



The Many Faces of Entropy*

HAROLD GRAD

1. Introduction

One of the reasons for the bewilderment which is sometimes felt at an unheralded appearance of the term entropy is the superabundance of objects which bear this name. On the one hand, there is a large choice of macroscopic quantities (functions of state variables) called entropy, on the other hand, a variety of microscopic quantities, similarly named, associated with the logarithm of a probability or the mean value of the logarithm of a density. Each one of these concepts is suited for a specific purpose. More confusing, however, than the lack of imagination in terminology is the fact that several of these distinct concepts, different in meaning and in numerical value, may be significant in a single problem. This remark is the theme of this paper. A given object of study cannot always be assigned a unique value, its "entropy". It may have many different entropies, each one worthwhile. The proper choice will depend on the interests of the individual, the particular phenomena under study, the degree of precision available or arbitrarily decided upon, or the method of description which is employed; and each of these criteria is largely subject to the discretion of the individual.

The fertility of this concept is in large part due to its flexibility and multiple meanings. On the other hand, much of the confusion in the subject is traceable to the ostensibly unifying belief (possibly theological in origin!) that there is only one entropy. Although the necessity of dealing with distinct entropies has become conventional in some areas, in others there is an extraordinary reluctance to do so. The widespread misconception that there exists a paradox in classical statistical mechanics associated with the name Gibbs, and the frequent difficulties bound to the classical reversibility-irreversibility dichotomy are directly traceable to this source.

It does not seem to be possible to give a precise mathematical definition of entropy or to create an abstract mathematical structure which is general enough to include all of the interesting applications. The strictly conservative mathematical attitude is to examine each situation separately. But it

*The work presented in this paper is supported under Contract AT(30-1)-1480 with the U. S. Atomic Energy Commission. Reproduction in whole or in part is permitted for any purpose of the United States Government.

will be illuminating to try to synthesize, even though imprecisely, the totality of related mathematical structures.

The unifying concept, always on the scene whenever entropy makes its presence felt, is that of irreversibility. It is sometimes claimed that irreversibility can appear only through the intervention of a stochastic or random model.¹ This is not so, and we shall give examples to illustrate this presently.² In a classically deterministic system, the relevant concept when viewing evolution of the system in time is the progressive weakening of the property of continuous dependence on initial conditions. As time goes on, the system appears to take on random features, loses sight of its initial state, and approaches a more-or-less universal equilibrium. An entropy which describes this evolution will be a measure of the deviation from equilibrium or a measure of the developing disorder or loss of memory (conventionally, the entropy increases with the disorder). Insofar as this behaviour approximates randomness, the ultimate equilibrium state will be, in some sense, the most probable.

Fortunately, one is frequently satisfied with a comparison of the end states alone. Thus one can put aside the subtleties of whether and how these end states are connected by a motion, either deterministic or stochastic. A given mathematical entropy structure will usually be founded on a set of inequalities, frequently as a consequence of the convexity of some function or functional of the state variables. The questions of time evolution and of the relation between different structures are more difficult.

It is sometimes possible to apply a definite linear ordering to some of the mathematical entropies which apply to a given physical system. One can then identify this set as "an" entropy if he wishes; the advisability of doing so will be discussed subsequently. The first prerequisite is that there be a linear ordering of state variables, i.e., descriptions of the state of the system, each one giving more detail and containing the previous one. Intuitively, a more detailed description is capable of describing a greater degree of order and should have a smaller entropy associated with it. We define the set of entropies, each expressed in terms of its state variables, to be *comparable* provided that such a set of inequalities exists, and provided that the maximum value of a specific entropy for all states which are compatible with a given lower order state is equal to the value of the lower order entropy. A set of comparable entropies does not assign a unique entropy value to a physical system but only to a selected mathematical description of the

¹For example, see M. Kac, *Probability and Related Topics in Physical Sciences*, Interscience, 1959, p. 86.

²An early example was given by H. Poincaré, *Réflexions sur la théorie cinétique des gaz*, Journal de Physique et Le Radium, Ser. 4, Vol. 5, p. 369, 1906.

system. Also, a low level description of a highly ordered state can have the same numerical entropy value as a more detailed description of a more disordered state. On the other hand, such comparisons are intuitively meaningful, and this is not the case for noncomparable entropy values.

We now come to the basic question, how to choose an entropy in a given situation. We claim that the interests of the individual are paramount. As a first example, observation of a portion of the sky might indicate a uniform distribution of stars with respect to number and brightness, thus a high entropy. A new observation, say of mass or color, could uncover a strong inhomogeneity. This would yield a lower value for a different (but comparable) entropy. Here the decision is clearcut; the adoption of a new entropy is forced by the discovery of new information. For another example we turn to aerodynamics. The existence of diffusion between oxygen and nitrogen somewhere in a wind tunnel will usually be of no interest. Therefore the aerodynamicist uses an entropy which does not recognize the separate existence of the two elements but only that of "air". In other circumstances, the possibility of diffusion between elements with a much smaller mass ratio (e.g., 238/235) may be considered quite relevant.

It is not permissible to insist that only the most detailed description of the system is correct, quite apart from the impracticality of this position. Such a stand requires either ultimate knowledge, which is, in principle, unattainable, or else maximal information according to the current state of science. With the latter criterion, the entropy of a system can change suddenly, with the observer unaware, as a result of a discovery just made in a distant laboratory. Even with our attitude, that the choice is up to the individual, it is sometimes expedient to change the definition of entropy, thus its value, without any concomitant physical change in the system under observation. But, such a change in entropy will only occur when some relevant facet of the problem at hand has changed, even if only in the mind of the observer. The change in entropy is therefore substantive to this observer and as significant to him as the result of an operation performed on the system.

We now adopt the specific rule to take, within a set of comparable entropies, the one of maximum value which is compatible with the information which is both known and considered relevant. Such a decision is, of course, subject to review. Although both are significant, one should be careful to distinguish a change in the value of the entropy resulting from altering its definition from a change produced by a physical process. However, one must be even more careful not to interpret as a difference in entropy the different values possessed by non-comparable entropies.

The rule for selecting an entropy is sometimes useful in discovering an entropy which is appropriate to a given specification of a state. If one has an

ordered sequence of states and an *a priori* definition of entropy for the most detailed state, then one can use the maximum property to define a lower order comparable entropy where a suitable definition was not otherwise evident.

It is these observations describing the proper choice of an entropy that eliminate the Gibbs paradox. Whether or not diffusion occurs when a barrier is removed depends not on a difference in physical properties of the two substances but on a decision that we are or are not interested in such a difference (which is what governs the choice of an entropy function). There is no paradox to any observer. When he is aware of a difference in properties, he observes diffusion together with an increase in entropy. When he is unaware of any difference, he observes no diffusion and no increase in the entropy which he is using. If two observers disagree, they must be interested in different phenomena, and there is no conflict.

It is also evident that the quantum resolution of this paradox is not only incomplete (since it cannot refer to purely classical problems of distinguishable particles such as in a cloud of stars), but it is incorrect. It is never permissible to refer to an absolute (and therefore unattainable) decision as to the identity of different particles. It is certainly permissible to say that they are identical for all practical purposes, but the same statement can be made with reference to oxygen and nitrogen under appropriate circumstances. We further note that it *is* possible to treat indistinguishable particles correctly by classical means. In the other direction, it is, in principle, possible to observe (and therefore identify) large molecules with an electron microscope. This is not to say that there is no difference between quantum and classical mechanics, but that the proper handling of distinguishability is separate from either.

The crucial point is that the correct classical interpretation, without paradox and without the need for overt philosophical interpretation, arises automatically in a consistent treatment of classical statistical mechanics.³ The basic error which produced the Gibbs paradox lay in a Procrustean attempt to synthesize two different physical situations into a single thermodynamic structure. The two situations can be described as follows. In the first we have a box containing definite particles labelled 1, 2, \dots , n . In the second experiment we observe a certain region of space together with whatever particles happen to be there at a given instant. The situations are different and, when treated properly, give rise to two different and even non-comparable entropy functions. Ignoring irrelevant dependences, the first entropy is approximately $S_1 = n \log V$ and the second is approximately

³H. Grad, *Statistical mechanics of dynamical systems with integrals other than energy*, J. Phys. Chem. **56**, p. 1039, 1952.

$S_2 = \bar{n} \log (V/\bar{n})$, where \bar{n} is the mean occupation. In the second case, the very description of the state implies that the particles are indistinguishable. This classical entropy function turns out to be essentially the same as in quantum mechanics. It is found as well in any classical situation where we *choose* to consider the particles to be indistinguishable.

