proof that the approach to equilibrium, in this new sense, will be monotonic, nor that the time scale will be the one we want to replicate for the ensembles in the actual approach to equilibrium in time of non-equilibrium systems in the world. (See Figure 2-9.)

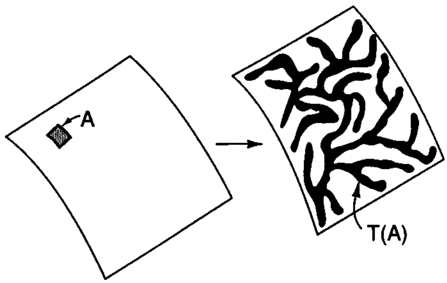

Figure 2-8. Coarse-grained spreading of an initial ensemble. Region A represents the collection of points in phase space corresponding to a collection of systems. Each system has been prepared in a non-equilibrium condition that is macroscopically the same. But the preparation allows a variety of microscopic initial states. As the systems evolve following the dynamics governing the change of microscopic state, A develops into T(A). The size of T(A) must be equal to that of A by a law of dynamics, but whereas A is a simple region confined to a small portion of the available phase space, T(A) is a complex fibrillated region that is spread all over the available phase space "in a coarse-grained sense." A uniform distribution over the available phase space is what corresponds to equilibrium in the theory. T(A) is not really uniformly spread over the phase-space, but may nevertheless be considered to represent a spreading of the initial ensemble that represents approach to equilibrium.

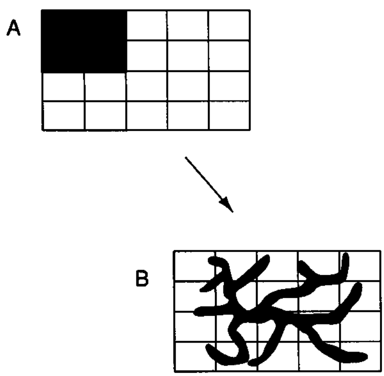

Figure 2-9. Coarse-grained entropy increase. Suppose an ensemble starts with the situation as noted in A – that is, each coarse-graining box is either totally full of ensemble points that are spread uniformly throughout the occupied boxes, or else totally empty of points. In B, the initial ensemble has become fibrillated and coarsely spread out over the available phase space. One can certainly show that the Gibbs coarse-grained entropy in B must be higher than it is in A, for by the definition of coarse-grained entropy it is minimal for the situation in which boxes are totally full or totally empty. But that, as the Ehrenfests noted, is a long way from showing that coarse-grained entropy will always increase monotonically, that it will approach the appropriate entropy for the new equilibrium condition after the constraints have been changed, or that it will approach this value in the appropriate time scale to represent observed approach to the new equilibrium.

Gibbs does point out the provable fact that the initial coarse-grained entropy can, at least, not be greater than the entropy of future states. But it is also the case that the fact that one can expect in the future ensembles of greater and greater coarse-grained entropy is matched by similar statements about the past. The "increase of coarse-grained entropy toward the future," if we have a right to expect it, is matched by a similar expectation of increase of coarse-grained entropy in the past of a system. That is, if we trace the dynamic evolution of the systems that constitute a non-equilibrium ensemble at a given time back into the past, we will generate ensembles that are more uniform in the coarse-grained sense than the ensemble at the specified time.

It is worth quoting Gibbs' response to this time-symmetry paradox in full, as it represents a theme we will see recur as the paradox of time-symmetry is faced by later investigators:

> But while the distinction of prior and subsequent events may be immaterial with respect to mathematical fictions, it is quite otherwise with respect to events in the real world. It should not be forgotten, when our ensembles are chosen to illustrate the probabilities of events in the real world, that while the probabilities of subsequent events may be often determined from probabilities of prior events, it is rarely the case that probabilities of prior events can be determined from those of subsequent events, for we are rarely justified in excluding from consideration the antecedent probability of the prior events.

We shall try to fathom what this means later on.

In Chapter XIII, Gibbs applies his results to the discussion of the appropriate ensemble description of various non-equilibrium processes. Here, Gibbs focuses not on the evolution of an ensemble representing an isolated system, but on the course of change to be expected as external constraints are varied. First, he considers constraints being abruptly varied. Here, changing the parameter does not effect the systems directly but does change the distribution from an equilibrium one to one that is now non-equilibrium in the newly accessible phase-space. In this case, one expects the ensemble to change in such a way that it finally, in the "coarse-grained" sense, approximates the equilibrium ensemble for the newly specified parameters. And one expects, therefore, the coarse-grained entropy to increase. In the case of slow enough change of parameter, one expects the ensemble to vary in such a way that even the coarse-grained entropy is constant.

Gibbs then takes up the problem of the bringing into energetic contact of systems originally energetically isolated, and argues, once again, that the resulting change of the ensemble representing the composite system can be expected to evolve, in a coarse-grained sense, in such a way that the resulting ensemble eventually comes to approximate the equilibrium ensemble appropriate to the new joint system subject to its overall external constraints. For example, heat having flowed from the originally higher temperature body to the one originally at lower temperature, the two components will be at equal temperatures as represented by the appropriate "modulus" constant of the ensemble.

IV. The critical exposition of the theory of P. and T. Ehrenfest

In 1912, P. And T. Ehrenfest published a critical review of the then current state of kinetic theory and statistical mechanics entitled "The Conceptual Foundation of Statistical Mechanics." Published in the *Encyclopedia of Mathematical Sciences*, the article provided a brilliantly concise and illuminating, if sometimes controversial, overview of the status of the theory at that time. This piece can be considered the culmination of the early, innovative days of kinetic theory and statistical mechanics, and the beginning of the critical discussion of the foundations of the theory that accompanied and still accompanies its scientific expansion and development. The piece is directed first to an exposition of the original kinetic theory and its changes in response to the early criticism, along with a discussion of remaining unsolved problems in the "statistical-kinetic" approach. It then moves to a critical exposition and critique of Gibbs' statistical mechanics.

The general argument of the piece is this: The development of kinetic theory from Krönig and Clausius to Maxwell and Boltzmann culminated in Boltzmann's kinetic equation and his proof of the H-Theorem. The H-Theorem led to the criticisms of the theory summed up in the Reversibility and Recurrence Objections of Loschmidt and Zermelo. These led Boltzmann to a reinterpretation of the description of system evolution given by his H-Theorem and kinetic equation that is perfectly consistent with the underlying dynamical theory. Despite this there are those who remained unconvinced that Boltzmann had avoided the fundamental problems of consistency posed by the Reversibility and Recurrence Objections. In order to show that he had, one must resolve many ambiguities in Boltzmann's expression of his views. This elimination of obscurities proceeds primarily by "transforming the terminology of probability theory step by step into hypothetical statements about relative frequencies in clearly defined statistical ensembles." On the other hand, the revised Boltzmannian theory of non-equilibrium does require for its justification postulates whose plausibility and even consistency remain in doubt.

Furthermore, "a close look at this process of development shows that the systematic treatment, which W. Gibbs attempts to give . . . covers only a small fraction of the ideas (and not even the most important ones) which have come into being through this process."

1. *The Ehrenfests on the Boltzmannian theory*

The older formulation of the theory. The Ehrenfests first offer a brief survey of the evolution of assumptions about "equal frequencies of occurrence" in the early development of kinetic theory. Krönig asserted that the motions of the molecules being very irregular, one could, "in accordance with the results of probability theory" replace the actual motion by a regular motion – for example, assuming the molecules all have the same speed. Clausius made a number of assumptions about equal frequencies of molecular conditions, some explicit and some only tacit. He assumed, for example, that the spatial density of the molecules was uniform, that the frequencies of molecular speeds did not vary from place to place, and that all directions of molecular motion occurred with equal frequency. He also, tacitly, assumed the *Stosszahlansatz.*

These are all postulates about "equal frequency" of conditions. What about frequency postulates where the frequencies are not all equal – the non-*gleichberechtigt* cases? Clausius had already assumed that at equilibrium there was a definite distribution of molecular velocities, even if he didn't know what it was. Maxwell posited his famous distribution law, guided, it seems, by the Gaussian theory of errors. Boltzmann succeeded in generalizing this distribution law to the cases of gases in external potentials and to the cases of polyatomic gases from Maxwell's law, which holds for monatomic gases.

But how can one derive the equilibrium velocity distribution law? Maxwell and Boltzmann, treating dynamic evolution of the system, showed the equilibrium distribution to be stationary. Boltzmann, by means of the H-Theorem, showed it to be the uniquely stationary distribution. The H-Theorem requires a mechanical assumption in its proof, but it also requires the kinetic equation, which assumes the truth of the *Stosszahlansatz.* So we can reduce the case of non-equal frequency assumptions to an equal-frequency assumption, the *Stosszahlansatz.* But clearly, then, the kinetic equation and the H-Theorem are the central components of the theory that we must critically explore.

Now Boltzmann had defined the quantity H and shown that if the evolution of the distribution function obeyed his kinetic equation, H must monotonically decrease. But Loschmidt's Reversibility Objection and Zermelo's Recurrence Objection show that this alleged monotonic decrease in H from any non-equilibrium distribution cannot be valid. This led to the new "Statistical Mechanical" or "Kineto-Statistical" approach to the theory.

The "modern" formulation of the theory of equilibrium. The revised approach to the theory of equilibrium, according to the Ehrenfest account, is one that takes as its basic arena the phase space of points, each of which represents a possible micro-state of the system compatible with its macroscopic constraints. The evolution of a system under the dynamical laws is represented by a path from a point in the phase space. If we measure volume in phase space in the standard way noted in Section III,1, a theorem of dynamics, Liouville's Theorem, will show us that the volume assigned to a set of points will be the invariant volume of the regions that set evolves into under dynamic evolution. If we restrict our attention to points all of which have a common energy – that is, to an energy hypersurface in the phase space – and to flows confined to that surface, a new measure of "area" on the surface – the micro-canonical measure discussed earlier – will play the role of invariant measure under dynamical flow. If the probability distribution over all phase space is an arbitrary function of the constants of motion of the systems, it will be constant along any path. And it will be stationary in time, always assigning the same probability to any region of the phase space as the systems evolve. In the energy-surface constrained case, a similar probability function, normalized by the surface measure, will similarly be a stationary probability distribution.

At this point, the Ehrenfests introduce their notion of an ergodic system as one whose path in phase space "covers every part of a multi-dimensional region densely." They attribute to Maxwell and Boltzmann the definition of an ergodic system as one whose phase point "traverses every phase point which is compatible with its given total energy." (Although, as historians point out, it isn't at all clear that either Maxwell or Boltzmann ever really defined the relevant systems in exactly this way.) If the systems obey this latter condition, then every system in the ensemble traverses the same path if we follow it over an unlimited time. Therefore, every system will have the same time average – defined by $\lim_{t\to\infty}\frac{1}{t}\int_0^t f(p,q,t)dt$ – of any quantity that is a function of the phase value of the system,

But, the Ehrenfests maintain,

> The existence of ergodic systems, *i.e.* the consistency of their definition, is doubtful. So far, not even one example is known of a mechanical system for which the single *G*-path, that is the path in the phase space under dynamic evolution, approaches arbitrarily closely each point of the corresponding energy surface. Moreover, no example is known where the single *G*-path actually traverses all points of corresponding energy surface.

But it is just these assumptions that are necessary for Maxwell and Boltzmann's claim that in the gas model, all motions with the same total energy will have the same time average of functions of the phase.

Next, special distribution functions are introduced: those over phase space that are a function of energy alone, of which the canonical distribution is a special case; and the standard micro-canonical distribution over the energy surface. It is these distributions that are constantly used to derive results in gas theory, despite the fact that they were originally introduced as justified by the assumption of ergodicity, even though ergodicity is not usually even mentioned by those who have taken up the use of these distributions in the foundations of their work.

At this point, the Ehrenfests offer a plausible reconstruction of what "may have been" the starting point of Boltzmann's investigation – that is, of his suggested plan to use the micro-canonical ensemble average as the appropriate representative of equilibrium values. Begin with the empirical observation that a gas goes to equilibrium and stays there. So the average behavior of a gas over an infinite time ought to be identical to its behavior at equilibrium. Can we calculate "time averages of functions of the phase over infinite time," in the limiting sense of course? In particular, can we show that the infinite time average of a gas corresponds to the Maxwell–Boltzmann distribution? Take any phase quantity f – that is, $f(p,q)$. Then consider the following equalities:

(1) Ensemble average of f = time average of the ensemble average of f
(2) = ensemble average of the time average of f
(3) = time average of f

Equality (1) follows from the choice of a stationary probability distribution over the ensemble. Equality (2) follows from the legitimacy of interchanging averaging processes. Equality (3), however, is crucial. To derive it, we must invoke the postulate of ergodicity, for it is that postulate that tells us that the time average of f is the same for each system in the ensemble. This is so because the dynamic path from each point is always the same path if it is true that each system eventually traverses each dynamical point. The Ehrenfests point out that the derivation, as they have reconstructed it, looks rather more like the work of Maxwell, Lord Rayleigh, and J. Jeans than it does like that of Boltzmann, because Boltzmann chooses to derive from ergodicity the fact that the time spent by the system in a region of phase space goes – given ergodicity and in the limit as t goes to infinity – to the size of the region. But, they say, and as we shall see in detail in Chapter 5, that conclusion is equivalent to the earlier one that proves the identity of ensemble phase averages with infinite time averages given the posit of ergodicity.

The Ehrenfests conclude with the observation that given the postulate of ergodicity, the facts about infinite time averages and about relative duration of a system in states corresponding to a specified region of phase space in the infinite time limit follow from dynamical considerations alone. No "probabilistic hypotheses" are called for. But if, on the other hand, we dispense with the posit of ergodicity, it is unclear what rationales the probabilistic hypotheses of the standard distribution of probability over the ensemble have in the first place.

The statistical approach to non-equilibrium and the H-Theorem. At this point, the discussion moves on to an attempt to make fully unambiguous the statistical reading of the kinetic equation and the *H*-Theorem. In the process the Ehrenfests offer their demonstration of the consistency of these important non-equilibrium results with the underlying dynamics.

First, apparatus is set up that will allow the Ehrenfests to clear up a number of misunderstandings resulting from Boltzmann's loose language. Consider the phase space appropriate for a single molecule – that is, the space such that points in it represent the position and momentum, in all degrees of freedom, of one of the molecules of the system. Coarse-grain this phase space into small boxes, each of equal extent in position and momentum. Now consider the position and momentum of each molecule of the gas at a given time. This results in a certain number of molecules having their representative points in this small phase space, called μ-space, in each i-th box. Call the number in box i, a_i, and a given specification of the a_i's a state-distribution, Z. To each Z corresponds a region of points in the original "big" phase space, now called Γ-space – that is, the region of all those phase points representing the total molecular state of the gas with a given corresponding state-distribution, Z.

First, note that the region of Γ-space corresponding to the Z that is the Maxwell–Boltzmann distribution is overwhelmingly large and dominates the allowed region of Γ-space open to the gas. If we assume ergodicity, this will justify Boltzmann's claim that for the overwhelming amount of time (in the infinite time limit) spent by the gas, it is in the Maxwell-Boltzmann distribution.

Now, consider any function F of Z. Considered as a function of p,q it will be discontinuous, changing its value when p,q change in such a way that an a_i changes its value. One such function is $H(Z) = \Sigma_i a_i \log a_i$, and it is this H that will play a crucial role in our discussion of the H-Theorem. Next, let us discretize time, looking at the gas only at a number of discrete time points separated by an interval Δt from each other.

With this apparatus we can now present a tidied-up version of Boltzmann's claims about the behavior of H over time. We can say that $H(Z)$ will remain for overwhelmingly long time intervals near its minimum value, H_0. If H_1 is a value of H much above the minimum, then the sum of the time intervals in which H has a value at H_1 or above decreases rapidly with the increasing value of H_1. If we now move to discrete times, and look at the value of $H(Z)$ at each time, we can argue that whenever *Ha*, *Hb*, and *Hc* follow one another in time and all are well above H_0, when we look at neighboring time moments we will usually find situations in time like this:

$$\begin{array}{ccc} & Hb & \\ Ha & & Hc \end{array}$$

Much less often, but with equal frequency, we will find:

$$\begin{array}{ccccccc} Ha & & & & & & Hc \\ & Hb & & & & Hb & \\ & & Hc & \text{or} & Ha & & \end{array}$$

Only very rarely will we observe the following pattern to be the case:

$$\begin{array}{ccc} Ha & & Hc \\ & Hb & \end{array}$$

So, from a state with an H much above its minimum value we will almost always find that the immediately succeeding state is one with a lower value of H. (And almost always the immediately preceding value of H will be lower as well!)

In what sense, then, could the H-Theorem be true? Pick a Z_a, at t_a such that $H(Z_a)$ is well above H_0. Corresponding to this Z_a is a region of points in Γ-space. From each point, a dynamic path goes off, leading to states that at later times, $t_a + n\Delta t$ have a value H_n. The claim is that there will be values of $H, H_1, \ldots H_n$ such that nearly all the systems will have at time t_i, H values at or near H_i. The set of values $H_1, \ldots, H_n$ we call the concentration curve of the bundle of H-curves. We then assert that this curve monotonically decreases from its initial high $H_a(Z_a)$ value, converges to the minimum value H_0, and never departs from it. At any time t_n, the overwhelming majority of curves of H will be at a value near H_n, but, as the Recurrence Theorem shows us, only a set of measure zero of curves of individual systems will themselves behave in a manner similar to the concentration curve. Next, we claim that the curve of the H-Theorem, derived from the kinetic equation that presupposes the *Stosszahlansatz*, is identical to this concentration curve. Note that neither of the claims made (that the concentration curve will show this monotonic decreasing behavior and that it will, in fact, be replicated by the curve of the H-Theorem) has been proven. The claim here is only that

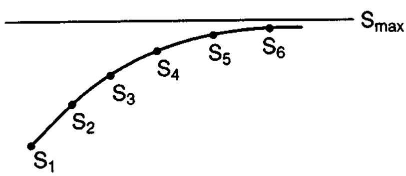

Figure 2-10. The concentration curve of a collection of systems. A collection of systems is considered, each member of which at time 1 has entropy S_1. The systems evolve according to their particular micro-state starting at the initial time. At later times, 2, 3, 4, 5, 6, . . ., the collection is reexamined. At each time, the overwhelming majority of systems have entropies at or near values S_2, S_3, S_4, S_5, S_6, . . ., which are plotted on the "concentration curve." This curve can monotonically approach the equilibrium value S_{max}, even if almost all the systems, individually, approach and recede from the equilibrium condition in the manner depicted in Figure 2-4.

a consistent reading of Boltzmann's claims has now been produced. An additional supplement, suggested by the Ehrenfests, is that a given observationally determined state of a gas must be posited to have one particular Z state overwhelmingly dominate the class of Z states compatible with that observational constraint. This is needed to connect the observational thermodynamic initial state with the posited definite initial Z_a in the earlier treatment of the H-Theorem. (See Figure 2-10.)

