MACROSCOPIC PREDICTION†

E. T. Jaynes
Wayman Crow Professor of Physics
Washington University, Campus Box 1105
1 Brookings Drive, St. Louis, Missouri 63130, U.S.A.

1. INTRODUCTION 1
2. HISTORICAL BACKGROUND 3
3. THE BASIC IDEA 4
4. MATHEMATICAL FORMALISM 7
5. THE MAXIMUM CALIBER PRINCIPLE 11
6. BUBBLE DYNAMICS 12
7. CONCLUSION 15
8. REFERENCES 16

1. INTRODUCTION

Our topic is the principles for prediction of macroscopic phenomena in general, and the relation to microphenomena. Although physicists believe that we have understood the laws of microphysics quite well for fifty years, macrophenomena are observed to have a rich variety that is very hard to understand. We see not only lifeless thermal equilibrium and irreversible approaches to it, but lively behavior such as that of cyclic chemical reactions, lasers, self–organizing systems, biological systems.

On a different plane, we feel that we understand the general thinking and economic motivations of the individual people who are the micro–elements of a society; yet millions of those people combine to make a macroeconomic system whose oscillations and unstable behavior, in defiance of equilibrium theory, leave us bewildered. Governments seem helpless to understand what is happening well enough to predict it, much less control it.

Why is it that knowledge of microphenomena does not seem sufficient to understand macrophenomena? Is there an extra general principle needed for this?

Our message is that such a general principle is indeed needed and already exists, having been given by J. Willard Gibbs over 100 years ago; but it is not fully recognized in the current thinking of either statistical mechanics or economics. A macrostate has a crucially important further property (entropy) that is not determined by the microstate. We hope to discuss the problem in some generality, give a mathematical scheme embodying that principle, and show a few of its consequences.

Of course, any particular mathematical scheme will apply only in a particular area, so we do not want to take a provincial viewpoint tied too strongly to the mathematics. Fortunately, understanding of the problem has advanced to the point where it is possible to explain the principles in a very simple way with almost no mathematics. Only from the simplest viewpoint can one see the full range of application of a principle.

† In *Complex Systems—Operational Approaches in Neurobiology, Physics, and Computers*, H. Haken, Ed.; Springer–Verlag, Berlin (1985); pp. 254–269

A moment's thought makes it clear how useless for this purpose is our conventional textbook statistical mechanics, where the basic connections between micro and macro are sought in ergodic theorems. These suppose that the microstate will eventually pass near every one compatible with the total macroscopic energy; if so, then the long–time behavior of a system must be determined by its energy.

What we see about us does not suggest this. Solids appear to maintain indefinitely whatever macroform they happen to have, although many others would have the same energy. In Egypt's Karnak temple there are stone columns that have supported heavy weights for thousands of years without flowing. In the crystalline texture of the granite facing stones on the building where I am writing this we see the record of the pressure and temperature changes when they cooled, preserved unaltered for over a billion years.

Persistence of macrostructure is not limited to solids. The atoms comprising an elephant could, if reassembled, make instead ten tigers, any twenty humans who have ever lived, or a half–million butterflies. A man who speaks English but not German would, by a convenient rearrangement of some atoms in his brain, be converted into one who speaks both languages.

All of these different configurations of the same atoms, with the same total energy, would preserve their structure and behavior for the life of the organism. Indeed, the macrostructure and behavior persists even when the original atoms are replaced by others.

In short, virtually all the phenomena of Nature are manifestations of the fact that the behavior of macrosystems, over any times of human concern, is *not* determined merely by their energy. Over all the observation time available to us, they remain in extremely small subregions of the energy shell. Surely, no fact is more familiar to us; then we shall not waste any time on ergodicity. Even if ergodic theorems should prove to be true over infinite times, our problem is of a different nature.

The problem to which we address ourselves is: We have some information about a few macroscopic quantities $(A_1, A_2, \ldots)$, for example, their values in certain space–time regions; call this A for short. Given this and any other relevant prior information I that we may have, what are the best predictions we can make of some other macroscopic quantities $(B_1, B_2, \ldots)$ in some other space–time regions? Call this B for short.