To compare the two possibilities more precisely, consider a closed system in which a physical barrier can be inserted or removed to create the two situations. The increase of entropy on removal of the barrier is $n \log 2$ for S_1 and is either $n \log 2$ or zero for S_2 depending on whether the particles in the separate halves are distinguished or not. Reinserting the barrier will return S_1 to its original value (independent of how many particles are caught in each box), but it will have no effect on S_2 whether or not the particles were originally distinguishable (assuming that the diffusion has been carried to completion). It may seem strange, at first sight, that the entropy $S_1 = n_1 \log V + n_2 \log V = (n_1 + n_2) \log V$ has the same value for any distribution of particles between the boxes. But, since the particles are labelled, one has complete information about each particle knowing in which box it is; there is the same amount of order if they are all in one box or equally distributed between the two. The total number of particles in each box is a relevant description only when their individuality is lost. Insertion of a barrier changes the information known about the system; once particle 1 has been seen in a box, it will never leave it. If the barrier is made permeable to particle 1, the entropy increases by an amount $\log 2$, and similarly for each particle, until complete removal of the barrier yields the increase $n \log 2$.

It is usually found that a maximal description, giving complete information, has an entropy which is constant in time. This is no surprise since complete specification of each identified particle leaves no room for any disorder (cf. S_1 above). Only an incomplete description will have an entropy which can vary with time (presumably increasing). There is no conflict between these two descriptions; a single physical system can exhibit some features which are apparently reversible and others which require the epithet irreversible. A mathematical example is discussed in Section 5. A very illuminating experimental example is given by the "spin-echo" effect.⁴ In this experiment, it is found to be possible to produce a highly ordered microscopic initial state and, at a later time, effectively reverse all velocities. To a person who has access to such equipment, a very high level "reversible" entropy will be appropriate; to one who has not, a lower order entropy will properly describe all phenomena.

It is true that the entropy S_1 is completely impractical in the amount of

⁴E. L. Hahn, *Spin echoes*, Phys. Rev. **80**, p. 580, 1950.

information it describes. It is, however, necessary to realize its existence and even its inevitability in a classical treatment in which the particles are labelled with indices at the outset.

In the remainder of this paper, we analyze properties and interrelations of the major sequences of comparable entropies, viz., the sequence of H -functions, and the various thermodynamic and fluctuating entropies. A brief outline of the structure is given in the following section.

2. Morphology

The maximum information is given in statistical mechanics by the n -particle distribution $f(z_1, \dots, z_n)$, where $z_i = (x_i, \xi_i)$ is the phase space of a single particle, x_i being its position and ξ_i its velocity. We adopt $H = \int f \log f dz_1 \cdots dz_n$ as the basic entropy function.

The one-particle distribution $f^{(1)}(z_1) = \int f dz_2 \cdots dz_n$ is frequently taken as an appropriate description of the state of a gas (it is even in a certain sense an asymptotically complete description).⁵ The comparable entropy function, minimizing H subject to given $f^{(1)}$, is simply $H_1 = \int f^{(1)} \log f^{(1)} dz_1$. For an isolated system, f satisfies Liouville's equation, and H is constant in time; however, H_1 is not. A strong indication of a proof can be given that H_1 approaches a minimum as $t \rightarrow \infty$.⁴ Thus, the maximal information contained in f can always be withdrawn at a later time, but with increasing difficulty as time goes on. With less than maximal information, viz., $f^{(1)}$, the irreversible behaviour is unambiguous. To a certain approximation, $f^{(1)}$ will satisfy Boltzmann's equation.

The next reduction in information consists essentially in reducing z_1 to either x_1 or ξ_1 . Each is appropriate under the proper circumstances. One possibility is to introduce the local thermodynamic state (i.e., mass density, fluid velocity, and energy density) at each point x_1 . The comparable entropy, given by the maximum of $-H_1$ at a given point x_1 , is the conventional thermodynamic entropy. The other possibility, taking ξ_1 as a parameter, is to introduce $\int f^{(1)} dx_1 = F(\xi_1)$ as the state variable, in which case the comparable entropy turns out to be

$$H' = \int F \log F d\xi_1 - \log V,$$

where V is the volume. The first procedure is appropriate in a gas with small mean free path in which $-H_1$ closely approximates the thermodynamic entropy. The second procedure is relevant in a gas with infinite mean free path, in which it can be rigorously proved that collisions with the walls

⁵H. Grad, *Principles of the Kinetic Theory of Gases in Handbuch der Physik*, Vol. XII, Springer, 1958.

(rather than between particles) produce spatial uniformity in which H_1 approximates H' (see Section 5). In the thermodynamic case, one has a choice of dissipative equations which yield a simple increase of the entropy with time or non-dissipative equations which one may expect to yield weak convergence (possibly through turbulence) to the state of maximum entropy. No simple equation is known for the function $F(\xi_1)$.

It is possible to insert states intermediate between $f^{(1)}$ and the local thermodynamic state; these are related to the expression (4.11).⁶

For a general fluid (as distinguished from a gas),

$$f^{(2)}(z_1, z_2) = \int f dz_3 \cdots dz_n$$

gives an appropriate description rather than $f^{(1)}(z_1)$. The comparable entropy is not simply

$$H^{(2)} = \int f^{(2)} \log f^{(2)} dz_1 dz_2,$$

which does not make appropriate use of the symmetry of f , but is a more complicated expression⁷ \tilde{H} (Section 4). No satisfactory equation for the time behavior of $f^{(2)}$ for a general fluid has been given. The next step, introduction of a local thermodynamic state, yields the conventional thermodynamic entropy (either by minimizing H or \tilde{H}). Note that the gas and fluid sequences are not comparable with one another. Although it is not relevant to statistical mechanics, $H^{(2)}$ is an appropriate comparable entropy (between H and H_1) if f is not required to be symmetric.

It is possible to introduce ordered structures within the framework of thermodynamics (in a sense proceeding laterally across the preceding structures). Given two isolated systems with fixed energies E_1 and E_2 , one can relax these constraints in favor of the sum, $E_1 + E_2 = E$, and obtain a comparable entropy for the combined system in thermal contact with E as the state variable. Similarly, one can allow two systems containing N_1 and N_2 particles, respectively, to diffuse, obtaining a comparable entropy which is a function of the sum, $N_1 + N_2 = N$, alone. A mathematically similar procedure is to take a single container with N_1 and N_2 molecules of two different chemical constituents and relax these constraints in favor of the sum $N_1 + N_2 = N$. This gives a comparable entropy for the system under the assumption that the two chemical constituents are indistinguishable.

Within statistical thermodynamics, one obtains separate structures for an isolated system, and for systems in thermal contact and in diffusional

⁶See H. Grad, *On the kinetic theory of rarefied gases*, Comm. Pure and Appl. Math. Vol. 2, p. 331, 1949; also J. Lebowitz, H. L. Frisch, and E. Helfand, *Non-equilibrium distribution function in a fluid*, Phys. Fluids 3, 1, 1960.

⁷J. E. Mayer, *Ensembles of maximum entropy*, J. Chem. Phys. 33, p. 484, 1960.

contact. For large systems, the first two are asymptotically the same, and they are asymptotically equivalent to the third with the exception of an additive constant which is indicative of the different assumptions as to distinguishability.

After relaxing a constraint such as energy or number of particles, it is possible to assign to the individual subsystems either instantaneous fluctuating thermodynamic coordinates or fixed values coupled with a more conventional structure (e.g., by assigning the mean value of the energy). If a fixed value is assigned, one must be careful to select a comparable structure; this will not be the case for diffusional subsystems unless an entropy based on indistinguishable particles is used. In the case of fluctuating thermodynamic structures, the deviation between the sum of the fluctuating component entropies and the fixed system entropy is found to be related to the logarithm of the probability density of the fluctuating quantity when suitably defined. For example, the energy fluctuation of a small component is canonical; as a density in phase space,

$$f^{(1)}(z_1) = \frac{e^{-\theta E_1}}{Z_1(\theta)},$$

where E_1 is a function of z_1 , but, as a density in energy space

$$P(E_1) = \frac{e^{-\theta E_1} \Omega_1(E_1)}{Z_1(\theta)}.$$

The fluctuating entropy is correctly given by $\log P$. The fixed thermodynamic entropy of the subsystem is given by $-\int f^{(1)} \log f^{(1)} dz_1$, i.e., by the mean value of $\log f^{(1)}$. On the other hand, the mean value of $\log P$ is approximately zero. Similar considerations apply to density and volume fluctuations.

It should be remarked that $\int f \log f$, where f is the n -particle distribution, can only yield the thermodynamic entropy for distinguishable particles. But, from $\int f^{(1)} \log f^{(1)}$, one can obtain either this entropy or the diffusional (i.e., quantum) entropy, depending on the normalization of $f^{(1)}$.