The picture presented here is time-symmetrical and perfectly compatible with the Objection from Reversibility. It is also compatible with the Objection from Recurrence. A further variant of the Reversibility Objection – that to each state there is the reverse state, and that if the former leads to an increase of H in a time interval, the latter must lead to a decrease of H in the same time interval of equal size – is noted to be fallacious, because there is no general argument to the effect that in a given time interval a state and its time reverse will lead to equal and opposite changes in the value of H for their respective systems.

Finally, the Ehrenfests take up the question of how to formulate properly the *Stosszahlansatz*, essential to the derivability of the kinetic equation and hence of the H-Theorem, in light of the new, statistical, understanding of the non-equilibrium situation. Boltzmann, they say, could be understood as making claims in his revised, statistical, view of irreversibility: (1) The *Stosszahlansatz* gives, for each time interval, only the most probable number of collisions, and the H-Theorem only the most probable value in the change in H; (2) the actual number of collisions (and the actual change in H) fluctuates around this most probable value. Here, once again, we must resolve obscurities and ambiguities by replacing "probability" language with a careful formulation in terms of relative frequencies in well-defined reference classes. While, the Ehrenfests say, statement (2) remained (at that time) to be clarified, J. Jeans, responding to criticisms of the *Stosszahlansatz* by S. Burbury, had made the meaning of statement (1) precise.

Consider a gas in state Z_a at a given time. How many collisions of a given kind will occur in time Δt? This depends on the specific phase of the gas at the time, and is under-determined by the Z-state. Look at the points in Γ-space corresponding to the Z-state. We can formulate the statistical *Stosszahlansatz* as the claim that the sub-region of the region of points in space corresponding to the Z-state in which the number of collisions is given by the *Stosszahlansatz* is the overwhelmingly largest part of the given region of Γ-space.

But to get the kinetic equation we need more. For we apply the *Stosszahlansatz* at every moment of time in deriving that equation. This leads to the Hypothesis of Molecular Chaos as a posit. Take the subset of the original region in Γ-space for which the *Stosszahlansatz* gave the right number of collisions in the time interval. Look at the systems those systems have evolved into at the end of that time interval, characterized by a new region of points in Γ-space. In order to derive the kinetic equation, we must assume that the overwhelming fraction of those points also represent systems whose collisions in the next interval of time will also be in accord with the *Stosszahlansatz*. And this hypothesis about the dominance of the *Stosszahlansatz* obeying region must be continuously repeated for the set into which that subset evolves, and so on.

We can get the result we want by considering the points reached by trajectories started in our initial region. These will correspond to different distributions of state, $Z_{B'}$, $Z_{B''}$, and so on. But each such $Z_{B'}$ region of points started in our initial region will correspond to only part of the Γ-space region corresponding to a given $Z_{B'}$ distribution, that part occupied by points on trajectories that started in our initial Z region. What we must assume is that in each such partial region corresponding to a Z-state, the largest region is that for which the *Stosszahlansatz* will hold, just as we assumed it for the whole region of Γ-space corresponding to a given initial Z-state.

Although this new Hypothesis of Molecular Chaos may need subtle modification (because of molecular correlation induced by finite size of molecules, and so on), it shows, according to the Ehrenfests, that a statistical formulation of the *Stosszahlansatz*, and therefore a statistical derivation of the kinetic equation and *H*-Theorem, exists. And the derivation is immune to the Objections of Recurrence and Reversibility. Note, once again, that no claim as to the provability or derivability of the needed posit from underlying dynamics is being made here.

2. *The Ehrenfests on Gibbs' statistical mechanics*

Whereas the Ehrenfests view Boltzmann, aided by his predecessors and Maxwell, as the seminal thinker in the field, the researcher whose work defines the field, even if it is sometimes ambiguously stated and needful of clarification and supplementation, the Ehrenfests' critical view of Gibbs is rather unfavorable. In successive sections they offer an exposition and critique of Gibbs' approach: (1) to the theory of equilibrium; (2) to the theory of the irreversible approach to equilibrium; and (3) to the association of Gibbs' statistical mechanical quantities with thermodynamic quantities by means of the "thermodynamic analogies."

Critique of Gibbs' equilibrium theory. Gibbs, according to the Ehrenfests, first recapitulates Boltzmann's investigations into the most general stationary ensembles, and then fixes his attention on two "very special" cases – the micro-canonical distribution, which is equivalent to Boltzmann's ergodic surface distribution, and the canonical distribution.

Gibbs is able to compute many average values of functions of the phase quantities when the systems are canonically distributed in phase. For a given modulus (θ) of a canonical distribution, the overwhelmingly large number of systems have about the same total energy, E_0, given the vast number of degrees of freedom of the system. For this reason, "it is plausible that in general the average over the canonical distribution will be very nearly identical with the average taken over the micro-canonical or even ergodic ensemble with $E = E_0$." On the other hand, that distribution will be almost identical to the one obtained by taking all those phase points in Γ-space that correspond to a number a_i of molecules in box i in coarse-grained Γ-space such that $\Sigma_i a_i \varepsilon_i = E_o$, where ε_i is the energy of a molecule "centered" in the i-th box, and then distributing probability uniformly over those phase points in Γ-space. But average values calculated by this last probability function will be values of phase quantities when the molecules are in the Maxwell–Boltzmann distribution. So, essentially, average values calculated using the canonical ensemble can be expected to be the same as values calculated by assuming that the gas is in one of its "overwhelmingly most probable" Maxwell–Boltzmann distributed states. "Thus, from the point of view of Boltzmann's presentation, the introduction of the canonical distribution seems to be an analytical trick" – that is, merely a device to aid in calculation.

And, the point is, Gibbs does not even mention the subtle problems of the posits of ergodicity needed to justify the claims that these methods are the appropriate methods for calculating equilibrium.

Critique of Gibbs on non-equilibrium and the approach to equilibrium. First, say the Ehrenfests, Gibbs defines a function $\sigma = \int \rho \log \rho \, dq dp$ and shows that this function takes on its minimum value subject to appropriate constraints (definite average energy for canonical case, definite energy for micro-canonical case) in the canonical and micro-canonical distributions. But this quantity is probably invariant under dynamical evolution, the proof being a simple corollary of Liouville's Theorem. Then Gibbs suggests dividing the Γ-space into small boxes, taking the average ensemble density in each box, P_i, and defining $\Sigma = \Sigma_i P_i \log P_i$ as a new measure of departure of an ensemble from its equilibrium distribution.

Gibbs concludes that every non-equilibrium ensemble will stream in such a way that the P_i's will become stationary in the infinite time limit and such that in the limit of infinite time, the value of $\Sigma(t)$ will assuredly be less than or equal to its value at the initial moment. His argument for this consists in the analogy with the mixing of insoluble liquids. Further, his arguments can be taken to show that in the infinite time limit, the P_i's will all be the same on a given energy surface.

Now one certainly can demonstrate the inequality noted here, for if we start with an ensemble all of whose coarse-grained boxes are either totally filled or totally empty, then $\Sigma(t)$ cannot increase in time, and may very well decrease.

But, say the Ehrenfests, Gibbs has not shown, by any means, all that needs to be shown. First, there is the question of the time the ensemble will take to spread out in such a way that $\Sigma(t)$ is at or near its minimum value. The Ehrenfests suggest that this can be expected to be many cycles of the time needed for a system that leaves a coarse-grained box to return to that box – that is, many of the " 'enormously large' Poincaré–Zermelo cycles." Nor does Gibbs' argument show in any way that the decrease of $\Sigma(t)$ with time will be monotonic. It could very well sometimes decrease and sometimes increase, even if the limiting inequality is true. Finally, Gibbs fails to show that the limiting value $\Sigma(t)$ will in fact be that value corresponding to the appropriate equilibrium ensemble for the now modified constraints. As far as Gibbs' proofs go, $\lim_{t \to \infty} \Sigma(t)$ could be arbitrarily larger than the fine-grained entropy of the appropriate canonical or micro-canonical ensemble. In his treatment of the thermodynamic analogies, Gibbs simply assumes, without noticing he is making the assumption, that one can identify the ultimate ensemble as that which approximates the equilibrium ensemble although, as just noted, he really has not demonstrated this, for he has only shown, at best, "a certain change in the direction of the canonical distribution."

What, finally, is the response of the Ehrenfests to Gibbs' observation that his "increase of coarse-grained entropy" result holds both forward and backward in time from an ensemble stipulation at a given time? Gibbs observed that we often infer future probabilities from present probabilities, but that such inference to the past was illegitimate. In their footnote 190, the Ehrenfests respond: "The penultimate paragraph of Chapter XII [the paragraph in which Gibbs makes the relevant observations quoted above], so important for the theory of irreversibility, is incomprehensible to us."

Critique of Gibbs on the thermodynamic analogies. As the final component of their critique of Gibbs' theory, the Ehrenfests offer a systematic comparison of how the components of the statistical-kinetic theory are to be associated with those of thermodynamics in the alternative approaches of Boltzmann and Gibbs.

First, they note that the Maxwell–Boltzmann distribution of molecules as they are distributed in μ-space is formally comparable to the canonical distribution function description of the distribution of systems in Γ-space in the canonical ensemble. If we think of slowly modifying the constraints on a system, we get, in the Boltzmann picture, a description of how the parameters determining the Maxwell–Boltzmann distribution vary with this parameter variation. This parallels Gibbs' description of how the parameters governing the canonical distribution will have their variation connected with variation of constraints in his approach. And both resulting equations will *formally* resemble the familiar equation governing the relation of thermodynamic quantities in the phenomenological theory.

For a system in equilibrium, Boltzmann thinks of the system as having a representative point in Γ-space, which characterizes its micro-state, as overwhelmingly likely to be in that dominating region of points that correspond to the Maxwell–Boltzmann distribution. From this distribution, one can then calculate the appropriate surrogates to the thermodynamic quantities. Gibbs, on the other hand, identifies averages over all phases in the appropriate canonical ensemble with the thermodynamic quantities.

How do Boltzmann and Gibbs treat the problem of two systems, initially each in equilibrium but energetically isolated, and then brought into a condition where they can exchange energy with one another? Boltzmann describes each system, prior to interaction, as having its appropriate Maxwell–Boltzmann distribution over its molecules. Once in interaction, they are in a non-equilibrium initial state. The kinetic equation and H-Theorem show that on the average (and with overwhelming frequency) the combined system will evolve to the new equilibrium Maxwell–Boltzmann state appropriate to the combined system. Gibbs will describe each component system initially by a canonical ensemble.

To represent the systems in thermal contact we pick a new ensemble each member of which is a combination of one system from each of the earlier ensembles. Unless the initial such ensemble is canonical, it will "smear out," in the coarse-grained sense, until the new ensemble "approximates" the appropriate canonical ensemble for the combined system.

How do they treat reversible processes? For Boltzmann, one can, given the slow variation in the parameters, treat the system as always in a Maxwell–Boltzmann distribution. Using the thermodynamic analogy relation obtained earlier, one can derive that entropy is (tentatively) generalizable to $-H = -\Sigma_i a_i \log a_i$ where a_i is the occupation number of the i-th box in coarse-grained μ-space. For Gibbs, slow change of parameters means that we can view the system as represented by a canonical ensemble at every moment. The analogue to entropy is given by $-\int \rho \log \rho \, dpdq$, where ρ is the density function of the canonical ensemble in Γ-space.

How do they treat the increase of entropy in an irreversible process? For Boltzmann, if the system does not have a Maxwell–Boltzmann distribution at a time, then by the H-Theorem or the kinetic equation, it will, in almost all motions, assume smaller values of H at later times. For Gibbs, it is the quantity $\Sigma_i P_i \log P_i$, defined by coarse-graining Γ-space and taking the P_i's as average ensemble density in the i-th box, whose decrease with time represents the changing ensemble and which, in the infinite time limit, must assume a smaller value than its initial value.

Concluding remarks. Summing up their critiques both of Boltzmann and Gibbs, the Ehrenfests remark that their conceptual investigation into the foundations of kinetic-statistical mechanical theory required that they "emphasize that in these investigations a large number of loosely formulated and perhaps even inconsistent statements occupy a central position. In fact, we encounter here an incompleteness which from the logical point of view is serious and which appears in other branches of mechanics to a much smaller extent."

But these foundational and conceptual difficulties have not prevented physicists from applying the basic modes of thought over wider and wider ranges of phenomena, generalizing from the theory of gases to radiation theory and other areas. How optimistic can one be that they will prove warranted in their optimistic use of the theory? Here the Ehrenfests are guarded in their prognosis, not least because of the notorious difficulties with the fundamental consequence of equilibrium theory – the equi-partition theorem – giving apparently wrong results in the case of the theory of poly-atomic molecules and absurd results (divergence of the distribution function with increasing frequency) in the case of radiation. These difficulties were not resolved until the classical dynamical underpinning of the theory was replaced with a quantum micro-dynamics.

V. Subsequent developments

In the three-quarters of a century since the Ehrenfest review article appeared, there has been an intensive and massive development of kinetic theory and statistical mechanics. To hope to even survey the multiplicity of ramifications of the theory would be impossible here. It will be helpful for our further exploration of foundational problems, however, to offer an outline of some of the high points in the development of the theory subsequent to its innovative first period described earlier.

Two of the following sections – 1 and 3 – will simply note some of the gains that have been made in extending the theory of equilibrium and of non-equilibrium to allow them to cover cases outside the range of the early theory. Whereas the early theory is primarily devoted to understanding dilute gases, the theory can be extended widely both to cover more difficult cases within the realm of kinetic theory proper (such as gases in which more than binary collisions are relevant, and so on) and to realms of phenomena such as magnetization of matter and the distribution of energy in radiation that although treatable in general statistical mechanics, go beyond the case of many-molecule gas systems in various ways.

The other two following sections – 2 and 4 – treat material more directly relevant to our project. Here I will offer the barest outline of the direction in which attempts have been made to continue the program of rationalizing or justifying the methods used in statistical mechanics, in particular in rationalizing and justifying the choice of probabilistic postulates such as the traditional ensemble probability distributions in equilibrium theory and the Hypothesis of Molecular Chaos in non-equilibrium theory. Such programs of rationalization and justification are also intimately connected with the problem of explaining *why* the posited probability assertions are true, if indeed they are. Or why, if they are not, the theory that posits them works as well as it does. Because the topics of these two sections will occupy our attention for three full chapters in Chapters 5, 6, and 7, all that is offered here is an outline to be fleshed out in greater detail.

1. The theory of equilibrium

From classical to quantum micro-theory. We have already seen how a major failure in the predictive reliability of equilibrium statistical mechanics cast doubt upon the validity of the entire theory. One fundamental result of the theory is that at equilibrium the thermal energy of a substance will be equally distributed over all of the degrees of freedom available to its components. But even in the case of diatomic molecules, the energy fails to be so distributed when the vibrational and rotational degrees of freedom of the molecule are added to its translational degrees of freedom. In the case of radiation, the equi-partition theorem leads to totally incoherent results, because the possibility of standing waves in a cavity having unlimitedly short wavelength leads to a distribution function that diverges with high frequency.

One of the origins of quantum mechanics is in Planck's study of the thermodynamics and statistical mechanics of radiation. He was able to get the empirically observed distribution curve by assuming that energy is transferred from radiation to matter, and vice versa, only in discrete packets, each having its energy proportional to the frequency of the relevant radiation. This was generalized by Einstein to the view that energy exists in such "quanta" even in the radiation field. Combined by Bohr with the existence of stationary states of electrons in atoms and the quantization of emission and absorption of radiant energy by them, we get the older quantum theory. Soon, difficulties of applying that theory to more general cases than the hydrogen atom, along with the desire for a general quantum kinematics on the part of Heisenberg and an exploration of "wave-particles" duality by de Broglie and Schrödinger, give rise to the full quantum theory.

From the point of view of this theory, the underlying dynamics of the components of a system (molecules of the gas, frequency components of the radiation field, and so on) is not governed by classical mechanics at all but by the quantum mechanical laws. This requires a total reformulation of kinetic theory and statistical mechanics. Where we previously dealt with ensembles of micro-states, taking micro-states to be specifications of the positions and momenta of each molecule, now we must deal with ensembles of quantum micro-states, where these are represented by rays in a Hilbert space or, more generally, by density matrices.

We shall, of course, not divert ourselves into the mysteries encumbent in a study of the meaning of the foundational concepts of quantum mechanics. Nor shall we even devote a great deal of attention to the ways in which a statistical mechanics founded upon an underlying quantum mechanical micro-theory differs in its structure and its predictions from a statistical mechanics founded upon an underlying classical mechanical micro-theory. But we will note here one curious apparent consequence of the theory and a few ways in which the change in the micro-theory does impinge on the study of the fundamental statistical assumptions of statistical mechanics.

The curious consequence of the theory arises out of its statistical mechanics of systems made up of a multiplicity of similarly constituted particles. For reasons that are, to a degree, made available by some results of quantum field theory, it turns out that in considering possible states of systems made of particles with half-integral spin, we must construct our ensembles only over state-functions that are anti-symmetric under permutation of the particles. In constructing ensembles for particles whose spin is integral, we must restrict ourselves to symmetric wave functions. In both cases, the simple way of counting micro-states in the classical theory is abandoned. This is the method that takes micro-states obtained from one another by a mere permutation of particles from one phase to another as distinct micro-states. Statistically, the result is the creation of two new distribution functions: the Fermi–Dirac distribution for half-integral spin particles, and the Bose–Einstein distribution for particles with integral spin. The Maxwell–Boltzmann distribution remains only as an approximating distribution for special cases.

Philosophically, the new way of counting "possible micro-states compatible with the macroscopic constraints" leads us, if we insist upon thinking of the particles as really being particles – which, if quantum mechanics is correct is a position hard to maintain – to say very curious things both about the "identity through time" of a particle, about the alleged Principle of Identity of Indiscernibles, and about the possibility of apparently non-causal correlations among the behaviors of components of a compound system. Because I believe these are puzzles that can only be coherently discussed within the context of a general study of the foundations of quantum mechanics, it will be best to leave them to the side.