Logically, the problem is one of inference (i.e., plausible conjecture) rather than deduction, since in almost all real problems of physics, biology, and economics our information is far too meager to permit any deductive proof that our predictions must be right. Indeed, it is often too meager to justify any definite predictions at all. But even in such cases (or rather, especially in such cases) a normative theory of inference could be useful if it indicates to us what kind of further information would be needed in order to make good predictions.

We are therefore concerned primarily with optimal information processing, only secondarily with physical law (although we shall of course not ignore any physical law that is known and relevant); so if there are general principles for solving it they ought to apply equally well in or out of physics. More specifically, we are asking whether probability theory can be applied to this problem, in the form: given prior information I and data A, what is the probability $p(B|A, I)$ that any particular prediction B will be right?

Such a question has nothing to do with the "random variables" that are so prominent in conventional probability theory. The reason we need inference is not chance or "randomness", but incomplete information. But if the question makes sense in probability theory, it should make sense also in biology and economics, as well as in physics.

Historically, however, such problems first arose in physics; in the next Section we recall briefly what was found there. Section 3 then explains the principles that we propose to use, in essentially verbal rather than mathematical terms. The reader who wants to get on with the new technical content of this work may turn at once to Section 4.

2. HISTORICAL BACKGROUND

It is now more than 160 years since Carnot started the science of thermodynamics with the first statement of what is perhaps the most fundamental of all principles for predicting macroscopic events. But we have not yet understood and exploited all its implications.

At first, progress came rapidly; from Carnot's principle (no heat engine E can be more efficient than a reversible one R with the same upper and lower temperatures t, t') it follows that all reversible engines must have the same efficiency depending only on t, t'. So, as Kelvin saw, the reversible efficiency $e(R)$ must define a universal temperature scale $T(t)$, independent of the properties of any particular substance:

$$e(R) = 1 - T'/T \,. \tag{1}$$

From the first law the actual efficiency is $e = 1 - Q'/Q$; so as Clausius saw, Carnot's inequality $e \leq e(R)$ implies the existence of a new function of the thermodynamic state

$$S = \int dQ/T \tag{2}$$

which he named "entropy", with the property that in a macroscopic change that begins and ends in thermal equilibrium, the total entropy of all bodies involved cannot decrease:

$$S_{final} \geq S_{initial}, \tag{3}$$

ergo, if it increases, the process is irreversible.

All this was accomplished within thirty years of Carnot's work; but further progress has been painfully slow. The theory has been hampered by a serious restriction; the definition (2) of entropy refers only to equilibrium states (the path of integration being a reversible path; i.e., a locus of equilibrium states). In consequence, the Clausius statement of the second law (3) gives us only minimal information about a change of macroscopic state; it predicts that it will go in the general direction of greater final entropy; but not how fast, or along what path. Nor does it predict what final equilibrium state will be reached. Indeed, as the more careful writers have noted, it does not even predict that the entropy will increase monotonically, since the entropy of an intermediate nonequilibrium state was not defined.

It required twenty more years for the next advance of Gibbs (1875–78) which gave the variational principle for determining the final equilibrium state; but the world was not yet ready for him, and only after another fifty years was the first level of understanding of Gibbs' work reached, by G. N. Lewis (1923). After still another fifty years, rereading Gibbs convinced me that G. N. Lewis extracted only about half of the deep insight in that work, which is by no means of mere historical interest today. After over 100 years, we can still learn important "new" technical facts from Gibbs.

We think that this is due mainly to the fact that Gibbs did not live long enough to complete his work. He was much too far in advance of his time to have left any students capable of carrying it on, and he was not the world's clearest expositor; so his thinking had to be partly re–discovered before we could recognize that Gibbs already had it.