As a final remark, we note that there are interesting cases in which the sum of a number of component or sub-entropies can be either equal to or greater than or less than the system entropy. The sum is larger if the combination of component entropies contains less information than the system entropy; for example $-nH_1 \geq -H$. The sum can be smaller if the combination has more detail. For example, the sum of the fluctuating entropies describes an instantaneous separation of E into $E_1 + E_2$ which is a more detailed description. An example of equality is given by splitting H_1 into two terms (see (5.17)), one of which expresses the deviation of $f^{(1)}$ from a locally Maxwellian distribution, the other being the thermodynamic entropy.

3. A Combinatorial Problem

Consider the problem of distributing n identical objects at random in r boxes which are assigned probabilities p_1, p_2, \dots, p_r , $\sum p_i = 1$. The probability of finding k_1 objects in box 1, k_2 in box 2, etc., where $\sum k_i = n$, is

$$(3.1) \quad P(k_1, \dots, k_r) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r}.$$

As a special case, the probability of observing the value k_1 regardless of the others is

$$(3.2) \quad P(k_1) = \frac{n!}{k_1! (n-k_1)!} p_1^{k_1} (1-p_1)^{n-k_1}.$$

For large n and fixed r , $P(k)$ behaves like the n -th power of a function of

$$(3.3) \quad y_i = \frac{k_i}{n}.$$

Using Sterling's formula in (3.1), with y fixed, we verify that

$$(3.4) \quad \lim_{n \rightarrow \infty} P^{1/n} = \left(\frac{p_1}{y_1}\right)^{y_1} \dots \left(\frac{p_r}{y_r}\right)^{y_r}.$$

The more familiar Gaussian approximation (which has a more restricted range of validity near the maximum of P) is obtained by writing

$$(3.5) \quad y_i = p_i (1 + n^{-1/2} \eta_i),$$

which yields

$$(3.6) \quad P \sim \text{const.} \exp\left\{-\frac{1}{2} \sum p_i \eta_i^2\right\}, \quad \eta = O(1).$$

The singular behavior of P , viz., that it is exponential in n , can be removed as in (3.4) by taking the n -th root, or more conventionally by taking a logarithm,

$$(3.7) \quad \lim \frac{1}{n} \log P = \sum y_i \log \frac{p_i}{y_i}.$$

Comparing (3.6) and (3.7), we see that the Gaussian distribution describes the (finite) likelihood of observing a state which deviates only slightly from the most probable, while the logarithmic probability (which suggests the H -function $\int y \log y$) can be used to describe, on an entirely different scale, finite (extremely improbable) deviations.

In order to be able to interpret the sum in (3.7) as an approximation to an integral, something must be said about the step size. If we interpret p_i as a measure of the size of the occupied domain, then the appropriate form is

$$(3.8) \quad \sum \frac{y_i}{\phi_i} \log \frac{\phi_i}{y_i} \phi_i \Delta i.$$

This is exactly in the form $\int f \log f$, where f is the ratio of the density function, y , to the most probable or mean density, ϕ .

The computations given above have more general validity than might appear at first glance. First, to relate these formulas to statistical mechanics, consider a constant probability density distributed on a unit cube $0 \leq x_j \leq 1$, $j = 1, 2, \dots, n$. Each interval $0 \leq x_j \leq 1$ is divided into subintervals of length ϕ_i , $i = 1 \dots r$, the unit cube into r^n parts. If at the point (x_1, \dots, x_n) , we define the random variable (i.e., phase function) $k_i(x_1, \dots, x_n)$ to be the number of coordinates x_j which lie in the interval ϕ_i , then $P(k_1, \dots, k_r)$ in (3.1) is the probability distribution function of the occupation numbers k_i . Moreover, it can be shown that (3.4) is a suitable approximation for large n even if the *a priori* probability distribution $f(x_1, \dots, x_n)$ is not a constant; it is sufficient if f is bounded by a constant, c_n , which does not grow exponentially with n . In this application to statistical mechanics, $P(k)$ can be interpreted as the probability of observing the density fluctuations in a given dynamical system which varies in time or in samples taken from separately prepared systems.

Another application of these formulas is provided by taking the probability distribution

$$(3.9) \quad f = \pi^{-n/2} \exp \left\{ - \sum_1^n x_i^2 \right\}$$

on the space $-\infty < x_i < +\infty$. The variables are independent, and we have merely to define ϕ_i as $\pi^{-1/2} \int \exp \{ -\frac{1}{2}x^2 \} dx$ over the appropriate interval to recover (3.1) and all succeeding formulas. In this case (3.8) approximates $\int f \log f d\mu$, where f is the ratio of the given density to the Gaussian and $d\mu$ is the Gaussian measure, $\pi^{-1/2} \exp \{ -\frac{1}{2}x^2 \} dx$.

Thus we see that the problem of the fluctuations in density of particles freely moving in a container can be analyzed exactly as the problem of fluctuations in the distribution of energy over the particles in a box. In one case, we measure the deviation from uniformity, in the other, the deviation from a Gaussian distribution. It is clear that the general framework does not depend on the special Gaussian distribution (3.9). Any independent probability $f = f_1(x_1)f_1(x_2) \dots f_1(x_n)$ can be treated. It is also possible to show that many of the results quoted hold for a microcanonical distribution in the limit of large n .

In some problems an entirely different H -function, $\sum \phi_i \log \phi_i$, will be significant. This represents the deviation of the one-particle marginal distribution $f^{(1)}(x_1)$ from uniformity (equilibrium) and may be time-depend-

ent. It is entirely distinct from $\sum (y_i/p_i) \log (p_i/y_i)$ which measures the likelihood of deviations of the fluctuating density y_i from its instantaneous mean value p_i .

4. Inequalities for the H -function

From the convexity of the function $\phi(f) = f \log f$, we conclude that

$$(4.1) \quad f \log f \geq f - 1, \quad f \geq 0.$$

There is strict inequality unless $f = 1$. Writing f/g for f , we have the alternative form

$$(4.2) \quad f \log f - f \log g \geq f - g, \quad f \geq 0, \quad g > 0.$$

Here we have strict inequality unless $f = g$.

The remainder of this section will be devoted to a study of the various integral inequalities which are obtained by taking f and g to be functions on some measure space and integrating. The spaces of interest will be intervals, finite or infinite, and products of such intervals.

For the first example, take f to be an arbitrary non-negative normalized function on a space of total measure unity,

$$(4.3) \quad \int d\mu = 1, \\ \int f d\mu = 1.$$

Integrating (4.1), we obtain

$$(4.4) \quad \int f \log f d\mu \geq 0.$$

Equality holds only if $f = 1$.⁸ This result can be interpreted variationally as the statement that the functional $H[f] = \int f \log f$ is minimized by the function $f = 1$. The simplest application is to functions $f(x)$ defined on the unit interval $0 \leq x \leq 1$. Clearly, H is a measure of the deviation of f from its mean value.

Next, consider the easy generalization to a space of finite total measure m and to functions of norm a ,

$$(4.5) \quad \int d\mu = m, \\ \int f d\mu = a.$$

By introducing the measure $d\nu = d\mu/m$ and the function $f' = mf/a$ to

⁸For measurable functions, a statement such as " $f = 1$ " or " $f = g$ " is to be interpreted as holding almost everywhere. All the results also hold for continuous functions, in which case such statements hold everywhere.

which (4.4) can be applied, we obtain

$$(4.6) \quad \int f \log f d\mu \geq a \log \frac{a}{m}.$$

There is equality if f is the constant a/m .

A similar inequality is obtained by integrating (4.2), taking f and g to be two functions of equal norm (the total measure of the space need not be specified and can be infinite),

$$(4.7) \quad \int f = \int g,$$

viz.,

$$(4.8) \quad \int f \log f \geq \int f \log g.$$

Equality follows only when $f = g$. This inequality hides a wealth of interesting possibilities depending on the choice of an appropriate comparison function g . We shall give several examples.

First, on the infinite line, $-\infty < x < +\infty$, we take

$$(4.9) \quad g(x) = ae^{-b(x-c)^2}$$

and require

$$(4.10) \quad \begin{aligned} \int x f dx &= \int x g dx, \\ \int x^2 f dx &= \int x^2 g dx, \end{aligned}$$

as well as $\int f dx = \int g dx$. Since $\log g$ is quadratic in x , we have

$$\int f \log g dx = \int g \log g dx.$$

The inequality (4.8) reads $\int f \log f \geq \int g \log g$. Thus, among functions f with given zero, first, and second moments, $H[f]$ is minimized by the Maxwellian function g which has the same moments (it is separately verified that these moments determine g uniquely). Exactly the same theorem holds if x is n -dimensional; the first moments and c are vectors, as is x .

An approximate generalization of this result is the following. We take

$$(4.11) \quad g = e^{-x^2} \sum_0^N a_n \mathcal{H}_n(x),$$

where \mathcal{H}_n are Hermite polynomials and a_n are given constants. We demand of f that

$$(4.12) \quad \int f \mathcal{H}_n dx = \int g \mathcal{H}_n dx, \quad n = 0, 1 \cdots N;$$

(if $N < 2$, the energy constraint, $n = 2$, is also included).