More relevant to our purposes are some special features of ensembles in the quantum mechanical statistical theory that do interfere with the smooth treatment of the problem of the nature and status of the fundamental statistical posits of statistical mechanics. In classical statistical mechanics, we can have an ensemble that has no "recurrence" over time in its distribution, even if the individual systems in the ensemble are all governed by the Poincaré Recurrence Theorem. In quantum statistical mechanics, this is not true. If we want a non-recurrent ensemble, we can only obtain it in some special way – for example, by going to the idealization of a system with an infinite number of constituents. Further, we can show that what we will call ergodicity and mixing results of various kinds obtainable in classical statistical mechanics cannot hold in the theory with a quantum micro-foundation.

Finally, as we shall see, a number of investigators have tried to found an explanatory theory of the nature of the fundamental statistical posits of the theory by arguing that these rest upon constraints in our ability to so prepare systems as to construct ensembles that fail to behave in the appropriate kinetic way. Their arguments frequently draw analogies between limitations on our preparation of systems, which result from purely classical considerations, with the alleged limitations on measurement and preparation familiar from the Uncertainty Relations in quantum mechanics. When we explore this approach to the rationalization of non-equilibrium ensemble theory, we will have to deal with these alleged similarities to the quantum mechanical situation.

Extending the equilibrium theory to new domains. In standard statistical mechanics, the methodology for deriving the thermodynamic features of equilibrium is routine, if sometimes extraordinarily difficult to carry out in practice. One assumes that equilibrium properties can be calculated by establishing the appropriate canonical ensemble for the system in question. The form of this ensemble is determined by the external constraints placed upon the system, and by the way the energy of the system is determined, given those constraints, by the generalized position and momentum coordinates of its component parts. Then one uses the now familiar argument that in the limit of a very large number of components, one can use the micro-canonical ensemble or the canonical ensemble to represent the problem with the assurance that the thermodynamic results obtained will be the same. Actually, the true rationale for this requires subtle argumentation, some of which we will note later.

Formally, what one calculates is the function Z, the *Zustandsumme* or partition function, where Z is defined by

$$Z = \int e^{-E(p,q)/T} dp_1 \ldots dp_n dq_1 \ldots dq_n$$

in the classical case, and by the analogous sum

$$Z = \Sigma_i e^{-Ei/T}$$

in the quantum mechanical case. E is the energy and T the temperature of the system.

By a Gibbsian thermodynamic analogy, one can identify a thermodynamic quantity, the Helmholtz free energy, F, as $F = -T \log Z$. Then, from the transformation laws of the thermodynamic quantities, one can derive the other thermodynamic functions. The end result of the process is a characterization of the equilibrium thermodynamic properties of the system in terms of the parameters constraining the system and the features of its internal constitution, such as the number of constituent components in the system. Here, the standard Gibbsian identification of ensemble averages with equilibrium thermodynamic features is presupposed.

Because the energy function can be a complex function of the positions and momenta of the components, depending in crucial ways, for example, on spatial separations of particles acting on one another by potential-derived forces and, in the case of electromagnetic forces, by forces that may depend on relative velocity as well, the actual calculation of the partition function as an analytically expressible function of the constraints and structural features may be impossible. This leads to a wealth of approximative techniques, series expansions, and so on. Cases such as L. Onsager's exact calculation of the thermodynamic functions for the Ising model of a two-dimensional lattice of spinning particles interacting by their magnetic moments are rare triumphs for the theory.

But if one is satisfied with approximative solutions, the wealth of systems whose equilibrium properties can be calculated by this orthodox method is great: dilute and dense gases of molecules; one-, two-, or three-dimensional arrays of spinning particles interacting by their magnetic moments; radiation confined to a cavity (which can be viewed as being "composed" of its monochromatic component frequencies); interacting systems of matter and radiation; plasmas of charged particles, and so on. All fall prey to the diligent application of sophisticated analytical and approximative techniques.

Phases and phase changes. One of the most characteristic macroscopic features of matter is the ability of macroscopically radically different forms of matter to coexist at equilibrium. Typical examples are the gas-liquid, gas-solid, and liquid-solid phases of matter at boiling, sublimation, and freezing points. But examples abound from the macroscopic magnetic states of matter and other thermodynamically describable situations as well.

We shall see in Section 2 how the study of the so-called "thermodynamic limit" throws light upon the existence of phases and the nature of the change from one phase to another. Here I wish only to note the existence of a theory that supplements the general principles of statistical mechanics and has proven enormously successful in shedding light upon the nature of the change from one phase to another, at least upon the nature of the transition to the solid phase. The special feature of solidity is long-range order. Whereas in a gas one expects at most short-range deviation from pure randomness in the state of the molecules the deviation to be expressible in a correlation of the state of nearby molecules because of their dynamic interaction, in a crystal one finds an astonishing regularity that holds to macroscopic distances.

Reflecting on the fact that small arrays and large arrays show the same kind of correlational order in the solid phase, a theory – renormalization theory – is invented that provides great insight into the nature of the solid phase and the transition to it from, say, the random gas phase. In particular, this theory is able to explain why it is that quite dissimilar systems (atoms interacting by van der Waals forces and spinning electrons interacting by means of their magnetic moments, for example) can show phase transitions to the solid state that are quite similar in their macroscopic characteristics. Renormalization theory is able to explain why it is that general features of the system such as its dimensionality and the number of degrees of freedom that are available to its constituents are what are important in the characterization of the nature of the phase transition. The specifics of the intercomponent forces are much less important.

2. Rationalizing the equilibrium theory

Ergodicity. We have seen how, from the very beginning of the ensemble approach to the theory, there has been a felt need to *justify* the choice of the standard probability distribution over the systems represented (in Gibbs) by the canonical and micro-canonical ensembles. In Chapter 5, we shall spend a great deal of time exploring exactly what it would mean to justify or rationalize such a probability distribution, as well as ask in what sense such a rationalization would count as *explanation* of why the probability distribution holds in the world. It will put matters into perspective, though, if I here give the barest outline of the progress of so-called ergodic theory subsequent to the Ehrenfest review. The details of this history and an explanation of exactly what was shown at each stage and how it was shown will be provided in Chapter 5.

Maxwell, Boltzmann, and Ehrenfest all propose one version or another of an Ergodic Hypothesis. In some versions, one speaks of the dynamical path followed from the micro-state at a time of a perfectly isolated system. In other versions, one speaks of the path followed by the representative point of a system subjected to repeated small interference of a random sort from the outside. In some versions of the hypothesis, one speaks of the point as traversing a path that eventually goes through each point in the accessible region of phase space. In other cases, the path is alleged to be dense in the accessible region, or to come arbitrarily close to each point in the region over an unlimited time.

What is supposed to be provided by the hypothesis that is otherwise not demonstrable? Although the micro-canonical ensemble is provably a stationary probability distribution, we do not, without an Ergodic Hypothesis, know that it is the unique such stationary probability distribution. Ergodicity is supposed to guarantee this. Ergodicity is supposed to guarantee that the infinite time limit of a function of the phase is equal to the average of that phase function over all accessible phase points,

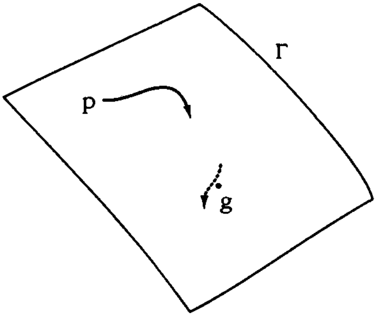

Figure 2-11. The Quasi-Ergodic Hypothesis. As a suggestion to replace the Ergodic Hypothesis, which, in its strong versions, is provably false, the Quasi-Ergodic Hypothesis is proposed. Let a system be started in a micro-state represented by some point p in the phase space Γ. Let g be any other representing point in the phase space. Then if Quasi-Ergodicity holds, the trajectory started from p will eventually come arbitrarily close to g. That is, given any initial micro-state of a system and any other micro-state allowed by constraints, the system started in the first micro-state will eventually have a micro-state as close as one likes to the given, second, micro-state.

where the average is generated using the micro-canonical probability distribution. Finally, it is supposed to follow from ergodicity that the limit, as time goes to infinity, of the fraction of time spent by the representative point of a system in a given region of phase space is proportional to the size of that region of phase space, when the size is measured using the unique invariant measure.

First, A. Rosenthal and M. Plancherel show that the strongest version of Ergodic Hypothesis – that the path of an isolated system will eventually actually go through each point in the accessible region of phase space – is provably false. This leads attention to be directed to quasi-ergodicity – that is, to the claim that the path will be, instead, dense in the phase-space region. It seems to be as difficult to prove any realistic model of a system quasi-ergodic as it was to prove it ergodic in the stricter sense. In any case, as we shall see, quasi-ergodicity, even if it holds, proves to be insufficient to prove such things as equality of time and space averages. (See Figure 2-11.)

Eventually, J. von Neumann and G. Birkhoff proved theorems that give necessary and sufficient conditions on the dynamical structure of an isolated system for the alleged *consequences* of ergodicity to hold. That is, the study of the path is dropped in favor of trying to prove facts about the system sufficient to show directly the uniqueness of the stationary distribution, the equality of time and phase average, and the proportionality of time in a region to size of region. But their necessary and sufficient condition – metric indecomposability – is itself extremely difficult to establish for realistic models.

For some years there have been attempts at performing "end runs" around ergodicity, offering weaker rationales for the probability distribution, rationales that are, at least, dependent only upon conditions that can be shown to follow from the dynamics of the system. Others drop the rationalization program entirely, proposing instead that the fundamental probability hypothesis be taken as an ineliminable basic posit of the theory.

Finally, as the culminating result of a long-term mathematical research program initiated by Birkhoff and von Neumann and carried on by E. Hopf, A. Kolmogorov, N. Krylov, V. Arnold, and others, Ya. Sinai is able to prove metric indecomposability for certain models that bear an interesting relation to realistic models of molecules in a box. But, at the same time, other important mathematical results at the hands of Kolmogorov, Arnold, and J. Moser provide rigorous limitations on the range of systems that could possibly satisfy the Birkhoff–von Neumann ergodic condition.

As noted, we shall follow out this history in detail in Chapter 5. More importantly, we shall there explore in detail just what one has a right to expect from ergodicity results, when they are available, in the manner of justifications of the probabilistic posits, rationalizations of them, or explanations of why they are so successful in predicting thermodynamic results.

The thermodynamic limit. We have seen continual reference in our historical survey to the important fact that the systems under consideration are constituted out of a vast number of micro-constituents. This undeniable fact has been invoked to make such claims as the assertion that fluctuations from equilibrium will be extremely rare over time and that the canonical and micro-canonical ensemble calculations will give the same values for thermodynamic quantities. Can this use of the large numbers of degrees of freedom of the system be rigorously justified? The attempt to put such claims on a firm foundation is known as the study of the thermodynamic limit.

Here we deal, if we are working with the typical case of a gas of molecules confined to a box, with systems whose volume is taken as very large, whose number of molecules is taken as enormous, but whose density is that of the system under consideration. This suggests the obvious idealization of dealing with systems in the limit as the volume and number of particles become infinite, with the density being held constant in the limiting process. The theory of the thermodynamic limit has been highly fruitful. Results may be rigorously obtained in interesting idealized cases. Furthermore, there is in this theory a nice sense of "control" over the limiting process in that one can not only prove things about the limit but get estimates on deviation from the limit in finite cases. This is something usually impossible to obtain in the situation of ergodic theory where infinite time limits can be obtained, but where, often, no hold on how much deviation from the limit one can expect in finite time intervals is available.

What are some of the results one desires from the study of the thermodynamic limit?

(1) We presuppose in statistical mechanics that the effect on the behavior of the system of the interaction of the molecules with the box at the edges of the gas is ignorable in calculating, for example, total energies due to intermolecular interaction. But what genuine right do we have to believe that this is so? Can we show that in the thermodynamic limit, the contribution of the molecule-box interaction to the potential energy stored in the gas goes to zero?

(2) If we demand that entropy be a function that is maximal for the standard equilibrium distribution when that distribution is compared with all distributions consistent with the specified constraints, and if we demand that it have the appropriate additivity properties for independent components of a system, then the unique statistical mechanical definition of entropy can be derived. Now in thermodynamics we take entropy to be an extensive quantity. A portion of gas in an equilibrium state that is like a portion in a similar state but twice the size is taken to have twice the entropy. Can we show in the thermodynamic limit that this will be so? That is, assuming the usual equilibrium ensemble distribution and the usual definition for entropy in terms of the distribution, can we show that in the limit of the vast system, the entropy will have the appropriate extensive nature?

(3) We generally assume that in the thermodynamic limit, an ensemble will describe a system whose fluctuations away from the dominating equilibrium state will become negligible. Whereas each ensemble provides a probability distribution over micro-states that is such that "most probable" distributions in Γ-space overwhelmingly dominate in probability those distributions in Γ-space that are far from equilibrium, what assurance do we have that the probabilities of the fluctuational Γ-space distributions will go to zero in the thermodynamic limit?

This problem is intimately related to the problem of the equivalence of ensembles. It is just this assumption of vanishing fluctuations that allows one to conclude that calculations done canonically and microcanonically will give the same thermodynamic results. Thermodynamic quantities are derivable from one another by appropriate transformations in the phenomenological thermodynamical theory. But different ensembles may be appropriate for deriving the different thermodynamical quantities in statistical mechanics. Only the proof of vanishing fluctuations – fluctuations whose characteristic nature varies from ensemble to ensemble – will allow us to legitimately select ensembles indiscriminately at our need or convenience in calculating the values of thermodynamic quantities while being assured that the resulting quantities will bear the appropriate thermodynamic transformational relations to one another.

(4) The existence of multiple phases in equilibrium with one another can plausibly be claimed to be modeled in statistical mechanics by a partition function that is non-analytic – that is, that fails at certain points to be a smooth function of its variable. But a simple proof is available to the effect that in a finite system the partition function is provably analytic everywhere. The thermodynamic limit approach to this problem is designed to show that such smooth behavior of the partition function can indeed break down in the thermodynamic limit, leading in that idealized case to a non-everywhere analytic partition function representing multiple phases. An alternative but similar approach to this problem recognizes in the path to the thermodynamic limit an analogy to the ergodic path to the infinite time limit from orthodox ergodic theory. This second approach attempts to correlate phase-change with a failure of this kind of "system size ergodicity."

But how can the results about the thermodynamic limit be obtained? Let us first note that the task here is not to justify the equilibrium ensemble distribution. This is presupposed. What additional presuppositions need be made in order to derive the desired results? In the case of a gas of interacting molecules, it is the nature of the interaction that is crucial. Only for certain energetic interactions of the molecules will the thermodynamic limit results be obtainable.

First, it is required that a stability condition be met. This demands that there be a lower bound to the potential energy produced by the interaction potential attracting one molecule to another. Next, there is a requirement on the interaction that the positive part of the potential converges sufficiently quickly to zero with increasing molecular separation. For some interactions, proving these results is not difficult. For other idealized interactions (such as the Coulomb interaction), various additional assumptions and clever analytical tricks are necessary. For gravitational interaction, as we shall see, the stability condition is not met, but for most realistic cases, one simply neglects the relatively insignificant gravitational attraction of molecules for one another.

Given these presuppositions, there are basically two ways to proceed. The first looks at the formulas connecting thermodynamic features to ensemble-calculated features for finite systems, and studies how those functions behave in the thermodynamic limit. Here, one is often able to prove the four results discussed as the aim of this theory.

There is an alternative to the approach that defines the thermodynamic functions for finite systems and then investigates how these functions behave in the thermodynamic limit. In this other program, one attempts to characterize an infinite system from the very beginning, seeking the appropriate representation to describe an idealized system of infinite extent and with an infinite number of components, but whose density is a specified finite value. When it comes to defining thermodynamic properties for such a system, one looks for the appropriate ensemble definition for quantities such as entropy density, rather than total entropy, which will be divergent.

One problem in characterizing such systems and their statistical mechanics is finding the appropriate restriction on initial micro-states to guarantee that wildly divergent behavior doesn't ensue. For such infinite systems, the possibility of solutions to dynamic evolution failing to exist after a finite time arises – say, by an infinite amount of energy being transferred to a single component in that finite time. Such possibilities do not arise in the finite-system case.

Once such constraints are discovered, the next task is to say what an equilibrium state is for such an idealized infinite system. One approach uses ergodicity as the defining criterion for a state being equilibrium. Although such equilibrium states for typical systems have been found, there remain difficulties in being sure all such states have been described. It is a curiosity of this approach that systems that when finite are never ergodic, such as the ideal gas, are ergodic in the infinite-system limit.

Other ingenious characterizations of equilibrium have also been proposed. A suggestion of Dobrushin, Lanford, and Ruell (the DLR condition) is that we think of equilibrium for a classical infinite system as being that state characterized by the fact that for any finite region, if we keep the particles outside of the region fixed, the particles inside the region will be describable by a Gibbs grand canonical ensemble. For quantum systems, an ingenious proposal from Kubo, Martin, and Schwinger (the KMS condition) provides a clear characterization of equilibrium state that can be extended to classical systems as well. Another way to characterize equilibrium in this approach to the thermodynamic limit is to think of equilibrium states as states that are stable under all local perturbations.

3. The theory of non-equilibrium

Solving the Boltzmann equation. The Maxwell transfer equations and the Boltzmann kinetic equation provide alternative, equivalent, means for describing the approach to equilibrium of a non-equilibrium dilute gas. But whereas the *H*-Theorem – at least if the presuppositions of the posit of Molecular Chaos can be established – gives us reason to expect, in the probabilistic sense, an approach to equilibrium from a non-equilibrium initial state, detailed knowledge of the nature of that approach requires a solution to the equation. Although Maxwell and Boltzmann were able to obtain limited results for the (unrealistic) inverse fifth-power forces – the so-called Maxwell potential – the solutions of the kinetic equation for realistic potentials were a long time in coming.