Turning now to macrophenomena, it is clear that far more than (3) can be said about how they proceed. We enunciate a rather basic principle, which might be dismissed as an obvious triviality were it not for the fact that it is not recognized in any of the literature known to this writer:

> *If any macrophenomenon is found to be reproducible, then it follows that all microscopic details that were not reproduced, must be irrelevant for understanding and predicting it. In particular, all circumstances that were not under the experimenter's control are very likely not to be reproduced, and therefore are very likely not to be relevant.*

Now in the laboratory we find that irreversible processes follow a definite course; control of a few initial macroscopic quantities is sufficient to yield a reproducible result. We observe this not only close to equilibrium as in heat conduction and viscous laminar flow, but also far from equilibrium as in shock waves, fast chemical reactions, and lasers.

Ergo, information about those initial macroscopic values, together with suitable general prior information I whose nature we have not yet specified, must be sufficient for theoretical prediction of the result. It cannot be necessary to specify the millions of microvariables that were not controlled and would not be the same on successive repetitions of the experiment.

But the Clausius rules (2), (3) presumably defined the initial state by specifying the macrovariables that the experimenter did control or observe; that is his definition of the initial "thermodynamic state", and in his applications it is what we have called the "data" A. Why then could he not predict far more than the increasing tendency of the entropy? He must have ignored a great deal of relevant information; but if he had the same data as we would have today, this could only be the unspecified prior information I.

In biology, it appears that a seemingly small amount of microscopic information (configuration of a few DNA molecules) is sufficient to determine thousands of details of the macroscopic structure of an organism. Biological phenomena are also highly reproducible; *ergo*, given proper theoretical understanding they must be also highly predictable.

One might have thought, then, that progress since Clausius would show us how to recognize and use this missing information I. Indeed, that is just what Gibbs did. But surprisingly, the Clausius statement of the Second law remains the culmination of the subject as taught to physicists. Recent works from which we might have learned something better [1,2] are still ignored in our pedagogy and research literature.

For decades the thermodynamics of physical chemists, trained instead in the Gibbsian principles as expounded by G. N. Lewis, has been ahead of that of physicists. But this theory – at least, the firmly established "exact" part of it that comes from Gibbs – is still limited to equilibrium predictions. A few extensions beyond that domain, such as the law of Mass Action, have the nature of useful approximate rules of thumb, rather than parts of an exact theory.

This is not to say that attempts to extend equilibrium theory have not been made. So many different approaches have appeared – all of the nature of *ad hoc* rules of thumb without foundation in first principles – that the field is reduced to chaos, even the meanings of such common terms as "entropy" and "reversible" being in contention.

3. THE BASIC IDEA

In the following we conjecture about common features that might be in any theory of macrobehavior; thermodynamic, biological, or economic. But the very simplicity of the idea we want to expound makes it difficult conceptually. Our first thoughts run thus: the macrostate is only a kind of blurred image, or projection, of the microstate, with the fine detail removed. *Ergo*, macrobehavior must be determined by microbehavior; there is no room for other considerations. Extra principles beyond those of microbehavior could not be even relevant, much less needed, to predict macrobehavior. As a physics student in the 1940's, the writer was so strongly indoctrinated with this view that it required decades of hard thinking to escape from it.

Of course, we do not suggest that macrosystems can violate the laws of microphysics, or that those laws can be suspended in the interest of any "final cause" or "vital principle". Our extra–microscopic principle is not of that nature at all; surely, if we knew the exact microstate and could integrate the equations of motion exactly, no additional principle would be needed.

The difficulty of prediction from microstates lies, not in any insufficiency of the laws of microphysics to determine macrophenomena, but in our own lack of the information needed to apply them. We never know the microstate; only a few aspects of the macrostate. Nevertheless, the aforementioned principle of reproducibility convinces us that this should be enough; the relevant information is there, if only we can see how to recognize it and use it.

The problem that Gibbs [3] faced in his *Heterogeneous Equilibrium* of 1875 was: given a few macrovariables defining a nonequilibrium state of a system, predict the final equilibrium macrostate that it will go to. But this problem is ill–posed; not enough information is given to determine any unique solution. Nevertheless, Gibbs gave a solution; he took a hint from the Clausius statement of the second law, but changed its logical status drastically. Where Clausius and Planck had tried to see the second law as an "absolute" law of physics like conservation of energy, Gibbs deprived it of that logical certainty – but at the same time made it a more powerful tool for prediction.