The conclusion, from (4.8), is that $H[f]$ is minimized by the function g .

$$(4.20) \quad H \geq H_1 + \cdots + H_n - (n-1)a \log a,$$

and equality is attained only when $f = a^{1-n} f_1 f_2 \cdots f_n$.

A slightly different version of this result is obtained by dividing the coordinates (x_1, \cdots, x_r) into two (or more) sets $x' = (x_1, \cdots, x_m)$, $x'' = (x_{m+1}, \cdots, x_n)$. For the marginal distributions

$$(4.21) \quad \begin{aligned} \int f'(x') &= \int f(x) dx'', \\ \int f''(x'') &= \int f(x) dx', \end{aligned}$$

we introduce

$$(4.22) \quad \begin{aligned} H' &= \int f' \log f' dx', \\ H'' &= \int f'' \log f'' dx'', \end{aligned}$$

and conclude that

$$(4.23) \quad H \geq H' + H'',$$

with equality only if $f = f' f''$.

If f is symmetric in its n arguments, the inequality (4.19) takes the simple form

$$(4.24) \quad H \geq nH_1.$$

Again, the inequality is strict unless f is a product. The implications of symmetry with respect to the inequality (4.23) are more subtle. Indeed, in contrast to all other inequalities (in particular (4.24)), this inequality is not the best possible since the minimizing function, $f = f' f''$, is not necessarily symmetric in the n variables. To pursue this, we restate the content of the inequality (4.24) as follows. Consider the class of symmetric functions $f(x_1, \cdots, x_n)$ which have a given function $f_1(x_1)$ as their common marginal distribution. The functional $H[f]$ is minimized in this class by the symmetric function $f = f_1(x_1) \cdots f_1(x_n)$, and the minimum value is nH_1 . Thus, if $H[f]$ is considered as an *a priori* entropy function for functions of n variables, we are led to the conclusion that $H_1[f_1]$ is a compatible entropy function when incomplete information, viz., f_1 , is given.

Now let us examine the same question for the two-coordinate marginal distribution,

$$(4.25) \quad f^{(2)}(x_1, x_2) = \int f dx_3 \cdots dx_n.$$

Considering symmetric f , for even n we can obtain

$$(4.26) \quad H \geq \frac{n}{2} \int f^{(2)} \log f^{(2)} dx_1 dx_2 \equiv \frac{n}{2} H^{(2)},$$

while for odd n ,

$$(4.27) \quad H \geq \frac{n-1}{2} H^{(2)} + H_1.$$

Equality is obtained in (4.26) with

$$(4.28) \quad f = f^{(2)}(x_1, x_2) f^{(2)}(x_3, x_4) \cdots f^{(2)}(x_{n-1}, x_n)$$

and in (4.27) with

$$(4.29) \quad f = f^{(2)}(x_1, x_2) \cdots f^{(2)}(x_{n-2}, x_{n-1}) f_1(x_n),$$

as well as with other such functions obtained by permutation of coordinates. If these functions do not happen to be symmetric in all n variables, then the right sides of (4.26) and (4.27) do not give the correct minimum values taken by H evaluated for *symmetric* functions f . Furthermore, it is easy to verify that these minimizing expressions are symmetric only when $f^{(2)}(x_1, x_2) = f_1(x_1) f_1(x_2)$. For, assuming symmetry and interchanging x_2 and x_3 , we obtain

$$f^{(2)}(x_1, x_2) f^{(2)}(x_3, x_4) = f^{(2)}(x_1, x_3) f^{(2)}(x_2, x_4);$$

integration with respect to x_3 and x_4 then yields the result.

The conclusion is that the two-coordinate H -function, $\int f^{(2)} \log f^{(2)}$, does not correctly incorporate the symmetry of f (except in the trivial case $f^{(2)} = f_1(x_1) f_1(x_2)$, in which case it is really H_1 that covers the situation). We pose the problem, to minimize H subject to $f^{(2)}$ being given. The solution, we claim, is given by a function f which can be written

$$(4.30) \quad \begin{aligned} \log f &= \sum_{i < j} q_{ij}, \\ q_{ij} &= q(x_i, x_j). \end{aligned}$$

This result is suggested by the formal procedure of performing a variation of the functional

$$(4.31) \quad H - \int q(x_1, x_2) f^{(2)}(x_1, x_2) dx_1 dx_2,$$

where q is a Lagrange multiplier. The difficulty is to relate the function q to the given function $f^{(2)}$, i.e., to invert the relation

$$(4.32) \quad f^{(2)}(x_1, x_2) = \int \exp \{ \sum q_{ij} \} dx_3 \cdots dx_n.$$

We conjecture that this can always be done. In the absence of a proof, we consider only those functions $f^{(2)}$ which can be so generated from functions q . For such an $f^{(2)}$ we can prove that the generator q is unique, and the related product f does minimize H . Given $f^{(2)}$ and any q related to it by (4.32), we construct

$$(4.33) \quad g = \exp \{ \sum q_{ij} \}.$$

That $\int g = 1$ follows from (4.32). Applying the inequality (4.8), we find

$$(4.34) \quad H \geq \frac{1}{2}n(n-1) \int f^{(2)}(x_1, x_2)q(x_1, x_2)dx_1dx_2 \equiv \frac{1}{2}n(n-1)\tilde{H},$$

where the inequality is strict unless $f = g$. The last remark implies that g is unique; a very simple argument then shows that q is also unique.

We can combine the inequalities (4.26) and (4.34) as follows:

$$(4.35) \quad H \geq \frac{1}{2}n(n-1)\tilde{H} \geq \frac{1}{2}nH^{(2)};$$

there is a similar result if n is odd. The second inequality is a consequence of the fact that the middle member is the minimum for a variational problem with fewer admissible functions than one for which the last member is the minimum. The first inequality is strict unless f is of the form (4.33); the second inequality is strict unless $f^{(2)} = f_1(x_1)f_1(x_2)$. In short, $\int f^{(2)} \log f^{(2)}$ is an appropriate entropy function for unrestricted functions f , whereas $\int f^{(2)}q$ is appropriate for symmetric functions f .

We remark that the variation of $H[f]$ with energy fixed, (4.14), gives the same result as the variation of $\tilde{H}[f^{(2)}]$ with energy fixed, provided that ε takes the form $\sum \varepsilon_{ij}$, $\varepsilon_{ij} = \phi(x_i, x_j)$. We have $q = \alpha\phi + \beta$, where α and β are adjusted to yield the normalization and correct mean value of ε .

We conclude with some general remarks regarding bounds for the marginal H -functions, H_r . If the domain of integration (x_1, \dots, x_n) is finite, then the existence of H alone in (4.19) implies the existence of each of the integrals H_r . For, with a finite total measure, (4.6) gives a lower bound for each H_r . Transposing all but one of the H_r in (4.19) yields an upper bound for the remaining one. With an infinite domain, it is possible for H_r to be either positively or negatively infinite even though H is finite. We shall illustrate these conclusions with several two-dimensional examples.

First consider $f(x_1, x_2)$, where $0 \leq x_r \leq 1$. From (4.4) and (4.19) we observe that

$$(4.36) \quad 0 \leq H_r \leq H, \quad r = 1, 2.$$

In the application to statistical mechanics, the equations of motion induce an area-preserving flow on the phase space (x_1, x_2) , with the values of f carried by the flow. We can interpret this more abstractly as a rearrangement of the values taken by f on (x_1, x_2) , without looking in detail at the possible motions. Clearly H is constant for any rearrangement, but H_1 and H_2 are not. H_1 can attain its maximum value, viz., H , only when $H_2 = 0$. This can occur only if $f_2 = 1$, consequently $f = f_1(x_1)$. Such a situation can be described as a rearrangement of the values of f to yield the maximum order

with regard to x_1 (all points with a given value of f are moved to the same x_1) and maximum disorder, or rather mixing, with regard to x_2 . The combined order, $H_1 + H_2$, is a maximum for any arrangement which satisfies $f = f_1 f_2$, of which $f = f_1, f_2 = 1$ is a particular example.

If f is symmetric, instead of (4.36) we have

$$(4.37) \quad 0 \leq H_r \leq \frac{1}{2}H, \quad r = 1, 2.$$

There is no possibility of obtaining maximum order for either H_1 or H_2 , but only a maximum combined order when f is a product. It is intuitively clear that any very complicated mixing will produce values of H_1 and H_2 which are close to zero.