In 1916–17, S. Chapman, working from Maxwell's transfer equations, and D. Enskog, working from the Boltzmann equation, were able to determine a class of special solutions to the equations. They did not look for a general solution to the equation from an arbitrary initial condition, an analytically hopeless task to solve. Instead, relying upon the phenomenological equations of hydrodynamics, they simply assumed that linear transfer coefficients existed, and looked only for solutions that presupposed such linear coefficients. The implicit assumption here is that even from the most wildly non-equilibrium initial condition, the gas will very quickly reach a state that although still non-equilibrium, is one whose future evolution obeys the familiar phenomenological regularities of hydrodynamics. The special class of solutions Chapman and Enskog found are usually characterized as normal solutions to the equations. Although the discovery of such solutions can hardly be taken to give a full account of why the phenomenological regularities hold, given that their form is presupposed, the theory has positive and novel results that outrun the resources of the phenomenological theory. One can, for example, compute numerical values for the linear transport coefficients, values that are simply inserted into the phenomenological theory. One can also derive such empirically confirmed results as the dependence of the coefficients on various thermodynamic parameters, such as the variation in viscosity of a dilute gas with its temperature.

The problem of non-equilibrium ensembles. In Chapter 6, we shall explore in great detail the structure of various attempts to generalize non-equilibrium statistical mechanics beyond its origins in Maxwell's transfer equations and in Boltzmann's kinetic equation. As we shall see, there is no single, coherent, fully systematic theory that all will agree to that constitutes the correct core of non-equilibrium theory. There is, instead, a collection of approaches to the fundamental problems. Although various approaches are plainly related to one another, exact or rigorous demonstrations of equivalence of approach are rare. Even the question of the limits of the statistical mechanical approach is not a settled issue.

It is simply not known in any definitive way what general classes of non-equilibrium phenomena will be susceptible to the apparatus of ensembles and their evolution, or of kinetic equations derived from this kind of ensemble dynamics.

Whereas equilibrium theory is founded upon a choice of the equilibrium ensemble, non-equilibrium theory presents a somewhat different problem. Plainly, our approach to the foundations of non-equilibrium will require an ensemble way of dealing with systems. If a kinetic equation or some other dynamical description of approach to equilibrium is to be possible at all, it can only be in a probabilistic sense – whatever that turns out to mean. But what should a theory of non-equilibrium ensembles look like?

The first thing to notice here is the joint role played by the choice of initial ensemble and by the dynamical laws of ensemble evolution that are entailed by the micro-dynamics of the individual systems in the ensemble. If an initial ensemble is picked at one time, the evolutionary course of that ensemble is not open to our choice, for it is fixed by the underlying dynamics. This is true despite the fact that, as we shall see in Chapter 6, the method for deriving an appropriate description of dynamical evolution often involves the positing of some "perpetual randomness" assumption that in one way or another, generalizes the Posit of Molecular Chaos. How the ensemble evolves will depend upon its initial structure. We will need to justify or rationalize any such assumptions about its structure in order to complete our descriptive theory. But if we derive our description of the evolution of this initial ensemble by superimposing on the micro-dynamical laws some posit of a statistical nature, we shall have to show that the posit is actually *consistent* with the constraints on evolution imposed by the micro-dynamical laws.

How should an initial ensemble be chosen? This immediately leads to a fundamental preliminary problem: What are to be the macroscopic-phenomenological quantities by which the system and its evolution are to be described? In the case of equilibrium theory, the answer to this question has always been clear. One chooses the energy (or its mean if one was doing canonical ensembles) or, in special cases, the energy plus the small number of additional knowable constants of motion. Then one generates the appropriate ensemble. In some cases of non-equilibrium statistical mechanics, where we are dealing with systems near equilibrium or started from an equilibrium that is then destroyed by a parameter being changed, the equilibrium methods guide us to our choice of non-equilibrium initial ensemble as well. But in the more general case we really have no systematic guide to what macroscopic parameters can be fixed by the experimenter, and hence no systematic guide to what kinds of initial ensembles we must consider. This issue will be crucial when the alleged reduction of thermodynamics to statistical mechanics is discussed in Chapter 9.

Once we have decided what the macroscopic constraints are to be, the further problem of choosing a probability distribution for the systems over the allowable phase space of systems arises. Here, the method of equilibrium theory that picks out the standard invariant ensemble is not available to us, although, once again, it provides direction in limited kinds of cases. We shall explore some general proposals for systematically picking the probability distribution given the macroscopic constraints.

What methods, given an initial ensemble, or even leaving the choice of this ensemble indeterminate, can we find for deriving the appropriate dynamical evolution equations that will lead to our macroscopic non-equilibrium behavior?

Ideally, one would like there to be methods that, given a sufficient characterization of the initial ensemble, could, using the micro-dynamics alone, allow one to derive a description of ensemble evolution from which generalized kinetic equations could be derived. Such a derivation does exist in a very limited realm. Even there it serves only to rationalize or justify a kinetic equation (in this case, the Boltzmann equation) derived by other means. The actual route to kinetic equations in the general case comes about through the positing of some probabilistic principle that, superimposed on the dynamical laws, permits the derivation of a kinetic equation in a manner similar to the way in which the Posit of Molecular Chaos permits the derivation of the Boltzmann equation.

One group of such methods appears in the form of a direct generalization of the Posit of Molecular Chaos. Here, the functions that describe the correlations of the motions of molecules of various kinds (positions, momenta, and so on) are posited to be simple functionals of the functions that express lower-order correlations. The result is a hierarchy of equations for the correlation functions that can be transformed, by a posit or *Ansatz*, into closed equations for the lower-order correlations. The method is a clear generalization of the basic posit of a "perpetual *Stosszahlansatz*" of the kind noted by the Ehrenfests.

By use of this method, one can seek for kinetic equations for dense gases, plasmas, and so on. It is worth noting here that even for the simple cases that go beyond the Boltzmann equation for a moderately dense gas, it has proven to be extraordinarily difficult to construct an appropriate kinetic equation. And, in fact, it is usually not only impossible to *solve* the kinetic equation derived, in the sense that Chapman and Enskog solved the Boltzmann equation, but even to prove an appropriate *H*-Theorem for the more general cases.

A very important class of approaches that has proven to be of value in deriving kinetic equations is applicable to cases where the system can be viewed as a large number of energetically, almost independent, subsystems that interact only to an energetic degree much smaller than their intrinsic energy. Many systems are of this nature – an almost ideal gas with small intermolecular potential energy, radiation coupled by a small "dust mote" that absorbs and reradiates, almost harmonic oscillators coupled by a small mutual inharmonic force, and so on.

Here, one can deal with states that are the invariant states of the uncoupled system, and one can assume that the introduction of the small coupling serves to give rise to transitions of the system from one distribution over the invariant states of the uncoupled system to another. The technique is subtle, for often one must deal not with values of functions that are determined by the invariant uncoupled states, but with averages of such functions over those states. But once again, an assumption that transitions from one such state to another are determined by fixed constant probabilities over time becomes another way of representing the system as a Markov process, and leads, by means of what is called a Master Equation, once more to a kinetic equation describing a monotonic trend toward equilibrium.

Other approaches to the problem of generalizing from the Boltzmann equation follow a plan derived in conception from Gibbs' treatment of the non-equilibrium ensemble. Gibbs suggested that we could deal with non-equilibrium, at least in part, by "coarse graining" the **Γ**-phase space and defining the new coarse-grained entropy. The approach to equilibrium would be represented by an increase in coarse-grained entropy, corresponding to the initial ensemble "spreading out" in such a way that although its volume remained constant, the proportion of each coarse-grained box occupied by it became more and more uniform with time, heading toward an ensemble in which each coarse-grained box was equally full and equally empty of systems.

But Gibbs did not offer a postulate sufficient to show that this monotonic increase in coarse-grained entropy would in fact occur, a deficiency clearly pointed out by the Ehrenfests. One can supply such a postulate, however. As we shall see in Chapter 6,III,3, the assumption that the system evolves as a kind of Markov process does the trick. The assumption, in essence, is that the fraction of systems in a box at one time that will be in a given box at the next specified later time is constant. Again, this assumption, like the generalized Posit of Molecular Chaos, amounts to a continuous rerandomization assumption about the dynamic evolution of the ensemble, an assumption whose consistency with the underlying dynamical laws remains in question.

In Chapter 6, we shall follow in some detail a number of approaches toward finding some general principles for construing non-equilibrium ensembles and for deriving, by means of various posits, kinetic equations of various sorts from them. This is an area of physics lacking in any single, unified, fully coherent approach. There is, rather, a wide variety of approaches. That they are related to one another is clear, but there is in general no systematic way of rigorously or definitively elucidating their relationships to one another.

4. Rationalizing the non-equilibrium theory

It is one thing to be able to describe initial ensembles in a systematic way and to be able to posit appropriate randomization hypotheses so as to be able to derive a kinetic description of non-equilibrium systems. It is quite another thing to justify the correctness of one's description. And it is another thing again to be able to explain why the description is successful, if it indeed is. The whole of Chapter 7 is devoted to just these issues but, once again, it will be useful to give here the briefest possible outline of what some approaches to the justification and explanation problem might be.

First, there are a number of approaches to this problem that lie outside the "mainstream" or "orthodox" approaches. One of these non-standard approaches argues for the reasonableness of the principles that give rise to kinetic equations and irreversibility through an interpretation of the probabilities cited in statistical mechanics as subjective probabilities. The rules for determining probabilities are taken as originating in general principles of inductive inference under uncertainty.

A second non-orthodox approach, whose origins can be traced back quite far in the history of statistical mechanics, seeks the rationale in the non-isolability of systems from the external world. Here it is claimed that at least part of the explanation of the success of the statistical mechanical methods arises from the fact that systems idealized as energetically isolated from the rest of the world are actually not so isolated. It is, in this view, the perpetual interaction of the system with its environment, small as that may be, that accounts for the success of the statistical method and the randomization posits. A third non-orthodox approach seeks the resolution of the difficulties in the positing of time-asymmetric fundamental dynamical laws.

The more common, orthodox, directions in which justification and explanation are sought look for their resources to identifiable physical features of the world arising out of the structure of the system in question and the structure of the micro-dynamical laws in order to ground the derivation of the kinetic description.

In one unusual case, the kinetic equation can be derived from a posited structure of the initial ensemble by itself using solely the micro-laws of dynamical evolution. In other cases, it is derived by means that rely both on the structure of the initial ensemble and on posited randomizing features of dynamic evolution. In both cases one would like to find some grounds for believing that all initial ensembles will have the requisite feature. Here, what one seeks are physical reasons for a limitation on the kinds of actual ensembles of systems that we can prepare. The resources to which one turns include the structure of the individual systems – the underlying micro-dynamical laws – plus additional constraints upon the ability of an experimenter, restricted in some sense to the macroscopic manipulation of the system, to constrain a system's initial conditions. Whether such resources will, by themselves, serve to rationalize the restriction on initial ensembles or whether, instead, some additional fundamental probabilistic postulates resting on different grounds will be needed as well is, as we shall see in Chapter 7, a matter of important controversy and is fundamental to the study of the foundations of the theory.

When one's derivation of a kinetic equation rests either in whole or in part upon randomization posits imposed on dynamic evolution, a rather different task of justification and explanation needs to be faced. Because the dynamic evolution of an ensemble is fixed by the underlying micro-dynamics – at least in the orthodox approaches in which isolation from the outside is retained for the system and in which one accepts the standard micro-dynamics as a given and not as some approximation to a more stochastic or probabilistic underlying theory – one must reconcile the posited randomization imposed on the evolution with the evolution as constrained by the laws of the micro-dynamical theory.

Here, one might try to rely solely upon some demonstration that the evolution described by the micro-dynamics is adequately representable by the probabilistic description, the justifiability of the alternative representation resting upon the structure of the laws of dynamical evolution alone. Or one might offer a justificatory or explanatory argument that utilizes both the structure of the micro-dynamical laws and some posited structure feature of the initial ensemble. We shall see both approaches used in our detailed reconstruction of these arguments in Chapter 7.

It is in this part of the program of rationalizing the non-equilibrium theory that the use is made of a series of important advances in our understanding of the micro-dynamical laws and the structure of the possible solutions. In particular, various demonstrations of the radical instability of solutions under minute variations of initial conditions for the equations, and of the connection of such demonstrable instabilities with various "mixing" features of the ensembles whose evolution is governed by these equations, play a central role in the part of foundational studies. Here, powerful generalizations of the ergodicity results that we shall study in Chapter 5 come into play. We shall explore these more general results – which play in the foundational study non-equilibrium a role both like in some respects and unlike in other respects the role played by ergodicity results in the equilibrium theory – in detail in Chapter 7.

VI. Further readings

For a history of thermodynamics, see Cardwell (1971).

Two clear books on thermodynamics emphasizing fundamentals and concepts are Buchdahl (1966) and Pippard (1961). Truesdell (1977) is historically and conceptually oriented.

On the approach to thermodynamics that directly embeds probability into the theory, see Tisza (1966).

For the theory of non-equilibrium systems close to equilibrium, two good sources are Kubo, Toda, and Hashitsuma (1978) and de Groot and Mazur (1984). Introductions to the concepts used to deal with non-equilibrium thermodynamics far from equilibrium can be found in Prigogine (1980) and (1984). See also, Truesdell (1984).

Brush (1983) is a brief and clear history of kinetic theory. Brush (1976) is a compendium of essays covering many areas of the history of kinetic theory and statistical mechanics. Brush (1965) contains many of the seminal works in the history of kinetic theory and statistical mechanics selected from the original sources and translated into English where necessary. Maxwell's Demon is exhaustively treated in Leff and Rex (1990).

An excellent introduction to statistical mechanics emphasizing foundational issues is Tolman (1938). Two later works covering both older and newer topics and emphasizing fundamental issues are Toda, Kubo, and Saito (1978) and Balescu (1975). Munster (1969) is a treatise covering both foundational issues and applications.

Gibbs (1960) is essential reading.

Ehrenfest and Ehrenfest (1959) stands as a masterpiece of critical analysis of the theory up to 1910. For quantum statistical mechanics, Toda, Kubo, and Saita (1979) and Tolman (1938) are excellent.

An excellent survey of the important developments in statistical mechanics in the period prior to its publication is O. Penrose (1979), a retrospective piece that can direct the reader to much of the original literature.

Toda, Kubo, and Saito (1979) and Balescu (1975) contain very accessible treatments of the phase-transition problem and descriptions of the exactly solvable models.

The issue of the thermodynamic limit is taken up with mathematical rigor in Ruelle (1969) and in Petrina, Gerasimenko, and Malyshev (1989). See also, O. Penrose (1979), Section 2.

For further readings on the foundational problems of the rationalization of the probabilistic posits in equilibrium and non-equilibrium theory, see the "Further readings" at the ends of Chapters 5, 6, and 7 of this book.

3

Probability

As is already clear from the preceding historical sketch of the development of foundational problems in statistical mechanics, the concept of probability is invoked repeatedly in the important discussions of foundational issues. It is used informally in the dialectic designed to reconcile the time-asymmetry of statistical mechanics with the time-reversibility of the underlying dynamics, although, as we have seen, its introduction cannot by itself resolve that dilemma. Again, informally, it is used to account for the existence of equilibrium as the macro-state corresponding to the "overwhelmingly most probable" micro-states, and to account for the approach to equilibrium as the evolution of micro-states from the less to the more probable. More formally, the attempts at finding an acceptable derivation of Boltzmann-like kinetic equations all rest ultimately on attempts to derive, in some sense, a dynamical evolution of a "probability distribution" over the micro-states compatible with the initial macro-constraints on the system. The picturesque notion of the ensemble, invoked in the later work of Maxwell and Boltzmann and made the core of the Gibbs presentation of statistical mechanics, really amounts to the positing of a probability distribution over the micro-states of a system compatible with its macro-constitution, and a study of the changes of such a distribution over time as determined by the underlying dynamics.

But what is probability? The theory of probability consists of a formal mathematical discipline of astonishing simplicity at its foundation and of astonishing power in what can be derived in it. Along with the formal theory, there is a wide and deep literature devoted to trying to explicate just what elements are described within the formal theory. As we shall see, the so-called problem of the interpretation of probability – the problem of understanding in a general way what probabilities consist of – is one that remains replete with controversy. What seems initially to be a simple matter, requiring only conceptual clarification in matters of detail, is actually a puzzling and difficult area of philosophy.

Although it will be the aim of this chapter to survey a number of the most crucial assumptions and puzzles in several modes of the interpretation of probability, it will not be my purpose here to explore this to the depth it deserves as a problem in its own right. Such an effort would take us too far afield from our main goal – the pursuit of the foundational problems in statistical mechanics. But a survey of these general approaches to interpretation will provide an essential background for studying the more specialized problems in the application of probability to physics that are encountered in statistical mechanics. We shall also see how some of these general interpretive issues may very well depend for their resolution on some of the particular results obtained in the study of statistical mechanics, because, as we shall see in Chapter 7,IV,1 and 2, it is sometimes unclear where the dividing line between purely conceptual issues and issues about the nature of the world as revealed in our best physics lies.

This chapter will first outline some essential formal aspects of probability theory. Next, it will survey some of the philosophical work on the interpretation of probability. The final section will deal with the special problems encountered when probability is introduced as a fundamental element in statistical mechanics.

I. Formal aspects of probability

1. *The basic postulates*

One begins with a collection, E, called a set of elementary events. Next, there must be a collection of subsets of E, called events, that is closed under set union, intersection, and difference, and that contains E. In general, this collection, F, will not contain every subset of E. Probability is then introduced as a function from the members of F into the real numbers, P. P assigns the value 1 to E – that is, $P(E) = 1$. Most importantly, P obeys an additive property: If A and B have no members in common, then P assigns to their union the sum of the values assigned to A and B – that is,

$$A \cap B = \phi, \text{ then } P(A \cup B) = P(A) + P(B)$$

When the set E is infinitely large, F can contain an infinite number of subsets of E. In this case, the additivity postulate here is usually extended to allow for denumerably infinite unions of subsets of F. If a collection of A_i's consists of sets, all of which have no members in common whenever $i \neq j$, then it is posited that $P(\cup_i A_i) = \Sigma_i P(A_i)$. This is called countable additivity or σ-additivity. This postulate is sometimes questioned in subjective probability theory, although denying it leads to peculiar behavior for probabilities even when they are interpreted subjectively. But in all of the standard applications of probability theory to statistical mechanics, it is assumed to hold.