Instead of the Clausius weak statement that the entropy "tends" to increase, Gibbs made the strong statement that it *will* increase, up to the maximum value permitted by the constraints (conservation laws and experimental conditions).

But Gibbs recognized also that this prediction cannot claim the certainty of deductive proof. There are initial microstates, allowed by the data, for which the system will not go to the macrostate of maximum entropy. There may be further constraints, unknown to us, that make it impossible for the system to go to that state; perhaps new constants of the motion. So Gibbs was only doing *inference* rather than *deduction*. He was predicting, not what must happen, but only what is most strongly indicated by the information he had. But that is really all that any scientist could ever pretend to do.

In spite of its seemingly shaky logical status, Gibbs' principle has proved to be as successful in practice as any of the "certain" laws of physics. Whenever the equilibrium macrostate is reproducible, Gibbs' rule predicts it with quantitative accuracy. Physical chemistry has been based on this variational principle for two generations [4].

Clearly, then, the missing information "I" that was needed to resolve the ambiguity of the ill–posed problem was contained somehow in the entropy function. Gibbs pointed out that entropy is a property only of the macrostate, not (like energy) of the microstate. But, cautious man that he was, Gibbs never undertook to tell us what entropy "really is" and why his principle works.

Today, macroscopic prediction has barely advanced beyond the level reached by Gibbs. Yet the way to generalize it has been staring us in the face for 80 years; for the meaning of entropy was seen already by Boltzmann, Einstein, and Planck before 1906. It is carved on Boltzmann's gravestone in Vienna:

$$S = k \log W. \tag{4}$$

The thermodynamic entropy of a macrostate (defined by specifying pressure, volume, energy, etc.) is essentially the logarithm of the classical phase volume (or in quantum theory the number of microscopic quantum states) consistent with it; i.e., the "number of ways" the macrostate can be realized.

Gibbs' variational principle is, therefore, so simple in rationale that one almost hesitates to utter such a triviality; it says "predict that final state that can be realized by Nature in the greatest number of ways, while agreeing with your macroscopic information."

But then the generalization we seek is equally obvious and trivial: to predict the course of a time–dependent macroscopic process, choose that behavior that can happen in the greatest number of ways, while agreeing with whatever information you have – macroscopic or microscopic, equilibrium or nonequilibrium. The prediction is not required to be right in the sense of deductive proof; but it is the best prediction we can make from the information we have; i.e., it is an inference.

From simplicity comes generality: everything we have said would seem to apply as well to biology and economics as to physics. The basic property of a macrostate that is needed for prediction is its multiplicity W. The second law of thermodynamics was only the first case of this to be discovered.

That is the entire content of our theory; but its simplicity makes it hard for sophisticated scientists to understand it and accept it. We were hoping to find a theory that would be subtle and recondite conceptually, elegant and intricate mathematically; how could one ever trust such a childishly simple rule?

First, we reassure the reader that, in spite of the conceptual simplicity, its full mathematical expression does prove to be elegant and intricate after all. To treat general space–time dependences we need a functional integral formalism very similar to that of quantum field theory.

Conceptually, we cannot understand why either Gibbs' rule or our generalization of it works until we recognize that principle of reproducibility. Our initial macroscopic information A confines the possible microstates to some class C of possibilities, containing $W(A)$ microstates. That information corresponds to a generalized entropy $S(A) = k \log W(A)$, equally meaningful for equilibrium and nonequilibrium cases. The subsequent macroscopic behavior could not be reproducible unless it was true that for *each* of the great majority of the microstates in C, we would have the *same* macroscopic behavior. But there remain a small minority of "dissenting" microstates that would lead to a different result, and therefore deny our rule the status of logical deduction.

Nevertheless, if our information determining C includes all the conditions that are needed in the laboratory to yield a reproducible result, our rule must predict that result; for it in effect takes a majority vote over class C, thereby suppressing that small minority. The smaller the minority, the more reliable we expect our predictions to be.