Now we turn to the case of two variables $f(x, \xi)$, where $0 \leq x \leq 1$ and $-\infty < \xi < +\infty$. Thus x could represent position and ξ velocity. In this example we introduce the special notation

$$(4.38) \quad \begin{aligned} \rho(x) &= \int f(x, \xi) d\xi, \\ F(\xi) &= \int f(x, \xi) dx, \end{aligned}$$

also

$$(4.39) \quad \begin{aligned} H_\rho &= \int \rho \log \rho \, dx, \\ H_F &= \int F \log F \, d\xi. \end{aligned}$$

As always, $H \geq H_\rho + H_F$. We have a lower bound on H_ρ , consequently an upper bound on H_F ,

$$(4.40) \quad \begin{aligned} H_\rho &\geq 0, \\ H_F &\leq H. \end{aligned}$$

It is easy to construct examples in which H_F does not exist or in which neither H_F nor H_ρ exists, but existence of H_F implies the same for H_ρ . There are many simple conditions which imply the existence of a lower bound on H_F , e.g., finiteness of the energy,

$$(4.41) \quad E = \int \xi^2 f(x, \xi) dx d\xi = \int \xi^2 F(\xi) d\xi.$$

Introducing the measure $d\mu = \pi^{-1/2} e^{-\xi^2}$ and the auxiliary function $F' = \pi^{1/2} e^{\xi^2} F$, we note that $\int d\mu = \int F' d\mu = 1$. The inequality (4.4) applied to F' yields

$$\begin{aligned} 0 \leq \int F' \log F' d\mu &= \int \left(\frac{1}{2} \log \pi + \xi^2 + \log F \right) F d\xi \\ &= \frac{1}{2} \log \pi + E + H_F, \end{aligned}$$

from which we obtain a lower bound for H_F . Thus even in the infinite domain, there exist both upper and lower bounds on the marginal H -functions in terms of constants of the motion, H and E .

5. Irreversible Behavior of a Gas

We examine the properties of a gas of independently moving particles in a rectangular box.⁹ The particles do not see one another and collide only with the walls. We take position coordinates $x = (x^1, x^2, x^3)$ and velocity $\xi = (\xi^1, \xi^2, \xi^3)$. The motion of a particle is given by

$$(5.1) \quad \begin{aligned} \dot{\xi} &= \xi_0, \\ \dot{x} &= x_0 + \xi_0 t \end{aligned}$$

in the extended domain x which is covered by reflections of the basic box, $0 \leq x^i \leq a^i$. It is easier to treat the doubled domain, $0 \leq x^i \leq 2a^i$, as basic, since the boundary condition is then periodicity. More precisely, if there is given an initial distribution $f_0(x, \xi)$ which is periodic in the doubled domain, then the solution to the originally posed problem is obtained from

$$(5.2) \quad f(x, \xi, t) = f_0(x - \xi t, \xi)$$

by "folding" once with respect to each space coordinate. Since this folding will be irrelevant to our purposes, we shall pursue the modified problem of the motion (5.1) in a basic domain $0 \leq x^i \leq a^i$ with $a^1 a^2 a^3 = 1$ (unit volume) but with periodic rather than reflection boundary conditions; in other words we assume that $f_0(x, \xi)$ has periods a^i , and the time evolution of f is given by (5.2).

The marginal distributions $\rho(x)$ and $F(\xi)$ and the marginal entropies H_ρ and H_F are defined in (4.38) and (4.39). It is evident that $F(\xi)$ is independent of the time. Thus H_F as well as H is constant in time. H_ρ is the only entropy which can vary, and it is restricted by the inequalities

$$(5.3) \quad 0 \leq H_\rho \leq H - H_F.$$

In order to reach the desired results with a minimum of mathematical manipulation we assume that $f_0(x, \xi)$ is uniformly bounded,

$$(5.4) \quad \begin{aligned} f_0(x, \xi) &< \bar{f}(\xi), \\ \int \bar{f} d\xi &= M, \\ \int \bar{f} \log \bar{f} d\xi &= \bar{H}; \end{aligned}$$

⁹Related problems are discussed in:

- a) H. Poincaré, *loc. cit.*
- b) H. Grad, *Handbuch der Physik, loc. cit.*
- c) H. L. Frisch, *An Approach to Equilibrium*, Phys. Rev. **109**, p. 22, 1958.

also $f_0(x, \xi)$ is uniformly continuous,¹⁰

$$(5.5) \quad |f_0(x, \xi') - f_0(x, \xi)| < \varepsilon$$

if

$$(5.6) \quad |\xi' - \xi| < \delta(\varepsilon).$$

First we prove that $\rho(x, t) \rightarrow 1$ and $H_\rho(t) \rightarrow 0$ as $t \rightarrow \infty$. We write

$$(5.7) \quad \begin{aligned} \rho(x, t) &= \int f_0(x - \xi t, \xi) d\xi, \\ \rho(x', t) &= \int f_0(x' - \xi t, \xi) d\xi \\ &= \int f_0(x - \xi' t, \xi' + \eta) d\xi', \end{aligned}$$

where

$$(5.8) \quad \eta = \frac{x' - x}{t}.$$

Subtracting (and changing ξ' to ξ),

$$\rho(x, t) - \rho(x', t) = \int [f_0(x - \xi t, \xi) - f_0(x - \xi t, \xi + \eta)] d\xi,$$

whence

$$\begin{aligned} |\rho(x, t) - \rho(x', t)| &\leq \int_{|\xi| < A} |f_0(x - \xi t, \xi) - f_0(x - \xi t, \xi + \eta)| d\xi \\ &\quad + 2 \int_{|\xi| > A} \bar{f}(\xi) d\xi. \end{aligned}$$

Given $\varepsilon > 0$, we choose A to make the term in \bar{f} smaller than $\varepsilon/2$. Then, by taking t large, i.e., $|\eta| < \delta$, the first integral, which is bounded by $\pi A^3 \delta/6$ can also be made smaller than $\varepsilon/2$. From

$$(5.9) \quad |\rho(x, t) - \rho(x', t)| < \varepsilon, \quad t > t_0,$$

and the normalization, $\int \rho dx = 1$, we conclude that

$$(5.10) \quad |\rho(x, t) - 1| < \varepsilon, \quad t > t_0.$$

It is an elementary consequence of this inequality that

$$|\rho \log \rho| < \varepsilon + \frac{1}{2}\varepsilon^2,$$

from which follows

¹⁰Theorems equivalent to those proved here for continuous functions can be proved for measurable functions as well, but the proofs are rather intricate and will not be given.

$$(5.11) \quad H_\rho < \varepsilon + \frac{1}{2}\varepsilon^2, \quad t > t_0.$$

The weak convergence of $f(x, \xi, t)$ to $F(\xi)$ is proved in the same way. Let I and I' refer to the intervals

$$(5.12) \quad \begin{aligned} I: \xi_0 &\leq \xi \leq \xi_1, \\ I': \xi_0 + \eta &\leq \xi \leq \xi_1 + \eta. \end{aligned}$$

We write

$$(5.13) \quad \begin{aligned} \hat{f} &= \int_I f(x, \xi, t) d\xi = \int_I f_0(x - \xi t, \xi) d\xi, \\ \hat{f}' &= \int_{I'} f(x', \xi, t) d\xi = \int_{I'} f_0(x - \xi t, \xi + \eta) d\xi. \end{aligned}$$

The difference between \hat{f} and \hat{f}' can be written as an integral over the intersection of I and I' plus an integral over the difference. The first integrand is estimated by the continuity of f , the second by the uniform bound \bar{f} , and both become arbitrarily small when $|\eta|$ is small,

$$|\hat{f} - \hat{f}'| < \varepsilon, \quad t > t_0.$$

From the fact that

$$\int \hat{f} dx = \int \hat{f}' dx' = \int_I F(\xi) d\xi \equiv \hat{F}$$

we conclude that

$$(5.14) \quad |\hat{f} - \hat{F}| < \varepsilon, \quad t > t_0.$$

We remark that this is a somewhat stronger statement than conventional weak convergence; the integral of f over a ξ -interval rather than a maximum dimensional (x, ξ) -interval is convergent.

The significance of the strict convergence of $\rho(x, t)$ to its limit, without interference from exceptional events, Poincaré recurrences, etc., may require elaboration. We have taken the position that $f(x, \xi)$ is a probability distribution, governed by repeated sampling. Let us now interpret it as an actual physical density. For a system containing a finite number of particles, n , f will be a sum of n δ -functions. The equation (5.2) for the time variation of f is the same whether f is an actual density or a probability density; properly interpreted, it is the same whether f is singular or smooth. To make use of the results we have obtained, we now take a sequence of singular initial distributions, $f_n(x, \xi)$ which converges weakly to the smooth function $f_0(x, \xi)$. It is a simple matter to verify that, in any fixed interval, $0 < t < t_0$, $f_n(x, \xi, t)$ will converge weakly to $f(x, \xi, t)$. For a fixed large value of t , $f_n(x, \xi, t)$ will therefore converge weakly to a function which is close to $F(\xi)$. Similarly, $\rho_n(x, t)$ (which is also singular) will converge weakly to $\rho(x, t) \sim 1$. However, this convergence is not uniform in t . For fixed n it

is known that complex behavior such as Poincaré recurrences will occur for large enough t . This phenomenon, resulting from the inversion of the limits $n \rightarrow \infty, t \rightarrow \infty$, is well understood, but this is one of the very few examples in which the mathematical analysis can be completed.