Next, one needs the notion of conditional probability, introduced in this approach by definition. If $P(A) \neq 0$, then the probability of B given that A, the conditional probability of B on A, written $P(B/A)$, is just $P(B \cap A)/P(A)$. This can be understood intuitively as the relative probability of B to the condition of A being assumed to be the case. Two events are probabilistically independent if $P(B/A) = P(B)$ and $P(A/B) = P(A)$. This can easily be generalized to hold for collections of events where each event is independent of all the others in the collection, rather than independence being merely between pairs of events.

A sequence of "trials" that constitute an "experiment" is an important notion. Here, one thinks of a sequence in which some elementary event is the outcome of each running of the trial. If the probability of an event on a given trial is independent of the outcomes of all other trials, the sequence is said to be a Bernoulli sequence. If the probability of an event in a trial differs, perhaps, from its probability conditional upon the outcome of the immediately preceding trial, but if this probability $P(A_n/A_{n-1})$ is then the same as the probability of A_n conditioned on all other additional outcomes, the sequence is called a Markov sequence. In a Bernoulli sequence, no knowledge of other outcomes leads to a change of one's probability attributed to the outcome of a given trial. In a Markov sequence, knowing the outcome of what happened just before the specified trial may lead one to modify one's probability for a given trial, but knowledge of earlier past history is irrelevant.

Finally, we need the notions of random variables and distribution. A random variable is merely a function that assigns a value to each of the elementary events in E in such a way that the set of all elementary events that have a value less than a specified amount is in F, and hence is assigned a probability. One can define the distribution function for the random variable as the function whose value at a is just the probability that the random variable has a value less than a. If this function is differentiable, its derivative, intuitively, gives the rate at which the probability that the random variable will have a value less than a is increasing at a. From the distribution function for a random variable, one can define its expectation. Intuitively, this is just the "sum of the product of the value of a random variable times the probability of that random variable." Naturally, for continuously distributed random variables, this requires formalization in terms of that generalization of the sum that is the integral. Crudely, the expectation of a random variable is its mean value.

The basic postulates of probability are of extraordinary simplicity. Naturally, much refinement is needed to make things precise in the general cases of infinite collections of events and so on. Although the postulational basis is simple, it is extraordinarily rich in consequences. We will need to note only a few of the most important of these consequences in the next section.

2. Some consequences of the basic postulates and definitions

A group of important consequences of the basic postulates of probability theory deal with sequences of trials that are independent – Bernoulli sequences. Suppose such a sequence of trials is performed, with a fixed probability for the outcomes of each trial and with the independence condition holding, so that for any collection of trials the probability of a joint outcome is merely the product of the probabilities of the outcomes of each trial, a condition easily seen to be equivalent to the earlier definition of independence.

One can take the outcomes of the first n trials, characterized by a random variable defined over the elementary event outcomes, add them up, and divide by n, getting a "sample mean" of the outcomes in a particular series of trials. The laws of large numbers relate these sample means in the limit as the number of trials "goes to infinity" to the expected value of the random variable as calculated by using the probability distribution for this random variable as fixed by the probabilities of the various outcomes on any one trial. This probability distribution, by the definition of the sequence as a Bernoulli sequence, is the same for each trial.

The Weak Law of Large Numbers tells us that

$$\lim_{n\to\infty} P(|y_n - \mu| > \varepsilon) = 0$$

where y_n is the sample mean of the random variable in question and μ is the expected value of it on a single trial. What this says is this: Suppose an infinite number of Bernoulli sequences of trials are run, and the sample mean calculated for each sequence at each n. For any deviation of this mean from the expected value, the proportion of sequences such that the sample mean at n, a specific trial number, will differ by more than ε from the expected mean can be made arbitrarily small by picking an n far enough out in the sequence.

This result is compatible, however, with the following: An infinite number of such infinite number of trials are run. In each, the deviation of the sample mean from the expected value becomes, as the Weak Law demands, sparser and sparser as the number of trials increases. But in each run, or in a large proportion of them, the sample mean never gets within ε of the expected value, *and stays there forever.* That this is not to be expected is given us by the Strong Law of Large Numbers. This says $P(\lim y_n = \mu) = 1$. What this comes down to is that when we look at this infinite collection of repeated trials, as the number of trials goes "to infinity" in each sequence, we can take it that the probability of a sequence getting a sample mean that settles down to within ε of the expected value forever has probability one, and the set of those that never do this has probability zero.

One can go much further. The Central Limit Theorem, for example, tells us that there is a limiting particular distribution function to which the distribution of such sample means in this infinite collection of infinite trials will converge.

These facts about large numbers of trials (or more correctly about the behavior of sample means of sequences of trials in the limit) will play a role in some of our attempts to outline just what probability is. For the moment the reader should only note how the notion of probability plays a crucial role in the statement of these results. The "convergence of the sample mean to the expected value" noted in all these results is some form of convergence "in probability." For future reference, it is important to note also how the crucial premise that allows these results to be derived is that the sequence is one of independent trials. In Chapter 5, we shall see in discussing the role of the Ergodic Theorem in attempts to provide a rationale for equilibrium statistical mechanics, that it is possible to provide a close analogue of these "large number" results even in some important cases where, intuitively, independence of trials is not the case, at least in the physical sense of trials whose outcomes are causally independent of one another. Indeed, we shall see that something like a law of large numbers can be derived in some cases even where the outcome of each successive trial is completely causally determined by the outcome of the initial trial.

One very simple consequence of the postulates and of the definition of conditional probability is Bayes' Theorem. Suppose $A_1 \ldots A_n$ constitute events in F so that each elementary event in E is in one and only one A_i. Then, for any event B in F, the following holds:

$$P(A_k/B) = \frac{P(B/A_k)P(A_k)}{\Sigma_i P(B/A_i)P(A_i)}$$

The theorem relates the probability of one of the A_i's conditional upon B to a quantity calculable from the unconditional probabilities of the A_i's and the probability of B conditional upon the specific A_i in question. We shall note in Section II,5 the central role this theorem plays in one view about the way probability assessments are to vary dynamically over time.

3. Some formal aspects of probability in statistical mechanics

The role of probability in statistical mechanics is one fraught with puzzles, many of which will be components of the central issues to be discussed in this book. Here, I want only to note a few basic aspects of the formalism by which probability is usually introduced into the theories we will survey.

The class of elementary events relevant to statistical mechanics is phase-space, the class of points representing each possible micro-state of the dynamical system in question. For the classical statistical mechanics with which we will be dealing, each point represents the full specification of generalized position and momentum coordinates for each dynamical degree of freedom of the system. The relevant phase-space for quantum statistical mechanics is of vital importance for statistical mechanics in practice, but not something we will have to deal with very often. The class of events, F, is just a class of sets of micro-dynamical states. Random variables will then be functions that assign numbers to micro-states in such a way that when the probability function over the relevant class of sets of micro-states, F, is defined, well-defined probability distributions for the random variables in question also exist.

To define one's probabilities, start with a *measure*, a generalization of the ordinary notion of volume, on the phase-space in question. The choice of this measure ultimately constitutes the determination of the relevant values a probability is to take. Justifying the choice of this measure and explaining why it works often constitute central parts of the foundational question. The standard measures are all derived from the measure on phase-space as a whole that works by taking a volume in phase-space to have its size be essentially the product of all its extensions in coordinate and momentum values. Usually one will be dealing with a lower-dimensional sub-space of the phase-space, a sub-space determined by the restriction of possible micro-states of the system to those compatible with some class of macro-quantities. The appropriate measure on this sub-space will not in general be the simple volume measure of the phase-space restricted to this lower dimension, but an appropriate modification of it. In the case of equilibrium statistical mechanics, this appropriate measure is a measure of the size of the regions in the sub-space provably invariant over time under the action of the dynamic evolution of the systems represented by trajectories from their initial representing phase-point. In non-equilibrium cases, as we shall see in Chapter 7, such measures also play a role, but one whose rationalization is rather less well understood.

The measure of regions then allows one to assign probability to appropriate regions. Here, it is important to note crucial facts about regions "of zero size," regions that are assigned zero probability. It is easy to show from the basic postulates that the empty set in F – the set containing no elementary events – must get probability zero. But having probability zero is not a property of this set alone. In general, many non-empty sets in F will have probability zero. In the ordinary measure for regions of the real line, for example, all sets of rational numbers only get measure zero. In a many-dimensional space with its ordinary volume measure, all subspaces of lower dimension will receive measure zero. In statistical mechanics, the role of sets of micro-states of probability zero is an important one, and we shall have a good deal to say about what can, and what cannot, be inferred from some class of micro-states being of probability zero.

Formally, the choice of a probability assignment to sets of micro-states that is such that probability zero is assigned to any set with zero size in the measure – an assignment of probability that is, in mathematicians terms, "absolutely continuous" with respect to the standard measure – makes the formal method for assigning probabilities simple. When the probability assignment is absolutely continuous with respect to the measure, we can derive all probabilities assigned to sets by positing a non-negative, measurable, function, f, over the micro-state points of the relevant subspace of phase-space. The probability assigned to a region A in the phase-space is then obtained by "adding up" the f values for all the points in A, or, more rigorously, by integrating the function with respect to the measure over the region A. That is, a single probability density function can be defined that "spreads the probability" over the appropriate region of phase-space so that the total probability assigned to a region is obtained by a measure of the amount of the total probability over the whole phase-space that is spread in the region in question. Most commonly in statistical mechanics it is "uniform" spreading that is posited, the probability being assigned to a region of micro-states just being proportional to the size of that region in the chosen measure. Of course, the grounds for choosing such a measure in the first place, and the explanation of why such a choice works as well as it does (when it does work), remain deep conceptual and physical issues.

II. Interpretations of probability

A single formal theory can be applied in many different domains. The same differential equation, for example, can apply to electromagnetic fields, sound, heat, or even to phenomena in the biological or social world. The postulates and basic definitions of formal probability theory are clear. But what is the domain that we are describing when we take the formal apparatus of set of elementary events, set of subsets of it called events, and additive function on these called the probability function to be characterizing what we take intuitively to be probability? As we shall see in the brief survey next, this has become one of those issues that seems at first easily resolvable, but that becomes slippery and elusive when a complete and clear explication is demanded.

1. Frequency, proportion, and the "long run"

Consider a finite collection of individuals with a number of properties such that each individual has one and only one member of the set of properties. The relative frequency with which a property is instanced in the population is plainly a quantity that obeys the postulates of probability theory. So, are probabilities in the world nothing but relative frequencies in ordinary finite populations?

One objection to this is its limitation of probability values to rational numbers only, whereas in many of our probabilistic models of phenomena in the world, we would want to allow for probability to have any value in real numbers from zero to one. But surely we can generalize beyond actual relative frequencies to actual proportions. Imagine a particle moving in a confined region that we think of as partitioned into non-overlapping sub-regions. Surely we understand the notion of the proportion of some finite time interval the particle spends in any one of the partitioning sub-regions. These proportions can, in general, have non-rational values. And it is transparent that the proportions will obey the laws of formal probability theory. The particle will spend all of its time somewhere in the region $P(E) = 1$, and for exclusive regions the proportion of time the particle spends in the joint region of A and B will just be the sum of the proportion of time spent in A and that spent in B.

But there is a standard objection to any proposal to take such clearly understood notions as finite relative frequency or ordinary proportion as being identifiable with *probability*. Our intuition tells us that although frequencies and proportions should in some sense cluster around probabilities, we cannot, in general, expect probabilities and proportions to be equal. The probability of heads on a fair coin toss is one-half, but only a small proportion of coin tossings have heads come up *exactly* half the time – and it will never be so if the number of tossings is odd. A simple identification of probabilities with frequencies or proportions, it is often argued, is too naive, too direct, and too simple-minded a theory to capture the subtle relation that probabilities hold to actual frequencies and proportions.

One familiar attempt to resolve this problem is to move to idealized "infinitely large" reference classes. Here, the attempt is made to use the close association of probability with observed frequency mediated by the laws of large numbers in probability theory to find something in the world more strictly identifiable with probability than the frequencies or proportions in finite reference classes. For example, it is sometimes argued that in the case of something like coin tossings, where only one of a finite number of outcomes is possible, the probability of an outcome ought to be identified not with the frequency of that outcome in some limited set of trials, but with the "limit of relative frequency in the long run" – that is, as the number of tosses increases without limit or "to infinity."

But such long-run relative-frequency accounts of probability are problematic in several ways, at least if they are intended as providing some definition of probability in terms that advert to actual features of the world and that themselves do not invoke probabilistic notions. One problem is that the limit of a series – the natural way to define probability in this approach being to use the usual mathematical definition of a quantity as a limit – can vary depending upon the order of the terms in the series. Although the finite relative-frequency approach to probability requires no implicit order among the events whose probability is to be determined, the limit of relative frequency in the long-run approach does demand such an order of events. But for events that are orderable in time or in some other natural way, this is not an insuperable conceptual problem.

More disturbing is the degree of unrealistic idealization that has been introduced in this "more sophisticated" approach to a definition of probability. Although the finite classes and sub-classes of the finite relative frequency view are taken as existing in the world, in what sense do the infinite sequences of events needed to define probability in the long-run approach really exist? Will there ever really be an infinite number of tossings of fair coins in the world? If not, then have we abandoned the ideal of finding something "real" and "in the world" to be the element described by the formal probability theory? If we take such infinite sequences not as real elements of the world, but as some "ideal" element, then exactly how is this ideal related to what is real (presumably actual relative frequencies in actual finite reference classes), and, given this idealization, what are we to take its probabilities to be?

Most important of all is the realization that going to long-run limits of relative frequencies doesn't really solve the problem it was introduced to handle – the possibility of deviation between relative frequency and what is taken, intuitively, to be probability. First, note that the association of probability with relative frequency guaranteed by the laws of large numbers holds only when the sequence is a Bernoulli sequence, a sequence of probabilistically independent trials. But this notion is one that requires an introduction of a probabilistic characterization of the physical situation in its definition, leading us to be quite skeptical that there is some way in which, by using infinite sequences of trials, we could "define" the notion of probability in terms that themselves do not presuppose the understanding of probabilistic notions.

A further consideration in this vein is even more important. Even given that the sequence has the proper probabilistic independence between trials, even the Strong Law of Large Numbers fails to identify the limit of relative frequency with the probability in an individual trial. We are told only that the probability that they will be equal is one, or that the probability of their being unequal goes to zero. But, as we have noted, probability zero is not to be identified with impossibility. Even if an infinite sequence existed, and even if we could be assured that it was a Bernoulli sequence, the laws of probability still leave open the possibility that the limit of relative frequency, even if it existed, could differ from the probability of the relevant outcome in an individual trial. It would seem, then, that we can associate limits of relative frequency with probabilities only relative to a modality that is itself probabilistic in nature. For this reason as well as for the others noted, many have become skeptical that the invocation of long runs, construed either realistically or in terms of some idealization, will do the job of mitigating the fundamental objection to the finite relative frequency (or finite proportion) account of probability – that is, the possibility of a divergence between the appropriate proportion, finite or long-run, and the probability we are seeking to identify with some feature of the world.

2. Probability as a disposition

One very important group of attempts to solve some of the problems encountered in the relative frequency (or proportion) and long-run relative frequency approaches is the one, beginning with seminal work of K. Popper, that offers an account of probability as a disposition or propensity. Although the frequentist takes probability as something "in the world" but attributable, at least primarily, to collections of trials or experiments, the dispositionalist thinks of a probability attribution as being fundamentally an attribution of a property to a single trial or experiment. For the frequentist, probability adheres to the individual trial only insofar as that trial is subsumable into some general kind or class, and frequentists generally take the correct probability attribution to an individual trial as something that is only relative to thinking of that trial as in some specific class or other. But such problems concerning the uniqueness or non-uniqueness of the correct probability attribution to an individual trial (or, on its relativization to thinking of that trial as a member of some general kind) are sometimes puzzles for the dispositionalist as well.

The property – or rather magnitude, for probability is taken to be a "quantity" that inheres in the individual event to some "degree" or other ranging from zero to one – that probability is taken to be is a dispositional property. Here, probability is assimilated to such properties as the solubility or the fragility of objects. Intuitively, a distinction is made between categorical properties and dispositional properties. The former are, in some sense, the "occurrent" properties of the object, the latter present in the form only of "conditional" properties, attributions that have their presence because of what the object *would* do were certain conditions met. Thus, a dry piece of salt is categorically cubical, but soluble only in the sense that if it *were* put in a solvent it *would* dissolve. Needless to say, that distinction between the categorical and dispositional properties is a hard one to pin down. Many, in fact, would doubt that any such distinction that holds in a context-independent way can be made, arguing that what is for the purposes of one discussion categorical can be for other purposes viewed as dispositional.

Dispositional theories of probability vary widely in what the disposition is attributed to and in their detailed analysis of how the attribution of probability as a disposition is to be precisely construed. We will have to survey the most general features of these accounts only briefiy.

As we have seen, the stipulation or definition of a dispositional property is usually by means of some counter-factual conditional: how the object would behave were certain test conditions satisfied. What would such an account of probability as a disposition look like? Usually this is thought of in terms of the propensity of the trial to reveal a certain relative frequency (or proportion) of outcomes on repetitions of it. Crudely, what it means to attribute a probability of one-half for heads on a coin toss is that were one to toss the coin (an infinite number of times?), a relative frequency of heads of one-half would be the result. Here, at least one problem of the ordinary relative frequency view is addressed. Even if there were no repeated trials – indeed, even if the coin were never tossed at all – it could still have a definite probability for an outcome upon tossing, for the relevant tossings are posited only in the mode of possibility.

There are many alleged problems with the dispositional account and, once again, we will only be able to note a few of them in outline. Intuitively, many think there are two distinct situations where probabilities are attributed to a kind of trial: the cases where the outcome on a single trial is determined by some underlying "hidden" parameters whose values remain unavailable to us, and the "tychistic" case where, as is frequently alleged of trials in quantum mechanics, there are no such underlying hidden parameters whose values actually determine the outcome of the trial in question. How does the dispositionalist address these two cases?

In the deterministic case, some dispositionalists would deny that any non-trivial probabilities are involved, the probability of the outcome being one or zero depending on whether it is determined that it occur or that it not occur. For these dispositionalists, the appearance of probability in deterministic cases is just that, an illusion. For them, where genuine indeterministic chance is not exemplified in the world, it is wrong to think of the outcomes as having non-trivial probabilities.