As is well known, the minorities are extremely small in the usual thermodynamic situations. For two macrostates A and B, if there is a tiny entropy difference $D = S(B) - S(A)$ corresponding to one microcalorie at room temperature, the ratio of multiplicities is about

$$W(B)/W(A) = \exp(D/k) = \exp(10^{15}) \,. \tag{5}$$

The macrostate of higher entropy can be realized in overwhelmingly more ways; this is the basic reason for the high reliability of the Gibbs equilibrium predictions.

In other applications we cannot expect the ratios to be so large; but there is a long way to go. If the ratio were only exp(10), as it might be in a problem of economics, we would still expect the predictions to be reliable enough for most purposes (in any event, to make any better ones would require more information).

Clearly, new information which does not cause an appreciable contraction of the class C cannot appreciably affect our predictions. Thus the cogency of new information – the degree to which it could help to improve our predictions – is indicated in a general way by how much further reduction in entropy it would achieve.

Now the equations of probability theory are usually presented as rules for calculating frequencies of "random variables" in "random experiments." But we are not forced to think of them in that way; abstract mathematics may be interpreted in whatever way serves our purpose. Probability theory is also the mathematical tool *par excellence* for locating the class C and taking that majority vote.

Indeed, probability theory was originally conceived, by James Bernoulli and Laplace, as a tool for conducting inference. A probability distribution may be used to describe a state of incomplete knowledge, where there is no "random experiment" involved. Its power for this purpose was demonstrated in massive detail by Sir Harold Jeffreys [5], and its uniqueness as the only consistent such tool was proved by R. T. Cox [6]. In recent years there have been exciting new applications of this viewpoint in many areas of science, such as spectrum analysis and image reconstruction; and indeed, any situation where we are obliged to draw the best conclusions we can from incomplete information [7].

4. MATHEMATICAL FORMALISM

In setting up our generalization of Gibbs' variational principle we shall use probability distributions over microstates. It almost never makes sense – and it can weaken our results – to think of the probability of a microstate as a real frequency in any "random experiment". In thermodynamics the imaginary experiment would have to be repeated for perhaps $\exp(10^{24})$ times before there was much chance of that particular microstate appearing even once. In geophysics or economics it is seldom possible to repeat an experiment at all.

Therefore we should think of our probability distributions in the original Bernoulli–Laplace sense, as representing simply our state of knowledge when we have the incomplete information A and some prior information I consisting of whatever we know about the laws of microbehavior and any other information that seems relevant. This gives us the freedom to take advantage of whatever information we have.

For our purposes the quantity W is not yet well defined. Specifying a macrostate is never so precise that it makes one microstate easily possible and an adjacent one impossible. Rather the probabilities on microstates that describe real macroscopic information must shade off to zero in some smooth way, and we need a refined definition of W that takes this into account.

We have noted before [8] that the asymptotic equipartition theorem of information theory has an application to this problem, relating Boltzmann's W to the Gibbs H function in classical theory, and to the von Neumann–Shannon information entropy $H = -Tr(\rho \ln \rho)$ in quantum theory. Consider the latter case; almost everything we say holds *mutatis mutandis* in the former. Given any density matrix ρ with eigenvalues $(r_1 \geq r_2 \geq \ldots)$, let $W(\epsilon)$ be the smallest integer for which

$$\sum_{i=1}^{W(\epsilon)} r_i \geq 1 - \epsilon. \tag{6}$$

Intuitively, $W(\epsilon)$ is the number of reasonably probable microstates, "reasonable" being defined by ϵ. Now we can associate $\log W(\epsilon)$ with H in various ways. The strongest supposes that we have N particles, and correlations fall off to zero at some distance. Then as $N \to \infty$ with the intensive parameters held constant, $N^{-1}[\log W(\epsilon) - H] \to 0$ provided ϵ is not 0 or 1. Remarkably, in the limit it does not matter what we mean by "reasonably probable". This theorem is discussed more fully by Feinstein [9].