A more difficult problem than the one considered up to now, but one that still can be analyzed rigorously, is that of the expansion or compression of the gas by movement of the walls. To be definite, we assume that the state at time $t = 0$ is one of equilibrium, viz., $\rho = 1$ and $f_0 = F_0$, and we assume further that the motion of the wall ceases at a finite time t_1 . Of course, the gas parameters will continue to vary after t_1 , but another equilibrium will ultimately be approached in which $\rho_1 = 1/V$ (V is the final volume) and f converges weakly to F_1 . H is still a constant of the motion, but neither F nor H_F are constant. Initially we have

$$(5.15) \quad H = H_\rho^0 + H_F^0, \quad H_\rho^0 = 0,$$

and finally

$$(5.16) \quad H \geq H_\rho^1 + H_F^1, \quad H_\rho^1 = \log \frac{1}{V}.$$

Thus the combined entropy, $-(H_\rho + H_F)$, has increased as a result of the change in volume. The inequality holds unless $f = \rho F$ is identically true at the instant $t = t_1$; it has not been proved, but it seems very likely that this factorization is impossible for any actual motion of the walls. What is true, however, is that equality in (5.16) is *approached* if the change of volume occurs more and more slowly (adiabatically). The proof is somewhat intricate but not difficult and will not be given here. It is based on the fact that there is an adiabatic invariant (viz., $\xi^1 a^1$, where a^1 is the length of the side which is varied) for an individual particle orbit. This invariant is constant to an arbitrarily high order in the slowness parameter ε provided that the motion is smooth.¹¹ However, it is not uniform in velocity; in particular, particles with small velocities of order ε are not adiabatic. Thus the change in $H_\rho + H_F$ turns out to be of order ε .

A natural question to ask is, what is the relation of the behavior of this highly idealized model of a gas to one with colliding molecules? The problem is extremely difficult and one can only conjecture, but these conjectures appear to be very soundly based. For simplicity, suppose the molecules are elastic spheres. The path of the system in $6n$ -dimensional phase space is exactly a rectilinear motion at constant speed interrupted by reflections at rigid walls. The location of the walls is extraordinarily complex however.

¹¹C. S. Gardner, *Adiabatic invariants of periodic classical systems*, Phys. Rev. **115**, p. 791, 1959.

In addition to the actual physical walls, there are walls given by the typical equation $|x_i - x_j| = d$, where d is the molecular diameter. The "speed" in this $6n$ -dimensional model is essentially the total energy of the system. It is likely that, except for very special initial conditions, the energy is the only constant of the motion and the path is ergodic on an energy surface. However, we do not have to depend on this property, and it is not the property which we have made use of in our simple model. Rather, it is the concept that two initial states with slightly different velocities will eventually end up very far apart. This can be true within an ergodic subspace if the motion is mixing, but this condition is not required. In our example, the flow within an ergodic subspace (constant ξ) is completely non-mixing. It is intuitive that intermolecular collisions can weaken the continuous dependence on initial conditions very effectively, much more so than collisions with the walls. It is also clear from our example that the Poincaré recurrence theorem is no impediment to an approach to equilibrium provided that we avoid singular distribution functions. Thus, in a realistic gas model with collisions we should expect the marginal entropy H_1 for a single particle (Boltzmann's H -function) to approach its minimum (i.e., equilibrium) value strictly. The averaging which is required to convert weak convergence of f into strong convergence is implicit in the integration which defines the marginal distribution f_1 .¹²

The analogy between our model and the colliding gas can be made stronger. For a gas which is governed by the Boltzmann equation, it is appropriate to consider

$$\begin{aligned}
 (5.17) \quad H &= \int f \log f \, d\xi \, dx = H' + H'', \\
 H' &= \int f_0 \log f_0 \, d\xi \, dx, \\
 H'' &= \int f \log \frac{f}{f_0} \, d\xi \, dx,
 \end{aligned}$$

where f_0 is the local Maxwellian corresponding to f . H'' is a measure of the deviation of f from the local Maxwellian and H' measures the deviation from uniformity in space (it is exactly the negative of the thermodynamic entropy). To make the comparison, in the non-colliding gas we write

$$\begin{aligned}
 (5.18) \quad H_\rho + H_F &= H' + H'', \\
 H' &= H_F - \log V, \\
 H'' &= H_\rho + \log V,
 \end{aligned}$$

¹²A more complete discussion of these points is given in H. Grad, *Handbuch der Physik*, loc. cit.

where V is the variable volume. In the one case, intermolecular collisions and, in the other, collisions with the wall make H'' approach zero. In an adiabatic process, H'' remains at all times close to zero, whereas H' decreases slowly. Note that in this analogy position and velocity are essentially reversed!

6. Macroscopic Thermodynamics

For an isolated system of fixed composition, a concise formulation of the thermodynamic structure is that there exists an entropy $S(E, V)$ which is a convex function of energy and volume. The pressure and temperature can be defined as the partial derivatives of this function by the schematic

$$(6.1) \quad TdS = dE + pdV.$$

This can be written alternatively in terms of densities per unit mass or volume. Since the total mass, M , is fixed, we merely divide through by M in (6.1) to obtain

$$(6.2) \quad Td\eta = d\varepsilon + pd\tau,$$

where

$$(6.3) \quad \begin{aligned} S &= \eta M, \\ E &= \varepsilon M, \\ V &= \tau M. \end{aligned}$$

In terms of volume densities,

$$(6.4) \quad \begin{aligned} S &= sV, \\ E &= eV, \end{aligned}$$

we obtain

$$Tds = de + (p + e - Ts) \frac{dV}{V}.$$

We write dV/V as $d\tau/\tau$ or $-d\rho/\rho$,

$$(6.5) \quad \rho = \frac{1}{\tau},$$

and introduce the quantity

$$(6.6) \quad \mu = \tau(Ts - e - p),$$

to obtain

$$(6.7) \quad Tds = de - \mu \frac{d\tau}{\tau^2} = de + \mu d\rho.$$

We remark that it is V alone that is varied in the differentials $d\tau$ and $d\rho$.

With composition variable, we postulate the existence of a convex function $S(E, V, M_i)$, where M_i are the component masses, and introduce the partial derivatives of S through

$$(6.8) \quad T dS = dE + \sum \mu_i dM_i + p dV.$$

In addition to the volume densities (6.4), we introduce

$$(6.9) \quad M_i = \rho_i V$$

and compute

$$T ds - de - \sum \mu_i d\rho_i = (p + \sum \mu_i \rho_i + e - Ts) \frac{dV}{V}.$$

It is now postulated that S is homogeneous of first degree in the variables (E, V, M_i) , or the equivalent, that s is a function of e and ρ_i alone:

$$(6.10) \quad T ds = de + \sum \mu_i d\rho_i,$$

or another equivalent,

$$(6.11) \quad \begin{aligned} p + \sum \mu_i \rho_i + e - Ts &= 0, \\ pV + \sum \mu_i M_i + E - TS &= 0. \end{aligned}$$

We note that the previous formalism with mass constant is encompassed within the present formalism as a special case only if the homogeneity assumption is made there. In practice it is not the entropy function which is directly measured but the partial derivatives p and T and to some extent μ_i . Thus S is determined only within an additive constant. In the general case of variable mass, the constant is a number; in the case of fixed mass, the constant could be an arbitrary function of the mass and thus give a non-homogeneous entropy function. But, since the added constant inherently contains only parameters which are of no immediate interest (they are not varied), it is permissible (although not necessary) to take the homogeneous entropy universally in this macroscopic theory.

It is an elementary consequence of the convexity of S that there exists a unique maximum for the sum

$$(6.12) \quad S'(E', V') + S''(E'', V'')$$

when E' and E'' are varied subject to the sum $E' + E'' = E$ being constant. This maximum is obtained when $T' = T''$. On relaxing the two constraints E' and E'' in favor of the single constraint E (thereby defining thermal contact), we obtain a system described by the coordinates E, V', V'' . The

entropy of this system, $S(E, V', V'')$, is defined to be the maximum value of $S' + S''$. It is convex and satisfies

$$(6.13) \quad TdS = dE + p'dV' + p''dV''.$$

The entropy has been defined to have the property that bringing two bodies into thermal contact increases it.