Other dispositionalists would be loath to drop the idea that even in such cases there would still be real, non-trivial probabilities. How would these be defined? By the relative frequency of the outcome that would result from repeated trials of the *kind* of experiment in question. But what would this frequency be? It would, given the determinism, be fixed by the *actual* distribution over the initial conditions that would have held. This leads to two problems. First, it seems to some that the dispositionalist account here may be parasitical on an actual relative frequency (or proportion) account, making it no real change over the latter. More importantly, wouldn't such an account be subject to the same fundamental objection as that made against the actual relative frequency account – that it would fail to do justice to our intuition that frequency and probability can diverge from one another, even in the "long run"?

What about the other case, where the outcome of a given trial is truly undetermined by any hidden parameter values? This is the case that many dispositionalists find most congenial, indeed some of them asserting that only in such cases does the trial have real non-trivial probabilities for its outcomes.

Here we have no actual distribution of underlying hidden parameters that fully determine the outcomes of the trials to rely on, nor do we have the worry that in any individual trial the "real" probability must be zero or one because the outcome, whatever it is, is fully determined by the values of the hidden parameters even if they are unknown to us. We do have, of course, the actual relative frequencies (or their limits or proportions) of outcomes in collections of trials that are (now, in all respects) like the trial in question.

Are these frequencies or proportions to be taken to be the probabilities? No. For the dispositionalist, the real probability – the dispositional feature of the individual tychistic situation – is connected to the manifested frequency, but not in a definitional manner. The real propensities are supposed to be, in some sense, causally responsible for the frequencies. The probabilities as dispositions in the individual case "generate" the relative frequencies in some sense of that term. And the manifested frequencies or proportions may be taken as evidence of just what the real dispositional probability values are. Manifested frequencies clustering about some value are an indication to us that the real probability has its value somewhere in the vicinity of the revealed frequency. But the manifested frequencies or proportions are not taken to be "constitutive" or definitive of what the probabilities are.

For many dispositionalists, the real probabilities are to be, once again, defined by the use of counter-factual locutions. The probability has a certain value if in a long run of repeated trials of the experiment in question the proportion or frequency of the value specified by the probability would now be obtained. The type of trial is now fixed as a kind by all of its features, there being no "hidden" features relative to which a further subdivision of kinds can be made into which the trial could be placed.

Yet that can't be quite right either. For, as we have seen, even the strongest laws of large numbers don't identify long-run proportion with individual case probability. Wouldn't this "slack" – a slack characterizable only in probabilistic terms – still hold even if it were possible but we were talking about non-actual runs of trials? One can imagine how to try and improve the situation from a dispositionalist stance. Think of a counter-factual being true if in some possible world like ours, but differing only in the satisfaction of the "iffy" part of the conditional, the "then" part holds. Now imagine the repeated trials being carried out in a vast number of such other possible worlds. In each such world, a certain proportion of outcomes would result. Take the disposition of probability as having a certain value if the distribution over all these possible worlds of the proportions of outcomes is as would be described by the laws of large numbers for the appropriate probability. So, in vast numbers of such worlds the proportion would cluster strongly around the proportion identical to the probability, but we could still allow for possible worlds where the frequency diverged from the probability, assuming such worlds were themselves infrequent enough.

But few are willing to have quite such a "realistic" attitude toward these other possible worlds. At this point, one begins to wonder if this "explication" of probability really is an explanation of what "in the world" probability is. Has it become, rather, a repetition of the facts about probability as described by the formal theory cast into a guise that seems to be providing an interpretation but that is now diverging from the original intent of finding some feature of the world, the actual world, that can be taken as described by the formal terms of probability theory?

3. *"Probability" as a theoretical term*

A group of important attempts to understand probability, still in the objectivist vein that we have been exploring in the last two sections, make a deliberate effort to avoid some of the problems encountered by frequentist and dispositionalist, problems they run into at least in part because of their desire to define probability in non-probabilistic terms. This alternative approach is in part inspired by views on the nature of theoretical terms in science. Early empiricist and operationalist views on the meaning of terms referring to non-directly observable entities and properties in science, wanting to legitimize the usage of such terms as they appeared in physical theories but wary of the introduction into discourse of terms whose meaningfulness was dubious because of their lack of associability with items of experience, frequently proposed that any term in science not itself clearly denoting an item of "direct experience" be strictly definable in terms of such items from the "observational vocabulary." But later empiricistically minded philosophers became dubious that our typical terms of physical theory could be so explicitly defined. Couldn't such terms be legitimized, though, so long as they played an essential role in a theoretical structure of sentences that was tied down to direct experience at some level, if only through a network of logical implications? A term, then, could be legitimate, and its meaning could be made clear, if it functioned in the overall network of scientific assertion, so long as that network as a whole led to testable empirical consequences.

The application of this general approach to probability would be to argue that attributions of probability to the world were also to be understood as functioning in a complex network of assertions, some of the components of which, including the attributions of probabilities to outcomes of experimental setups, were tied to the immediate data of observation only in an indirect way.

One component of the overall structure that fixes the meaning of probability attributions would be the rules of inference that take us upward from assertions about observed frequencies and proportions to assertions of probabilities over kinds in the world, and downward from such assertions about probabilities to expectations about frequencies and proportions in observed samples. These rules of "inverse" and "direct" inference are the fundamental components of theories of statistical inference.

Needless to say, there is no reasonable plausibility to the claim that there is a body of such rules that is universally accepted. The appropriate structure for a theory of inference to statistical or probabilistic generalizations from observed frequencies and proportions is one fraught with controversy at very fundamental levels. Even the apparently simpler problem of inferring from an accepted probabilistic assertion about a population to a legitimate conclusion to be drawn about a sample of one or more individual cases is wrapped in controversy. But the claim we are looking at here is only this: Once one has adopted some principles for making such inferences – from sample to probabilistic assertion and from probabilistic assertion to expectations for samples from the population in question – the very adoption of those rules of upward and downward inference is constitutive, at least in part, of what the probabilistic assertion means for you. The principles of "warranted assertion" that state what one will take to be legitimized inferences to and from probability assertions to assertions about frequencies and proportions in finite sample classes do at least part of the job of determining what we mean by the probabilistic assertions themselves. And fitting the probabilistic assertions into a network of inference in this way, we can "fill in" their meaning without requiring us to give an explicit definition of "probability" in terms of proportion or its limit, or in terms of these even in other possible worlds (as the dispositionalist suggests).

One puzzle with this approach is that it seems, so far, to leave out some crucial elements. Suppose two statistical schools rely on differing upward and downward rules of inference, but whose rules, when combined, lead to the same inferences from proportions in samples to proportions anticipated in other samples. Do they merely disagree on the meaning of the "intervening" statistical assertions, because they accept different such assertions on the same sample evidence, or do they have a disagreement about which upward rule of inference is correct? The latter seems to require some notion of another constraint on the truth of the statistical assertions, something that goes beyond its being the assertion warranted on the upward rule chosen. And doesn't it seem plausible to claim that the truth of the statistical assertion has at least something to do with the actual frequency or proportion of the feature in question in the total population, even if, as we have seen, a naive identification of the probability with that proportion can be challenged?

One direction in which to seek additional elements in the world pinning down the meaning we give to objectivist probability assertions is to look toward the origin of the actual frequencies and proportions in the total population in the general features of the world described by our background scientific theories. The idea here is that although some frequencies and proportions have a merely accidental or contingent aspect, others can, in one way or another, be shown to be "generated" out of the fundamental nature of the world as described by our most general and foundational lawlike descriptions of the world.

The ideas here are connected with those that arise in the philosophy of science when one seeks for a notion of a law of nature as opposed to a mere true generalization. Whereas "$F = ma$" is supposed to have a lawlike status, "all the coins in my pocket are copper" does not. The former generalization is inductively inferable from a sample of the cases it covers, the latter knowable only by an exhaustive inspection of each coin in the pocket. The former grounds counter-factual inferences about possible but non-actual situations, the latter hardly allows one to infer "if something else were a coin in my pocket it would be copper." What constitutes the lawlike force of some generalizations? Reliance on such notions as laws being true not only in the actual world but in all "physically possible worlds" just seems to beg the question.

One general approach focuses on the place the generalization holds in our hierarchical system of general beliefs about the world. Some beliefs of a generalized nature play a fundamental role in that they are the simple axiomatic beliefs from which other beliefs of great generality and explanatory importance can be derived. The idea is that lawlikeness is not a feature of a generalization having some semantically differentiable content when compared with ordinary generalizations, but rather that the lawlike generalizations are those fundamental axiomatic generalizations that ground our overall explanatory structure of the world, or those more restrictive generalizations whose lawlikeness accrues from their being derivable from the more fundamental laws.

The connection of these ideas with objective probability goes through a claim to the effect that a quantity that appears in such a fundamental lawlike generalization, and that has the formal requisites to obey the laws of probability theory, is what counts as an "objective probability in the world." The idea is that we connect up this quantity, obeying the formal postulates of probability theory and placed in the fundamental laws of nature to the world of experience, by taking its specification of probabilities to be our guide to proportions in samples of the population according to the usual probabilistic inferences. So although such a theoretical probability is a guide to frequencies in the world, and may be, at least in part, inferred to have the value it does by our observation of frequencies in the world, there is no strict identification of probability with frequency or proportion, even in the whole population in question or in that population in some dominant class of possible worlds either.

The wave-function of quantum mechanics, or rather that function multiplied by its complex conjugate, provides, in specific physical situations, a formally probabilistic distribution over the possible outcomes of an experiment. From this we can formulate our expectations not only of what proportion to find in some run of the experiment, but even of the way in which proportions themselves will be distributed over repeated runs of repeated experiments. Naturally we expect to find closer concentration of proportions and frequencies around theoretically posited probabilities as the number of trials in an experimental run increases. But we don't claim that the probability is the actual frequency even in the totality of trials, or even in that overwhelmingly large number of totalities of trials we could have run.

Something like this idea of probability will indeed fit well with much of the way in which probability functions in statistical mechanics. Needless to say, this "probability as a theoretical quantity" view is not without its puzzles as well. Some skepticism that we have genuinely fixed a unique meaning for "probability" still affects those who find general fault with the holistic idea that meanings can really be fixed by putting a term into a theoretical web that connects somewhere or other with experience, faults engendered by problems of referential indeterminism that come along with such holistic accounts of meaning. Again, the connection between theoretical assertions about probability and assertions about relative frequencies in finite samples is not completely clear here. Have probabilistic notions been presupposed when we say that we have interpreted the probabilistic assertion at the theoretical level by drawing inferences – inferences of what kind? – about proportions in finite populations from it?

One puzzle to consider is that the virtue of this approach – its ability to distinguish probability from proportion even in the total population – may be in some sense a vice as well. Could it be the case that probability and actual proportion in the entire population radically differed from one another? Of course we might doubt, in the case of wide variance of probability and frequency, that what we took to be the probabilities in the world really had the value we supposed. But from the point of view of the theoretical notion of probability we have been looking at, could it not be the case that that probability really did have the original value even though the proportion in the population as a whole radically differed from this value? Such an outcome would itself be "improbable" in that view, but not in any way impossible. But do we want to let the notion of "what probability is in the world" become *that* detached from the frequencies or proportions that actually manifest themselves? Perhaps the idea could be filled out that probability is that proportion of the total population posited as an "ideal" by the simple but very general postulates that we take to be the fundamental laws of nature.

As we shall see in Section III of this chapter, and as we shall again see a number of times, especially in Chapter 8, where the details of probability in statistical mechanics are discussed, this problem of radical divergence of probability from proportion or frequency is not merely one that occurs in an "in principle" discussion of the meaning of probability in general. Some vexing issues within statistical mechanics proper, and in its interpretation, hinge on the question of how we ought to respond to the puzzle of accepting a theory that posits a probability for an outcome while seeming to simultaneously maintain that the actual proportionate outcome of the feature in question in the world we inhabit radically diverges from the probability. The germ of this issue has already been noted in our discussion in Chapter 2 of Boltzmann's response to the problem of explaining the apparent non-equilibrium nature of the world as a whole, in the face of his contention that equilibrium is the overwhelmingly most probable state of a system.

Additional questions abound concerning the relation of probability construed as a theoretical property and the other features of the world described by our theories. Suppose we watch a piece of salt dissolve in water. Asked why the salt dissolved we could say, "because it is soluble," but we believe a much more interesting answer is available in terms of the composition of the salt out of ions, the nature of water on the molecular scale, and so on. Similarly, we might "explain" an observed proportion in nature by reference to a probability that leads to some degree of expectation for that proportion to occur in such a sample. But, once again, we feel that some other, deeper, answer in terms of the laws of nature governing the phenomenon in question, the distribution of initial conditions of hidden parameters in nature, and so on is available. What is the relation of "probability" to these more familiar features of the world as described by our scientific theories?

Some have suggested that we think of probability as a "temporary place-holder" in our theories, to be eliminated by future, deeper, theories that will dispense with it in terms of the underlying physics of the situation. So, it is suggested, "soluble" holds a place for descriptions of atomic constitution of solute and solvent and their mutual interaction, and "probability" is, similarly, a temporary and dispensable place-holder in our theory.

Much of this book will be concerned with the issues surrounding the question of just what underlying facts of nature ground the attributions of probability encountered in statistical mechanics. Whether one concludes that ultimately "probability" ought to be somehow defined in terms of these deeper physical elements of nature (as some have suggested), or, instead, that it be treated as a term "holding a place for them in theory" to be replaced by them as our deeper understanding progresses (as others have suggested), or even that it is, in its own terms, an ineliminable component of our ultimate theory (which also suggests itself as the right expectation to others), one still must get clear just what the nature of the world is, on the microscopic scale and as described by our physics, that plays the role relative to probability for thermal processes analogous to the role played by ionic constitution in grounding notions such as solubility for solids. Getting an agreed answer to this question at present will be too much to expect. The issues surrounding the importance of fundamental dynamical laws, of distributions of initial conditions, of the interaction of a system with its environment, of the nature of the means by which systems are constructed and prepared, and of the basic cosmological facts about the universe considered as a global whole in "grounding" the fundamental probability attributions needed to get the standard conclusions out of the theory remain replete with controversy.

4. *Objective randomness*

At the time attempts were being made to define objective probability in terms of limits of frequencies in long-run sequences of events, it was noticed that the order of outcomes in the sequence could be important for reasons that went beyond the dependence of the value of the limit on the order. A sequence of two outcomes coded by zeros and ones, for example, might have the limiting frequency of zeros being one-half. But suppose the zeros and ones alternated in regular order. Would we want to say that the probability of a zero outcome in a trial was one-half? If we knew the immediately previous outcome, wouldn't we be sure that the zero would or would not occur? Clearly, at least for the application of our knowledge of frequencies and their limits, some assurance that the outcomes occurred "randomly" in the sequence was essential. For some, randomness became a requisite of there being a genuine probability in the sequence. For others, it was thought to be merely a condition of its applicability. But both camps required some understanding of what randomness *is*.

Whereas some argued that questions of randomness and order ought best be considered a matter of the "subjective" knowledge of the observer, others looked for objectivistic characterizations of a sequence being random in the world. R. von Mises suggested that randomness consisted in the same limit of relative frequency holding in the original sequence and in any sub-sequence derived from it by a characterization that did not refer to the actual outcomes of the trials. For this rule to select at least some sequences as random, some restriction of the rule for selecting out sub-sequences needs to be imposed. A. Church made the important suggestion that a sequence is random if the same limiting relative frequency occurs in any sub-sequence selected by an effectively computable function generating the indexes of the selected trials. This ingenious characterization proved a little too weak, because it included as random sequences in which, for example, there were more zeros than ones in every finite sequence initiating an infinite sequence, even though the probability of zeros in the sequence as a whole was one-half. Strengthened versions of it do a better job, however.

Other ingenious characterizations of objective randomness have been developed. Some rely on the intuition that "almost all" sequences ought to be random, and that the orderly sequences ought to be sparse in the sequences as a whole. A. Wald and P. Martin-Löf have suggested that a property of randomness be defined as any property that "almost all" (sets of measure one) sequences be characterizable in a specific way (i.e. in some limited language), and that the random sequences be those with all the random properties. On this definition, the random sequences are, indeed, of measure one in the set of all sequences. But the definition moves rather far from our intuitive grasp of randomness.

Martin-Löf has developed another approach. Here, one looks at the way statisticians reject hypotheses. A hypothesis is rejected at a certain "significance level" if the outcome observed is sufficiently improbable given that the hypothesis is true. Effectively characterizable tests for randomness are described, and it is then shown that they can all be combined into a universal effectively characterizable test for randomness. The random sequences are those not rejected by this universal test at any significance level.

Another ingenious notion of randomness utilizes the notion of how long a computer program it would take to program an effective computer to generate the sequence in question. Although the length of a computer program will depend upon the programming language chosen, A. Kolmogorov has shown that the problem can be discussed in terms of a universal programming language. Intuitively, for finite sequences, the randomness of the sequence is measured by the relative length of the shortest program that can generate it, random sequences being those that, in essence, can be generated only by simply stipulating the outcomes in their order in the sequence. To a degree, the results can be extended to infinite sequences, although the natural way of so extending it fails, and several, inequivalent, alternative ways can be formulated that extend the notion in different ways. Furthermore, getting agreement between our intuition as to what is to count as random and what gets defined as random by the method of computational complexity requires our thinking of the sequence as itself being generated by some mechanism characterized in a probabilistic way. Once again this makes it implausible that we can define probability in non-probabilistic terms using the notion of objectively random sequences.

The multiplicity of definitions of randomness do not all coincide, but it is not a great surprise to discover that the initial vague intuitive notion had a number of distinct (although related) formal explications. We shall not be concerned much with these notions of randomness here. In Chapter 7,II,3, however, we will discuss in some detail other notions of randomness that have their origin in the way in which a collection of systems, each of which has a strictly deterministic evolution, can be characterized as displaying randomizing behavior when the systems are given descriptions only at the macroscopic level. That is, when we coarse-grain the phase space in the manner suggested by Gibbs, it turns out that for systems suitably constructed, the evolution of a system from coarse-grained box to coarse-grained box may generate a sequence of numbers, characterizing the box it is in at a given time, that has features closely related to the features we would expect of sequences generated in purely stochastic ways. One of these features will be the fact that "almost all" such sequences of occupation numbers for the members of the ensemble started in different initial conditions will be of the "random" sort. We shall also briefly mention later work that has shown that systems can be "chaotic" in their behavior as described by the macroscopic equations of evolution as well. That is, that the equations of hydrodynamics, for example, can have solutions that can be characterized as chaotic in nature. Work has also been done relating chaos in this sense (which comes down to ineliminable large scale deviation of future behavior on infinitesimal variation of initial state) to randomness of sequences of the kind noted in this section.