Numerical experimentation shows that a similar result holds under wider conditions than have been proved. But rather than going into all the minute details of more rigorous theorems about this connection – which would be in the end irrelevant to our problem – we now recognize that the definition $W \equiv \exp(H)$ is itself the appropriate refinement of the notion of "number of ways" that takes into account the smooth variation of probability. This is obviously correct in the case where W was exactly defined (i.e., when the probabilities r_i are uniform on a subset, zero elsewhere); and any other choice would lead us into conflict with masses of known results in equilibrium statistical mechanics, where the H of the canonical density matrix gives the experimental entropy to quantitative accuracy; and therefore $S = k \log W$ will remain valid with this definition of W.

This gives us a mathematically well posed, and formally elegant, variational principle; given incomplete information A, the best predictions we can make of other quantities are those obtained from the "ensemble" ρ that has maximum information entropy H while agreeing with A. By "agreeing" with A we mean, of course, that the information A can be extracted back from ρ by the usual rule of prediction: $\langle A \rangle = Tr(\rho A)$. This represents the taking of that majority vote.

If A consists of values of constants of the motion (energy, mole numbers, angular momentum, etc.) this prescription leads, as is well known, back to the Gibbs canonical, grand canonical, and rotational ensembles. Thus conventional equilibrium statistical mechanics, with the mathematical apparatus of partition functions, etc., is contained in our proposed rules as a special case.

Extensions to other kinds of information A are straightforward mathematical generalizations of that standard apparatus. We indicate briefly in two stages what is described more fully elsewhere [10]. Let A stand for a set of m real quantities $\{A_1, \dots A_m\}$ such as energy, pressure, magnetization, concentration gradient, etc. and denote their observed values at time $t = 0$ (the data) by A'_k, $(1 \le k \le m)$. We define an m–component vector $\lambda \equiv \{\lambda_1 \dots \lambda_m\}$ of Lagrange multipliers, the scalar product $\lambda \cdot A$, and the partition function

$$Z(\lambda) \equiv Tr \exp(-\lambda \cdot A). \tag{7}$$

Then the density matrix that agrees with the data A' while assuming nothing beyond that – i.e., which spreads the probability as uniformly as possible over all microstates subject to the constraint $Tr(\rho A) = A'$, is

$$\rho = Z^{-1} \exp(-\lambda \cdot A). \tag{8}$$

The Lagrange multipliers are found from the m conditions

$$A'_k = -\frac{\partial}{\partial \lambda_k} \log\, Z, \qquad 1 \le k \le m \tag{9}$$

and from this information the "best" (in the sense that it minimizes the expected square of the error) prediction of any other quantity B is

$$\langle B \rangle = Tr(\rho B) \tag{10}$$

which we may think of as the majority concensus of the likely microstates. If some of the A_k are not constants of the motion, specifying their values at only one time $t = 0$ would not in general lead to equilibrium predictions of other quantities. The density matrix, no longer a function of constants of the motion, would be time dependent in the Schrödinger representation, and the above algorithm would give time dependent predictions.

The situation is not essentially different if some of the $A_k = A_k(x)$ are functions of position, and their observed values $A'_k(x)$ specified in certain regions. The Lagrange multipliers then get promoted to functions $\lambda(x)$, the partition function to a partition functional $Z[\lambda_k(x)]$, all generalizing the above relations in the most obvious way. This is also a rather common algorithm, often miscalled the "local equilibrium" theory on the grounds of a loose analogy with the grand canonical ensemble if the space–varying A's are particle densities. It has a puzzling "induction time" phenomenon; irreversible fluxes do not seem to be going at the initial time $t = 0$, but require a short transient startup period, after which they often settle down to "plateau" values.

Of course, if data referring to time $t = 0$ are all we have, the "local equilibrium" algorithm will represent the best predictions we are able to make. But now knowing the values of the A's at other times becomes relevant and can improve our predictions. In fact, real macroscopic information can hardly refer to only one instant of time; we really know that macroscopic quantities vary slowly on the microscopic scale, so that the values A' have persisted for a short time in the past. Taking this seemingly trivial, unimportant information into account removes the induction time phenomenon, the corrected algorithm then yielding irreversible fluxes by direct quadratures over the initial ensemble. An example [11] shows calculation of the diffusion coefficient; others are entirely analogous.