If we adopt the conservative attitude that thermodynamic functions are properly defined only in terms of fixed constraints or parameter values, then the individual systems in thermal contact do not possess separate entropies, since the energies, E' and E'' , are fluctuating quantities. What is conventionally done is to consider T (or $\theta = 1/T$ which is more convenient) to be a common property of either of the systems in thermal contact. The appropriate thermodynamic function, when θ is the independent variable rather than E , is given by the Legendre transformation

$$(6.14) \quad \begin{aligned} F(\theta, V) &= \theta E - S, \\ dF &= E d\theta - \theta p dV, \end{aligned}$$

of which the inverse is

$$(6.15) \quad \begin{aligned} S(E, V) &= \theta E - F, \\ dS &= \theta(dE + p dV). \end{aligned}$$

The potential F is held to be defined for each of the systems in thermal contact and is seen to be additive,

$$F(\theta, V', V'') = F'(\theta, V') + F''(\theta, V'').$$

The convexity of S implies convexity of F , and it is an elementary consequence of this property that

$$(6.16) \quad \begin{aligned} F(\theta) &= \min_E [\theta E - S(E)], \\ S(E) &= \min_\theta [\theta E - F(\theta)]. \end{aligned}$$

In the first expression the minimizing E is the value associated with θ , and in the second expression the minimizing θ is that associated with the given E .

These relations supply a simple and elegant proof of the minimum property of $S' + S''$ that was referred to above. We minimize

$$(6.17) \quad \theta(E' + E'') - S'(E', V') - S''(E'', V'')$$

with respect to both E' and E'' , where θ is a Lagrange multiplier. The minimum is attained at $\theta = 1/T' = 1/T''$ and the minimum value is $F(\theta, V', V'')$. At the same time, $\theta E - S(E)$ is minimized for each component separately.

Now, from the potential $F'(\theta, V')$ of one component, we can formally compute a value for E' , viz., $E'_0 = \partial F'/\partial\theta$ and a value for S' , $S'_0 = \theta E'_0 - F'$. Clearly this value for E' is not “the” energy of the component since it has no fixed energy. Also, in computing $\theta E' - F'$, we would seem to have a choice of inserting the actual fluctuating energy or the fixed value E'_0 . The conventional thermodynamic procedure is to extend the range of validity of the purely thermodynamic formalism by assigning the constant values E'_0 and S'_0 to the fluctuating component. The alternative possibility, of introducing a fluctuating entropy, will be considered in the next section. We note only that the sum of the fluctuating component entropies is smaller than the system entropy,

$$\begin{aligned} S'(E') + S''(E'') &= \theta E + [S'(E') - \theta E'] + [S''(E'') - \theta E''] \\ &\leq \theta E - F'(\theta) - F''(\theta) \\ &= S(E). \end{aligned}$$

One can introduce Legendre transformations and minimum principles for the conjugate pairs of variables $(\theta p, V)$ and $(\theta \mu_i, M_i)$ just as for (θ, E) . Specifically, we remark that removing a constraint with separate masses M'_i and M''_i in two containers in favor of their sum $M'_i + M''_i = M_i$ increases the entropy. Physically, the process is diffusion. But the same mathematical apparatus shows that removal of the two constraints M_i and M_j in a single container in favor of their sum yields an increased entropy. This process consists in deciding to identify two different chemical components. Relaxation of the constraints V_1 and V_2 subject to $V_1 + V_2 = V$ can refer to the action of releasing a wall and allowing it to move freely.

7. Log Ω as Entropy



Consider a dynamical system described by the Hamiltonian $\mathcal{H}(q, p, \lambda)$, where λ is a parameter. We assume that \mathcal{H} is the only time-independent constant of the motion and that a motion will, in general, be ergodic on a constant energy surface. Now we suppose that λ is a slowly varying function of the time. By this we mean that both λ and the energy change only slightly in the time required for an orbit to wander all over an energy surface. Over such a time interval, we can plausibly replace a time integral by an average on an energy surface,

$$\delta \mathcal{H} = \int_{t_0}^{t_1} \frac{d\mathcal{H}}{dt} dt = \int_{t_0}^{t_1} \frac{\partial \mathcal{H}}{\partial \lambda} \frac{d\lambda}{dt} dt \sim \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle \delta \lambda,$$

in this way obtaining the adiabatic approximation

$$(7.1) \quad \frac{d\mathcal{H}}{d\lambda} = \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle = \phi(\mathcal{H}, \lambda).$$

However, we can integrate this differential equation quite generally using only the fact that the system is Hamiltonian. We recall that the volume element in phase space is an invariant measure. Therefore the volume within an energy surface, $\mathcal{H} < E$, is the same as the volume within the image surface into which the points of $\mathcal{H} = E$ are carried by the motion. But, by our adiabatic hypothesis, the image of an energy surface remains an energy surface. Thus the function

$$(7.2) \quad \mathcal{V}(E, \lambda) = \int_{\mathcal{H} < E} dq dp$$

is an approximate constant of the motion, i.e., an adiabatic invariant. It is an exact integral of the adiabatic equation (7.1).

Proofs of this type of theorem are meagre. Essentially the only cases that have been treated rigorously are problems in which the ergodic manifold is a closed curve, or other closely related problems (cf. the example of Section 5).

For systems with many degrees of freedom, it is likely that the reference to ergodicity is unnecessary. It appears to be a frequent occurrence that the value of a phase function (possibly $\partial \mathcal{H} / \partial \lambda$) deviates appreciably from its mean value only on a set of very small measure.

Of course, any function of \mathcal{V} is also an adiabatic invariant. For large systems, \mathcal{V} is approximately exponential in n ; thus to obtain an additive quantity which is proportional to the size of the system, we should take $\log \mathcal{V}$ as the entropy. To complete the thermodynamic structure we write

$$(7.3) \quad d(\log \mathcal{V}) = \frac{\partial}{\partial E} (\log \mathcal{V}) dE + \frac{\partial}{\partial \lambda} (\log \mathcal{V}) d\lambda$$

and identify $\partial \log \mathcal{V} / \partial E$ with θ and $\partial \log \mathcal{V} / \partial \lambda$ with θX , where X is the "force" conjugate to the parameter λ . To show that these identifications are proper requires treatment of interacting systems which we defer. But, to obtain more familiar expressions we introduce

$$(7.4) \quad \Omega(E, \lambda) = \frac{\partial}{\partial E} \mathcal{V}(E, \lambda)$$

and note the formula¹³

$$(7.5) \quad \Omega(E) = \int_{\mathcal{H}=E} \frac{d\Sigma}{|\nabla \mathcal{H}|}$$

¹³A. I. Khinchin, *Statistical Mechanics*, Dover, 1949.

where $d\Sigma$ is the “area” element on the energy surface, $d\Sigma/|\nabla\mathcal{H}|$ is the invariant measure on this surface; $d\mathcal{H} d\Sigma/|\nabla\mathcal{H}|$ is the volume element. Now, taking E fixed and λ variable, we can evaluate

$$d\mathcal{V}(\lambda) = \int \frac{d\Sigma}{|\nabla\mathcal{H}|} d\mathcal{H}(\lambda) = \int \frac{d\Sigma}{|\nabla\mathcal{H}|} \frac{\partial\mathcal{H}}{\partial\lambda} d\lambda$$

or

$$\frac{\partial\mathcal{V}}{\partial\lambda} = \int \frac{\partial\mathcal{H}}{\partial\lambda} \frac{d\Sigma}{|\nabla\mathcal{H}|} = \Omega \left\langle \frac{\partial\mathcal{H}}{\partial\lambda} \right\rangle.$$

Summarizing, we have



$$(7.6) \quad \theta = \frac{\Omega}{\mathcal{V}},$$

$$X = \left\langle \frac{\partial\mathcal{H}}{\partial\lambda} \right\rangle.$$



From the exponential behavior of \mathcal{V} for large n , it is easily seen that $\log \Omega$ is asymptotically equal to $\log \mathcal{V}$. For reasons that will appear shortly, it is more convenient to use $\log \Omega$. Also it can be shown in cases where there are integrals other than energy, that the generalization of $\log \Omega$ is the more appropriate choice.



To study fluctuations, we consider two systems in weak thermal contact. The adjective weak implies that the combined Hamiltonian can be approximated numerically by the sum of the values of the individual Hamiltonians ignoring interactions. Thermal contact implies that the Hamiltonian is not exactly a sum; more precisely we make the assumption that the combined Hamiltonian is the only constant of the motion, which motion is ergodic on the energy surface $H = \text{constant}$. Thus the distribution is microcanonical. It is an easy consequence of this fact that the probability distribution of the energy of one component is given by¹³



$$(7.7) \quad P(E') = \Omega'(E') \frac{\Omega''(E-E')}{\Omega(E)}.$$

Here Ω' refers to the component under discussion, Ω'' to the other component (the reservoir) and Ω to the combined system; E' is the (fluctuating) energy of the component and E is the constant total energy.