5. Subjectivist accounts of probability

Whereas objectivists look for some feature of the world that can be identified as the probability formally captured by the familiar axioms of probability theory, subjectivists focus on how our acceptance of a given probability assertion will govern our beliefs and our behavior. What is it to hold that a certain outcome in a kind of trial has a specified probability? Isn't it to hold a "degree of partial belief" that when the trial is undertaken that outcome will result? And what is it to have such a "degree of partial belief"? Can't we understand that in terms of the familiar phenomenon of betting behavior? In the face of reward and punishment for having our expectations come out confirmed or disconfirmed by experience, we will act as though some outcome will occur depending both on the gains and losses we would suffer were the outcome to actually occur or not, and on our degree of certainty that the outcome will in fact turn out to be the case. The higher the probability we hold the outcome to have, the lower the odds we will demand from the bookie (human or nature) before we will bet on that outcome being the case. Turning this around, can't we simply understand the notion of probability as being a measure of our confidence in an outcome as defined by the minimum odds at which we would bet (act as if) that outcome were really going to be the case?

Much rich and profound work has been done, from several perspectives, in trying to understand probability from this subjectivist point of view. Probability is here viewed as a measure of degree of belief fixed in its functional role in our network of psychological states by its place in an account of our action in the face of risk. One problem any such subjectivist account faces is to explain to us why, given probability as such "partial belief," probability ought to conform to the usual formal axioms of probability theory. For the objectivist, this conformity follows either directly or indirectly from the facts about proportions. But why should our partial beliefs obey the axioms, in particular the addition postulate?

One ingenious and philosophically underexplored group of arguments begins with the work of R. Cox. In these approaches, probability is a number assigned to propositions, perhaps conditional one on the other, so that we are looking for $P(i/h)$, the "probability of the inference i on the hypothesis h." The usual Boolean algebra of propositional logic on the propositions is assumed. Next, axioms are posited such as: (1) the probability of i on h determines the probability of (not i) on h; (2) the probability of i and j on h is determined by the probability of i on j and h and the probability of j on h; (3) the probability of $i \cup j$ on h (when $i \cap j = \phi$) is a continuous and monotonic function of the probabilities of i on h and j on h. These axioms of functional dependence, plus some assumptions about the smoothness (differentiability) of the functional dependence, lead to a generalization of the usual probability axioms. It is then argued that the standard form of the axioms can be obtained by making some conventional stipulations. Here, then, the formal aspects of probability are taken to arise out of our intuitions that some partial beliefs (probabilities) ought to depend in a functional way only on a limited class of others.

More familiar and more widely explored alternative justifications of the usual axioms focus directly on our betting behavior. The so-called "Dutch Book" arguments have us reflect on making bets against a bookie on the outcome of some trial. We pick our degrees of partial belief in the outcomes, and the bookie offers us the minimum odds we would accept appropriate to those degrees of partial belief. An ingenious but simple argument can show us that unless our subjective probabilities conform to the usual postulates of probability theory, the bookie can offer us an array of bets that we will accept, but that guarantee that when all stakes are collected he wins and we lose, no matter what the outcome of the trial. In order for us to avoid having such a "Dutch Book" made against us, our probabilities must be coherent – that is, accord with the usual postulates. Even if our probabilities are coherent, we might still be put into a position where we cannot win, no matter what the outcome, but we might lose or break even. If we wish to avoid that we must make our probabilities "strictly coherent" – that is, coherent and with a zero probability credited only to outcomes that are impossible. Normally in the physical situations we won't want strict coherence because classes of events that are non-empty but of probability zero are a familiar, if puzzling, part of our usual physical idealizations of phenomena.

An alternative and more general approach considers an agent offered choices of lottery tickets in which gains and losses are to occur to the agent conditional on an outcome of a trial occurring or not occurring. If the agent's preference ordering among the lottery tickets is transitive, and if some other conditions of the choices ordered being rich enough are met, then one can show that the agent acts "as if" he had a subjective probability over the trial outcomes that obeyed the standard axioms of probability and that was uniquely determinable by his preferences, and as if he had a valuation or utility or desirability function over the rewards and losses unique up to a linear transformation. It goes something like this: Suppose the agent always prefers x to z whenever x is preferred to y and y to z. Suppose also that if the agent prefers x to y, he will also prefer a given probabilistic mixture of outcomes involving x to the identical mixture with y substituted for x. And suppose there is a sufficiently rich collection of preferences. Then we can assign a probability to each outcome, p, and a utility to the gains or losses incumbent upon the outcome occurring or not occurring, u_1 and u_2, such that if each lottery ticket is assigned an "expected value," $pu_1 + (1 - p)u_2$, then the preference ordering among lottery tickets instanced in the agent can be reproduced by ordering the lottery tickets in terms of their expected value. It will be "as if" the agent assigned a unique probability to the trial outcomes consistent with the laws of probability theory, and a utility to the gains and losses, unique up to a linear transformation (which amounts to picking one gain or loss as "neutral" and fixing a scale unit for units of desirability), and then ordered his preference for the lottery tickets by calculating their expected values.

This approach to subjective probability is important as a component of various "functionalist" accounts as to what partial beliefs and desirabilities are, as well as a crucial component of the theory of rational decision making. Naturally the full theory is a very complicated business, involving many difficulties in associating the "psychological" states to behavior, and in formulating a theory of normative rational action immune to intuitive paradox. But the overall strategy for showing why probabilities, if they are to be taken as measures of partial belief, ought to obey the usual probability axioms (as a consequence of "rationality of choice as evinced by transitiveness") is clear.

An alternative derivation of subjective probabilities starts with the notion of comparative believability – that is, with the idea of a relation among propositions that, intuitively, can be thought of as one proposition being "as worthy of belief" as another. Intuitively plausible axioms are imposed on this notion of comparative believability. For example, we may demand transitivity, so that if A is as believable as B and B as believable as C, it follows that A is as believable as C. Additional constraints can be placed on the notion of comparative believability that are sufficient to guarantee that one's believability structure can be represented by a probability assignment to the propositions. Such a probability representation assigns real numbers between zero and one (inclusive) to the propositions. And the assignment is such that A will be as believable as B just in case the probability number assigned to A is at least as great as that assigned to B.

The subjective theory of probability is concerned not only with our holding partial beliefs, but with our changing them in the face of experience. How should we modify our distribution of partial beliefs in the face of new evidence? The usual rule suggested is conditionalization. We have, at a time, not only probabilities for propositions, but conditional probabilities as well, the probability of h given e, whenever the probability of e is non-zero. Suppose we then observe e to be the case. Conditionalization suggests that we take as the new probability of h the old conditional probability it had relative to e. This rule has been nicely generalized by R. Jeffrey and others to handle cases where we don't go to e as a certainty, but instead take the evidence as merely modifying the older probability of e. Conditionalization is a conservative strategy. It makes the minimal changes in our subjective probability attributions consistent with what we have learned through the evidence, and it generates a new probability distribution as coherent as the one we started with.

Much effort has gone into giving a rationalization for changing probabilities by conditionalization as persuasive as the standard rationales for having subjective probabilities at a time obey the usual formal axioms. An argument reminiscent of the Dutch Book arguments and due to D. Lewis has the agent confronting a bookie and making bets on the outcome of one trial, and then additional bets on other outcomes of other trials should the first trial result in one specific outcome or the other. Only if the second set of bets is made on odds consistent with conditionalization on the outcome of the first trial can a compound Dutch Book situation be avoided by the agent. P. Teller has shown that conditionalization follows from demanding that one's new probability distribution has $P(h) = P(k)$ after the evidence e has come in if $P(h) = P(k)$ before the observation and if h and k both implied the evidence assertion, e. And B. van Fraassen has shown that conditionalization is the only method of changing one's probability distribution that will make the probability dynamics invariant under reclassifications of the elementary events into event classes. A variety of other rationalizations for conditionalizing can also be given. On the other hand, because conditionalization is a conservative procedure, intuitively changing one's subjective probabilities only to the extent that the change is directly forced by the change in the probability attributed to the evidence induced by the result of the trial, there could be circumstances where the subjectivist would doubt its reasonable applicability, say if the new evidence indicated the agent's earlier irrational state of mind when the initial probability distribution was chosen.

In Chapter 7,III,3, we shall see that some of those who advocate what they call a subjectivist approach to probability in statistical mechanics seem to evade the strictures on conditionalizing with respect to all known evidence in some of their uses of probability attributions.

Suppose one accepts the legitimacy of probability as interpreted as a measure of partial belief, in the manner we have outlined. What, from this subjectivistic perspective, is the place of the version of probability as "a feature of the objective world" in one of the guises previously outlined? Here, defenders of subjective probability take a number of different positions. Some would allow for their being two kinds of probability (often labeled by subscripts as probability$_1$ and probability$_2$), objective and subjective, just as most defenders of objective probability are perfectly happy to countenance a legitimate subjectivist interpretation of the formalism as well.

The advocate of objective probability as actual short or long-run frequency or proportion will usually think of subjective probabilities as estimates on our part of the real proportions of the world. Naturally he will seek principles of rational inference to and from proportions in samples to those in populations, and hence to and from proportions in samples to subjective probabilities. The believer in objective probability as a propensity of an individual trial situation, it has been suggested by D. Lewis, will hold to a "Principal Principle," that if the chance of an outcome on the trial is taken to have a certain value, the subjective probability of that outcome must have the same measure. At least this will be so if the propensity chance is one that supposes the absence of underlying hidden variables that would, if known, change the propensities assigned. Subjective probabilities are, for some of the probability-as-a-theoretical-feature proponents, the basic interpretive device by which the theory structure fixing objective probabilities gets its connection to the world. It has been suggested, for example, by S. Leeds, that we can give an interpretation to the state-function of quantum mechanics by simply taking the rule that the values thought of as probabilities upon measurement computed from it are to be understood as subjective probabilities. From this point of view, subjective probabilities generated by a theory so interpreted partially fix any meanings for objective probabilistic expressions appearing in the theoretical network, leaving it open, of course, that these objective probabilities may be further pinned down by their association with non-probabilistically characterized features of the world, or even eliminated in terms of them. (The "fairness" of the coin that leads to subjective probabilities of one-half for heads and tails upon tossing, for example, would be replaced by its structural symmetry, and perhaps the distribution in the world of initial conditions over tossings, which underlie its being a fair coin.)

Other subjective probability theorists find no room whatever for the notion of objective probability as we have been construing it. For them, probability *is* subjective probability. Not that there aren't, of course, frequencies in the world, maybe even long-run limits of them. And, of course, there are also all those other features of the world, such as the balance of the coin, the distribution of initial conditions, the quantum state of the electron, that are causally connected to the frequencies we observe. But, from this point of view, there is no need to think of probability itself as anything over and above degree of partial belief. Our partial beliefs may very well be determined for us by various beliefs we have about frequencies and structures in the world, but there is no need to think of probability as being anything itself "in the world" generated by or identifiable with these familiar objective features of things.

A very interesting proposal in this vein, initiated by B. de Finetti, tries, from a subjectivist perspective, to explain to us why we might be tempted to think of "objective probability in the world" when no such concept can be made coherent (as the usual objectivist intends it, at least). Imagine a sequence of tossings of a coin the objectivist thinks of as biased – that is, of having some definite propensity to come up heads that might not have the value one-half. The objectivist thinks of us as learning from experience. We observe a long run of trials, and on the basis of the observed proportion of heads in the trials come to an estimate of the real propensity of the coin to come up heads, the objective probability of heads. How can a subjectivist understand this learning from experience without positing objective probabilities?

De Finetti asks us to consider someone with a subjective probability distribution over finite sequences of outcomes of trials of the coin tossing. Suppose this subjective probability distribution is exchangeable – that is, suppose the subjective probability given to a sequence of heads and tails is a function only of the proportion of heads in the sequence and is independent of the order of heads and tails. Then, de Finetti shows, this agent's probability distribution over the sequences can be represented "as if" the agent took the sequences to be generated by independent trials of tossing a coin that is biased to some particular propensity for heads, but with an unknown bias. It will be "as if" the agent generated his subjective probabilities for the sequences by having a subjective probability distribution over the range of possible biases of the coin.

Furthermore, let the agent modify his subjective probability distribution over the sequences by conditionalizing on the observed tossings of the coin. Then, if the agent's original distribution was such that when represented as a distribution over propensities or biases it gave no propensity probability zero or one, the evolution of the agent's probability distribution over sequences can be represented as the agent having his subjective probability converge on an ever narrower range of biases, indeed converging to a propensity equal to observed relative frequency in the limit as evidence tossings go to infinity. And two such agents will then behave "as if" they started with different subjective probabilities over biases, but, learning from experience, both converged to the "real objective propensity" in the long run. For de Finetti, of course, there is no such "real" probability of heads for the coin. All that exists are the convergences to observed relative frequency, convergences themselves dependent on the particular structure of initial subjective probabilities of the agent (these being exchangeable) and on the agent's learning from experience by conditionalization.

This result of de Finetti's is generalizable in interesting ways. The key to the representation theorem is the symmetry of the agent's initial subjective probability, the probability given to a sequence being invariant over any permutation of heads and tails in the sequence, so long as the number of heads and tails remained invariant. A general result shows that such symmetries in the agent's subjective probability will lead to his acting as if he believed in an objective propensity to which observed relative frequencies would converge. Even more general results can be proven to show that agents who start in agreement about which sets of outcomes receive probability zero, and who modify their probabilities in a conditionalizing manner, will converge on identical probability distributions "in the limit." These results constitute important formal facts about subjective probabilities that are closely analogous to other results usually interpreted in an objectivist vein. We will discuss these related results in the context of the equilibrium theory of statistical mechanics in Chapter 5 when we examine the important Ergodic Theorem of classical equilibrium statistical mechanics.

An important concept, emphasized by B. Skyrms, of "resiliency" should also be noted in this context. A subjective probability is, basically, resilient if the agent would not modify the value he gives the outcome in question (or at least would not modify it much) in the light of additional evidence. Resiliency, then, is closely analogous to the kind of objectivist demand that "real" probabilities be those that are genuinely tychistic, so that no exploration of hidden variables could divide the kind of trial in question into sub-classes in which the probability of the outcome would differ from the probability in the class as a whole. Resiliency is a kind of subjectivistic "no hidden variables" for the probability in question. It can be argued that what appear to be objective probabilities, can be, at least in part, reconstrued from the subjectivist point of view as resilient subjective probabilities. Again, this subjectivistic notion is related to a kind of objectivist resiliency (metric indecomposability) that we will explore in Chapter 5.

6. *Logical theories of probability*

The subjectivist places the restraint of coherence on an agent's initial probability distribution. If we take conditionalization to be justified by the arguments for it, then additional constraints exist on how subjective probability distributions are constrained by rationality to change in the light of new evidence. But is there any further constraint of rationality on an agent's initial probability distribution, on his probability distribution "in the light of no evidence," or, as it is frequently designated, on his "a priori probability?" Pure subjectivists frequently answer "no," coherence being given, that one a priori probability distribution is just as rational as any other.

Others deny this. "Objective Bayesians," as they are sometimes called, maintain that there are indeed further constraints of rationality that can be imposed on an a priori probability distribution. Their claim is closely related to the claims made by those who would offer what was once thought of as an alternative interpretation of probability. Deductive logic presents us with a relationship among propositions, in particular with the idea that the truth of one proposition can assure the truth of another. A premise can entail a conclusion, so that if the premise is true, so must the conclusion be true. Could there not be a similar but weaker relationship among propositions so that the truth of one, although not guaranteeing the truth of the other, could provide some degree of assurance, less than certainty perhaps, that the other was true as well? Qualitative relationships of this kind were explored by J. Keynes and others. Later attempts – in particular, the extended program of R. Carnap – sought to develop a quantitative theory of so-called "inductive logic."

But how can the "degree of support" or "degree of confirmation" of one proposition grounded in another be determined? First, it is usually assumed that these degrees of confirmation must obey the usual formal postulates of probability theory. In his later expositions of the work, Carnap took the line that these probabilities were to be thought of, ontologically, in the subjectivist vein. Then the rationale for holding them to be constrained by the familiar laws of probability theory would be the rationales for subjective probabilities being coherent that were discussed in the last section. But these constraints were plainly insufficient to uniquely fix the degree of probability of one proposition relative to another, and further constraints were sought.

A natural way of viewing the problem is to think of propositions as classes of possible worlds, the proposition being identified with the class of possible worlds in which it is true. The degree of confirmation of proposition *h* on proposition *e* could be thought of, then, as the measure of the proportion of worlds in which *e* is true and in which *h* is true as well. We can get the results we want, a logical probability obeying the formal theory of probability, by distributing an initial probability of one over the distinct possible worlds. The naive way of doing this that suggests itself – letting each possible world have an "equal share" in the total probability – leads to an inductive logic in which we don't learn from experience. A subtler method of first dividing the probability evenly over a class of kinds of worlds, and then evenly over the worlds in the kinds (kinds that don't have equal numbers of individual possible worlds in them), gives probabilistic weighting to "orderly" worlds, and leads to an inductive logic that raises our expectation that a property will be instanced as we experience its instancing in our observed sample, a kind of inductive projection from the observed into the unobserved.

Trying to rationalize a unique confirmational measure as the only one justifiable, or to find criteria of adequacy for a probabilistic measure that will pin down the possibilities for confirmation functions to a small number of choices, and then to rationalize those criteria of adequacy, is a task that Carnap only partially achieves to his own satisfaction. He relies frequently on "intuition" even to get that far. And finding a way of extending the method originally designed for simple and finitistic languages to the richer realm of higher order languages and to worlds with an infinite number of individuals proves problematic as well. But, most importantly for our purposes, another hidden difficulty with the program exists. Notice, though, that if we had achieved the program to our satisfaction, the problem of a rational constraint on an agent's a priori probability distribution would be solved. Each proposition has its "logical probability" relative to any other proposition, and indeed to the empty set of propositions. This "probability of *h* relative to null evidence" would then be the rational probability with which to hold a proposition before any evidence came in. Probabilities after the evidence could then all be arrived at by conditionalization, using the a priori probabilities to compute, in the usual way, the conditional probabilities needed to know how to modify one's initial probability as new evidence was uncovered.

The basic rule invoked for determining the a priori probabilities – treat all symmetric propositions equally when distributing probability – is a modern instance of one of the oldest ideas in probability theory, the Principle of Indifference. The idea is that "probabilities are to be taken as equal in all similar (or symmetric, or like) cases." A priori, heads has a probability of one-half, there being only two possible symmetric outcomes. Later tossings of the coin might indicate bias and lead us to modify the probability we give to heads, but it is the symmetry of cases that provides the a priori probability where we start. But *why* should we believe, a priori, in equal probabilities for the symmetric case?

In fact, the Principle of Indifference has more problems than its lack of apparent rationalization. It is, without further constraint, incoherent. For it depends in its probability assignment upon *how* the possible outcomes are categorized or classified. A die has six faces, so the a priori outcome of a one coming up is one-sixth. But the die can either come up with a one or else with a "not-one." So there are *two* cases, and by the Principle of Indifference, the probability of a one coming up ought to be taken to be one-half. In this case, we might resort to the fact that "non-one" can be decomposed into five cases, each of which is, in some sense, indecomposable, leading us back to the natural one-sixth and away from the counter-intuitive one-half. But what if the number of possible outcomes is infinite? Here, each indecomposable outcome has probability zero, and each interesting class of these (an interval on the real line, for example) has an infinite number of elementary outcomes contained in it.

This dependence of the a priori probabilities on the way cases are categorized has been emphasized by what are frequently generically referred to as "Bertrand's Paradoxes." Imagine, for example, a container with a shape so that the surface area on the inside that is wetted varies non-linearly with the volume of the container that is filled. An a priori probability of "amount of container filled" that distributed probability uniformly over the possible volumes of fluid would give radically different results from one that distributed probability uniformly over allowable interior surface area wetted by the fluid. What principle of rationality, designed to fix the unique one of the coherent possible a priori probabilities as the rational one for an agent to adopt, will tell us which categorization of the possible outcomes to choose? Only given a principle for selecting the right categories in which to view the problem can we then apply a Principle of Indifference or symmetry over those possibilities so construed.

H. Jeffreys initiated a program of selecting from among the categorizations by examining the invariance of probabilistic conclusions under various transformations of the data. Sometimes it seems as though we would want our probabilistic results to remain invariant under certain transformations. We feel, for example, that in some cases picking a designation of one value of a quantity as zero point as opposed to another, or picking one interval as unit scale as opposed to another, should not modify our expectations. Under those circumstances, one can sometimes fix upon a single categorization, with its associated uniform probability distribution militated by the Principle of Indifference so applied. But, as we shall see in Chapter 5,III,5 and 7,III,3 when we examine attempted applications of the Principle of Indifference (or its modern day reformulation, the "Maximum Information Theoretic Entropy Principle"), such a rationalizable invariance rule to fix our a priori probabilities is only rarely available to us. And when it is, what we can obtain from it may be less than what we would like.

When investigators in the foundations of statistical mechanics allege that they are understanding probability in statistical mechanics as subjective probability, it is usually, rather, a belief on their part that there is a legitimate applicability of the Principle of Indifference to physical situations that can be applied to ground the positing of initial probabilities so essential to getting what we want in statistical mechanics that is the core of their position.

III. Probability in statistical mechanics

A great deal of Chapters 5 through 9 will be directed toward problems that, at least in part, will be explored by examining in some detail the role played by probabilistic assertions in the description and explanatory account of the world offered by statistical mechanics. It will be of use here, though, to give a preliminary survey of how probability attributions are embedded in the statistical mechanical picture of the world, and of some of the peculiarities of probability in statistical mechanics that lead to difficulties in fitting an account of its role in that theory into one or another of the philosophical views about probability that have been outlined in this chapter.

In our survey of the history of the foundational problems of statistical mechanics we saw how attempts were made to account for the equilibrium state as being the "overwhelmingly most probable" state of a system. Here, we view the system as subject to some fixed macroscopic constraints (say volume and pressure for gas) and look for some justification of the claim that of all the microscopic dynamical states of the system compatible with these macro-conditions, with overwhelmingly great probability they will correspond to the state of the gas being at or near its equilibrium value of some other macroscopic constraint (say the temperature of the gas).

In trying to understand the non-equilibrium behavior of a system, we imagine it as having been prepared out of equilibrium, again subject to some macroscopic conditions. We then try to show that the combination of a reliance upon the detailed facts about the microscopic constitution of the system, the facts about the dynamical laws governing these microscopic constituents, and some other assertions framed in the probabilistic guise can lead us to the derivation of a description of the approach to equilibrium of the system as characterized by the macroscopic hydrodynamic equations of evolution appropriate to the system in question. In some accounts, the probabilistic assertions will be about probabilities of initial microscopic states of the system compatible with its initial macroscopic condition. In other accounts, they will be about the probability of future interactions of the micro-components with one another. Once again, the invocation of the probabilistic claims is one that cries out for justification.

As we shall see in Chapters 5 and 7, the justification needed for the probabilistic claims will vary quite radically depending on just what the claim under consideration is. And getting clear on that will sometimes require some serious preliminary conceptual clarification. At this point I will make only a few general remarks about some kinds of conceptual puzzlement we will run into. Here I will be concentrating, in particular, on those issues that will be problematic because they introduce features of probabilistic description of the world in statistical mechanics that most directly confront the various philosophical accounts of probability we have just seen outlined. The kinds of problems noted here are not meant to be either exhaustive or exclusive of one another. Some of the puzzles noted here are closely intertwined with one another, and an acceptable understanding of the world will require a grasp of issues that does justice to answering a number of questions simultaneously.

The origin and rationalization of probability distributions. Some problems arise out of the questions concerning the physical ground on which a given correct probability attribution rests. This is intimately connected, of course, with the questions concerning the rationale we can offer if a claim we make that a probability distribution ought to take a certain form is challenged. The problem is complicated in statistical mechanics by the fact that probability distributions are introduced on quite different grounds, with quite different rationales, and for quite different purposes in different portions of the theory.

In equilibrium theory, or at least in the version of it that seems clearest and most defensible, the aim will ultimately be to show that a unique probability distribution can be found that satisfies a number of constraints. One of these constraints, time invariance, comes from the very notion of equilibrium itself. The other, absolute continuity with respect to the usual phase space measure, is a constraint harder to justify, assuming as it does a portion of the original probability attribution we meant to justify. The rationalizing ground is sought in the constitution of the system and in the dynamical laws of evolution at the micro-level. The role played by this probability distribution in our description and explanatory account of the world is itself somewhat problematic.

In non-equilibrium theory, the understanding and rationalization of one probabilistic postulate of the usual theory plays a role in attempts to ground the posits that are the immediate descendants of Boltzmann's *Stosszahlansatz.* The kinetic equations of non-equilibrium theory are usually obtained by one version or another of some rerandomizing posit governing the interaction of the micro-constituents of a system. But the dynamic evolution of a system, and hence even of an ensemble of systems (or probability distribution over systems subject to a common macroscopic constraint), is fully fixed by the dynamical laws of evolution governing the micro-components of the systems. So a rationalization of the probabilistic posit here will come down to an attempt to show the *consistency* of the probabilistically formalized rerandomization posit with the accepted dynamical laws and the facts about the constitution of the systems. Here, probabilistic postulation is thought of much as a device to generate the correct result (the kinetic equation of evolution for a reduced description of the ensemble), a device to be instrumentalistically justified by reference to the real laws governing the dynamical evolution, and hence not directly a fundamentally independent component of the full description of the facts of the world we are seeking.

But another role for probability distributions in non-equilibrium theory is even more fundamental to the theory. This is the appropriate distribution to impose over the microscopic initial conditions of a system started in non-equilibrium subjected to some macroscopic constraints. In some accounts, the rationalization of rerandomization will itself depend upon this distribution. And if we want a full explanatory account of the structure of the approach to equilibrium of a system (its relaxation time, the form of its equation of evolution toward equilibrium on a macroscopic scale, and so on), it seems clear that such an assumed distribution will need to be posited. But here, much is still opaque in the theory. There is, in fact, and as we shall see, no universally accepted account of even what such distributions should be taken to be, much less a universally agreed upon account of their origin and rationalization. Microdynamical laws, constitutions of systems, places of systems in interacting external environments, modes by which systems are prepared, a priori distributions determined by principles of general inductive reasoning, and cosmological facts have all been invoked in one or another of the competing accounts of the physical basis of these essential initial probability distributions over possible micro-conditions.

Probability and tychism. We have outlined a debate between those who would take probability to be a mode of description of a world that might be underlain by a fully deterministic picture – that is, who would take non-trivial probabilities to be consistent with an underlying description of the situation that would so bring factors into account that with their full specification, the outcome would either definitely occur or not occur – and those who would hold that probability exists in the world only if there is a genuine failure of determinism – that is, only if such hidden factors fail to exist.

Quantum mechanics, with its "no hidden variables" theorems, has typically provided the context in which pure tychism and irreducible probabilistic descriptions not underpinnable by determinism at a lower level are defended as the correct picture of the physical world. Should the initial probabilities over initial conditions in non-equilibrium statistical mechanics be construed in a similar manner, or should they be viewed as merely the accounting of actual distributions over the underlying possibilities, one of which constitutes the real and fully determining condition in any particular instance of a physical system?

As we shall see in Chapters 7,III,6 and 9,III,1, this is a controversial matter. Particularly interesting in this context are attempts to argue that the probability of statistical mechanics is neither the "pure chance without hidden variables" of quantum mechanics, nor the "distribution over fully determining hidden variables" of the usual "ignorance interpretation" variety, but, rather, a third kind of probability altogether. The argument will be that the instability of the dynamic evolution of the systems with which we are concerned in statistical mechanics makes the characterization of the system as having a genuine microscopic dynamical state a "false idealization." The arguments to be offered rest upon the contention that no matter how close we get to any initial condition in systems of these kinds, we will find other initial conditions whose future evolution, as described by the familiar trajectory in phase space starting from this other condition, will diverge quickly and radically from the trajectory starting from the first initial state we considered. Under these conditions, it will be alleged, it is misleading to think of any individual system as really having a precise initial "pointlike" dynamical state and "linelike" dynamical trajectory. Rather, it will be argued, the individual systems in question ought to have their states characterized by probability distributions over phase-space points and their evolution characterized in terms of the evolution of such distributions over phase-space.

This view is like the "pure tychism" view in taking the probability distribution to be irreducibly attributable to the individual system, but unlike the "no hidden variables" view in that the grounds for denying any further deterministic specifiability of the system are quite unlike those that are used to deny hidden variables in quantum mechanics. In the latter case, we are presented with alleged proofs that the statistics in question (those predicted by quantum mechanics) cannot be represented as measures over underlying phase-space regions of points. But in the statistical mechanical case, there plainly is such a representation of the probability distribution. It is worth noting that although we will be discussing this issue in the context of classical statistical mechanics, the view takes on exactly the same aspects in the context of a statistical mechanics whose underlying dynamics is quantum mechanics. The "irreducible probability" in both cases is *sui generis.* In the quantum mechanical case, it is quite independent of the pure tychism that might be alleged to hold of the underlying dynamics.

Applying the posited probability distribution. Only occasionally are probability distributions used directly to determine observable quantities in statistical mechanics, although they sometimes do function in that way as in predictions of molecular velocity distributions and attempts to confirm these predictions by direct surveys of velocity distributions in samples of the population of molecules. More commonly, the probability distribution is used to calculate some value that is then associated with a macroscopically determinable quantity.

Care must be taken, however, to make sure that the quantity calculated can plausibly play the role demanded of it in the explanatory-descriptive account in which it is to function. Sometimes it is easy to lose sight of what one was interested in showing in the first place, and to think that a goal has been accomplished when a quantity is derived having the formal aspects sought, but whose role in the overall account is less than clear. Often a precise disentangling of the notions involved, and the application of derivable connections between quantities will go a long way to clarifying the structure of the theory. Thus, in the equilibrium theory the careful distinction must be made between time averages of quantities, computed using a probability distribution over the reference class of time states of a given system, and averages for these quantities taken over all possible micro-states of different systems compatible with given constraints and relative to some probability measure over these possible states. These latter averages are called phase averages. Also important is the distinction of the latter averages from most probable values of the quantities in question (calculated using the same set of available micro-states and the same probability distribution). The justifications of asserted relations among these values by means of Ergodic Theory and by means of results dependent upon the vast numbers of degrees of freedom of the system, play important parts in clarifying the formal and philosophical aspects of equilibrium statistical mechanics.

In some cases, it remains a matter of controversy which probabilistically generable quantity is the correct one to associate with the macroscopic feature of the world we are trying to understand. For example, in the non-equilibrium theory, most accounts of the statistical mechanical regularity to be associated with the monotonic approach to equilibrium described by thermodynamics follow the Ehrenfests in associating the curve of monotonic behavior of the macroscopic system with the concentration curve of an ensemble evolution – that is, with the curve that goes through the "overwhelmingly most probable value of entropy" at each time. And these accounts agree with the Ehrenfests that this curve should not be thought of as representing in any way a "most probable evolution of systems." But other statistical mechanical accounts do in fact think of the monotonic approach to equilibrium as being representable in statistical mechanics by just such an "overwhelmingly most probable course of evolution." Such conflicts of understanding point to deep conflicts about the fundamental structure of the theory insofar as it is designed to allow us to understand montonically irreversible thermodynamic behavior.

Probability versus proportion. In our discussion of foundational views on probability, we noted important doubts that probability could, in any simple-minded way, be identified with actual frequency or proportion in the world. Indeed, even moving to a view of proportion in possible worlds, doubts remained that one could define probability as proportion – even in the "long run."

Statistical mechanics presents us with additional puzzles concerning the relationship between probability and proportion. These are exemplified by the paradox considered by Boltzmann and discussed in Chapter 2: If equilibrium is the overwhelmingly probable state of a system, why do we find systems almost always quite far from equilibrium?

We will need to consider a variety of possible answers to questions of this sort. One might be to deny the correctness of the claim that equilibrium really is overwhelmingly probable. Another might be to deny that probability has anything to do with realized proportion, something not too implausible from some subjectivist or logical probability viewpoints. Still another group of attempts at resolving such a puzzle might be to argue that we have simply looked at too small a reference class in seeking the proportion in the world to associate with the asserted probability. Such a suggestion is that offered by Schuetz and Boltzmann to the effect that our region of the universe is only a small portion of its extent in space and time, and that if we looked far and wide enough in space and time we would indeed discover equilibrium to be dominant in proportion in the universe as a whole. Modern variants of such a "cosmological way out" will, as we shall see in Chapter 8, suggest that we would need to look far and wide indeed to find the appropriate proportion to identify with the probability of our theory. We shall also need to examine those modern variants of Boltzmann's suggestion as to why proportion differs so wildly from probability in our local region of the larger world – that is, updating his argument that such a deviation from the norm ought to be expected by us for the simple reason that in the dominant regions of equilibrium, no sentient observer could exist to experience this dominant proportion of equilibrium in his neighborhood.

Idealization and probability. In exploring how physical theories deal with the world, we are accustomed to the fact that our theories deal only with "idealized" systems. In mechanics, we talk of systems composed of point masses, or of frictionless systems, or of systems in which all interactions save one can be ignored. Idealization plays a prominent role in statistical mechanics, sometimes in ways that require rather more attention than a mere nod to the fact that the results hold exactly only in the ideal case.

Some of the idealizations we will encounter include going to the thermodynamic limit of a system with an infinite number of degrees of freedom, or to the Boltzmann–Grad limit in which the relative size of molecule to scale of the system becomes vanishingly small, or to the limit as time "goes to infinity" invoked in the ergodic theorems of equilibrium statistical mechanics or in the mixing results of the non-equilibrium theory. In some cases, the role the idealization is playing will be clear and relatively uncontroversial. Thus, for example, the role played by the thermodynamic limit in allowing us to move from average values of some quantities to the values being identified with the overwhelmingly most probable values of these quantities will be rather straightforward. In other cases, however, the place of idealization in the overall scheme can be quite controversial. One derivation of the kinetic equations of non-equilibrium – that of Lanford, for example – makes a radical use of the Boltzmann-Grad limit to obtain results puzzlingly incompatible (at least in a conceptual way) from those obtained by the "mixing" type results. And these latter rely on the "time going to infinity" idealization for their accomplishments in a way objected to by the proponents of the Lanford scheme. Which idealization ought to be taken as "really representing how things are in the actual world" is by no means clear.

Here, I want only to note how the very concept of probability introduced in statistical mechanics is sometimes heavily determined by the idealizing context. Take, for example, what is sometimes called a probability in equilibrium statistical mechanics viewed from the perspective of the Ergodic Theorem. Here, one starts off with systems idealized with reference to their structure, molecules interacting only by collision, and then perfectly elastically, for instance, and a system kept in perfect energetic isolation from the outside world. Next, one shows that for "almost all" initial conditions the proportion of time spent by a system with its representative point in a phase-space region will be proportional to the "size of that region in a standard measure. But one shows this only in the limit as time "goes to infinity." There is of course nothing wrong with such a notion. Indeed, it fits nicely with those accounts of probability that, we saw, emphasized proportion in the limit of an infinite number of trials. But the relation of such a probability to such things as the actually observed behavior of systems in finite time intervals is only indirect. And unless one is careful to keep in mind the essential role played by the idealization, one can think that more has been shown than is actually the case.

IV. Further readings

An excellent elementary introduction to the theory of probability is Cramér (1955). Feller (1950) is a classic text on the subject. The axiomatic foundations are explained in Kolmogorov (1950).

A survey of philosophers' ideas on the nature of probability can be found in Kyburg (1970). For the dispositional theory, a good source is Mellor (1971). Kyburg (1974) surveys many variants of the dispositional account. For the view of probabilities as "theoretical quantities," see Levi (1967) and (1980) and Fine (1973). Cox's derivation of probability is in Cox (1961). See also, Shimony (1970).

Fundamental papers on the subjectivist approach to probability are in Kyburg and Smokler (1964). The relation of subjective probability to comparative believability can be found in Fine (1973).

For one version of a "logical" theory of probability, see Carnap (1950). For other versions, so-called "objective Bayesianisms," see Jeffreys, H. (1931) and (1967), Rosenkrantz (1981), and Jaynes (1983).

On Humeanism and its opponents, see Chapter V of Earman (1986). Lewis (1986), Chapter 19, is important on the relation of subjective and dispositional probability.

For an introduction to the variety of ways of characterizing "objective randomness," see Chapter VIII of Earman (1986). Fine (1973) has detailed and comprehensive treatments of the major approaches to this issue.