It is immediately evident that we can interpret $\log P$ as a fluctuating entropy,

$$(7.8) \quad \log P = \log \Omega' + \log \Omega'' - \log \Omega.$$

We have previously identified $\log \Omega$ as the entropy of an isolated system. The quantities $\log \Omega'$ and $\log \Omega''$ are exactly the respective entropies that would be assigned to the two components if they were isolated at the instant when the energies happened to have the values E' and $E'' = E - E'$. Thus the logarithmic probability of observing the values E' and E'' is given by the difference between the instantaneous fluctuating entropy and the system entropy.

Furthermore, in a large system the expression P is peaked near its maximum, since P is approximately the n -th power of some fixed function ϕ . Since P is a normalized probability, the maximum value of ϕ must be approximately unity (otherwise $\int P dE'$ would become exponentially small or large). Thus the maximum value of $\log P$ is approximately zero, and we conclude that the fluctuating entropy is smaller than the system entropy with equality for the most probable state.

Because P is an approximate n -th power, we also know that it can be approximated near its maximum by a Gaussian distribution. This approximation is equivalent to a quadratic dependence on E' for sufficiently small excursions from its most probable value in (7.8).

8. Statistical Thermodynamics

We have described the thermodynamic structure for an isolated system in terms of $\log \Omega$ as the entropy. We continue with a discussion of systems in thermal or diffusional contact and recall that the complete thermodynamic structure is determined if we know one of the functions $S(E, V, N)$, $F(\theta, V, N)$, or $\phi(\theta, \mu)$ (in a microscopic analysis, the number of particles, N , is a more convenient variable than the mass, $M = mN$).

For two systems in thermal contact we recall (7.7),

$$(8.1) \quad P(E') = \frac{\Omega'(E')\Omega''(E-E')}{\Omega(E)}.$$

The normalization $\int P dE' = 1$ yields the convolution relation

$$(8.2) \quad \Omega(E) = \int \Omega'(E')\Omega''(E-E')dE'.$$

We accept these relations as postulates for the moment. They should be considered as approximations depending on an estimate of the coupling energy between the two systems. The convolution property implies

$$(8.3) \quad \log Z = \log Z' + \log Z'',$$

where

$$(8.4) \quad Z(\theta) = \int e^{-\theta E} \Omega(E) dE$$

and similarly for Z' and Z'' . The conventional thermodynamic structure results from an identification of $\log Z$ with $-F(\theta, V)/k$.¹⁴

For two systems which are free to exchange particles as well as energy, we have¹⁵

$$(8.5) \quad P(E', N') = \frac{N!}{N'!(N-N')!} \frac{\Omega'(E', N')\Omega''(E-E', N-N')}{\Omega(E, N)},$$

$$P(N') = \frac{N!}{N'!(N-N')!} Z'(\theta, N')Z''(\theta, N-N').$$

The last formula should be compared with (3.2). We have the normalization $\sum_{N'} \int P dE' = 1$ as well as the convolution relations

$$(8.6) \quad \frac{1}{N!} \Omega(E, N) = \sum_{N'} \int \frac{\Omega'(E', N')}{N'!} \frac{\Omega''(E-E', N-N')}{(N-N')!} dE',$$

$$\frac{Z(\theta, N)}{N!} = \sum_{N'} \frac{Z'(\theta, N')}{N'!} \frac{Z''(\theta, N-N')}{(N-N')!}.$$

This function Ω is the customary one but with N exhibited (the parameter V , or rather the domain of integration, has been suppressed). If there are several species of particles, we replace $N!$ by the product $N_1! N_2! \cdots$. The convolution property implies additivity,

$$(8.7) \quad \log \mathcal{E} = \log \mathcal{E}' + \log \mathcal{E}'',$$

where

$$(8.8) \quad \mathcal{E}(\zeta, \theta) = \sum_{n=1}^{\infty} \frac{\zeta^n Z(\theta, n)}{n!}.$$

The thermodynamic structure arises from an identification of $\log \mathcal{E}$ with θpV and $\log \zeta$ with $-\theta \mu m$ (m is the mass of a particle, $mN = M$).

We have presented three distinct but related thermodynamic structures in which $\log \Omega$, $\log Z$, and $\log \mathcal{E}$ are additive in the respective situations to which they apply. In particular, for fixed ζ and θ , $\log \mathcal{E}$ defines an additive set function which is invariant under translation (provided that we are sufficiently far from any wall potentials); thus it is proportional to the volume, which establishes the homogeneity of this particular thermodynamic structure.

Two problems arise: first to relate the three structures (since macroscopically we find only one); second to reestablish similar results with the correct error estimates for the energy coupling.

¹⁴For dimensional consistency, θ as used in this section is $1/kT$ rather than $1/T$. Thus $\log \Omega$ is S/k , $\log Z$ is $-F/k$, and $\log \mathcal{E}$ will be pV/kT .

¹⁵H. Grad, J. Chem. Phys., *loc. cit.*

If it is known that Ω is an approximate N -th power, then an elementary argument, based on the fact that the integrand in (8.4) is peaked, yields the asymptotic result

$$(8.9) \quad \log Z(\theta, N) \sim -\theta E_0 + \log \Omega(E_0, N),$$

where E_0 is the value which maximizes $-\theta E + \log \Omega$. We recognize the asymptotic approximation to the Laplace transform as the Legendre transformation connecting $F(\theta)$ with $S(E)$. Exactly the same type of estimate in the discrete Laplace transform (8.8) establishes the asymptotic equivalence of the third structure, with the exception of the added term $\log(N_0!) \sim N_0 \log N_0 - N_0$. Thus, by examining those quantities which are additive in a given situation, we are led to the entropies $\log \Omega$ or $\log Z + \theta E$ under circumstances of distinguishable particles and the different entropy $\log \mathcal{E} + \theta E + N \log \zeta$ in a case when particles cannot be distinguished. We remark that, although $\log Z$ is additive in the case of thermal contact, it is not additive, cf. (8.6), when there is diffusional contact. Instead, $\log(Z/N!)$ becomes approximately additive in the limit of large N .

Actually, it is Z which is known to be an N -th power (using $Z = Z' Z'' Z''' \cdots$ for a large system). The reverse procedure, viz., to invert (8.4) to obtain the asymptotic behavior of Ω , is usually done by use of contour integration or the central limit theorem. What is needed is a very weak condition, viz., that the distribution $P_N(E)$ of the sum of N independent random variables has the property $(1/N) \log P_N(E_0) \rightarrow 0$, where E_0 is the mean (actually $P_N(E_0) \sim N^{-1/2}$). Applying this condition to the normalized convolution distribution

$$(8.10) \quad P_N = e^{-\theta E} \frac{\Omega(E, N)}{Z(\theta, N)}$$

yields the equivalence relation (8.9).

For the case of thermal contact, one might postulate that there exists a wall separating the two systems such that the coupling energy is negligible. But, for diffusional contact there is no real wall and one must actually perform the estimate. The asymptotic additivity of $\log(Z/N!)$ has been shown¹⁶. The remaining arguments, indicated above under the assumption $Z = Z' Z''$, can be repeated with error estimates that take into account the coupling. We note only that the error terms have the rather large order $N^{-1/3}$. Thus the conventional asymptotic expansion, of which (8.9) gives the lowest order, cannot without special precautions be used to any higher order, since $N^{-1/3}$ is more important.

¹⁶L. van Hove, *Quelques propriétés générales de l'intégrale de configuration d'un système de particules avec interaction*, Physica **15**, p. 951, 1949. Some of this analysis can be simplified.

We conclude with some further remarks concerning fluctuations. The distribution of energy of a small component in contact with a large reservoir is given by the canonical distribution (8.10).¹⁷ If we further assume that the small component is itself large (to be able to identify $\log \Omega'$ with the entropy), then we see that

$$(8.11) \quad \begin{aligned} \log P' &= -\theta E' + \log \Omega'(E') - \log Z' \\ &= \log \Omega'(E') + \theta E'' + \text{const.} \end{aligned}$$

We verify that the fluctuating entropy of the reservoir is simply $\theta(E'' - E''_0)$, and, by examination of the constant in (8.11), we obtain (8.9) again. This result for the entropy fluctuation of the reservoir is intuitive. Since the energy fluctuation of the reservoir is relatively small, the entropy fluctuation, $dS = \theta dE$, can be evaluated with θ kept constant. An appealing heuristic procedure is to accept the intuitive formula for the entropy fluctuation of the reservoir and the postulate that $\log P$ is the fluctuating entropy to "derive" the formula

$$P(E') = \text{const. exp}\{-\theta E' + S'\}$$

for the energy distribution of a small component.

Received March, 1961.

¹⁷Khinchin, *loc. cit.*