This illustrates why we lay such stress on the interpretation of probabilities. If one believes that a probability distribution describes a real physical situation, then it would seem wrong to modify it merely because we have additional information. Indeed, the very idea that a probability distribution describes our state of information is foreign to almost all recent expositions of statistical mechanics; but this precludes any possibility of a full prediction of irreversible processes, as may be seen as follows.

The local equilibrium density matrix, that has maximum information entropy subject to given macroscopic values A' at only one instant of time, assigns equal probability to every possible microstate compatible with A', regardless of its past history. Thus it represents a mixture of every possible history by which the system could have reached the macrostate A'.

But the characteristic feature of an irreversible process, which one would think it the main purpose of theory to predict, is the appearance of fading memory effects; the behavior of a system depends on its past history. Since in the local equilibrium distribution all memory of the past has been thrown away, it is in principle incapable of predicting those parts of future behavior that depend on the past. Instead, its predictions are a kind of weighted average over all possible macrohistories.

Indeed, the local equilibrium distribution cannot even distinguish the past from the future; from symmetry, every possible micromotion and its reversed motion are present with equal weight, so the expectation of every flux is zero. The induction time phenomenon is thus only an artifact of the incomplete information being used; but it cannot be corrected until one recognizes that our probability distributions describe information.

The extension to take into account time dependent information is straightforward. If the given information consists of the values of

$$A'_k(x,t), \qquad 1 \le k \le m \tag{11}$$

in the space–time regions R_k, then the corresponding Lagrange multiplier functions are defined in the same regions, so that

$$\lambda \cdot A = \sum_{k=1}^{m} \int_{R_k} d^3x \, dt \, \lambda_k(x,t) \, A_k(x,t) \tag{12}$$

in which $A_k(x,t)$ is the Heisenberg representation operator, and the rest of the formalism is extended in the obvious way.

However, when time dependent information is used, a new terminology will help to avoid confusion. When information entropy is maximized subject to constraints A', the maximum attained is of course a function of the constraints:

$$S(A') = \log Z + \lambda \cdot A' \tag{13}$$

If the A_k are ordinary thermodynamic parameters, this is the Clausius experimental entropy of thermodynamics. For the "local equilibrium" case, it becomes a functional

$$S_0 = S[A'_k(x)] \tag{14}$$

still conventionally called "entropy". The same functional, with the $A'_k(x)$ taken at any time t, then defines a time–dependent entropy S_t which is a property of the macrostate at time t and for which various inequalities can be proved [10].

The notion of entropy varying with time is so ingrained in our thinking that the next stage of generalization, in which the maximum information entropy becomes a functional

$$\sigma_A = \sigma[A'_k(x,t)] \tag{15}$$

of the entire space–time history of the macroscopic process (over the regions R_k where we have information), calls for a new name. We have ventured [10] to all it the *caliber* of the process, since it measures the cross–section of a tube, a bundle of world–lines in "phase space–time", each line representing the time development of a possible microstate consistent with all the given information.

We have shown [11] that this space–time generalization of the canonical ensemble, stated in terms of the partition functional

$$Z[\lambda_1(x,t)\ldots\lambda_m(x,t)] \tag{16}$$

leads automatically to such known results as the Wiener prediction theory and the Kubo formulas for transport coefficients, but in a more general form free of restrictions to quasi–stationary or near–equilibrium conditions. For example, a single formula for predicted particle density encompasses both static diffusion and ultrasonic dispersion and attenuation.

Indeed, near equilibrium the second functional derivatives of $\log Z$

$$\begin{aligned} K_{ij}(x,t;x',t') &= \frac{\delta^2}{\delta\lambda_i(x,t)\,\delta\lambda_j(x',t')}\log Z, \\ &= \int_0^\beta du\,[\langle A_i(x,t-i\hbar u)\,A_j(x',t')\rangle - \langle A_i(x)\rangle\,\langle A_j(x')\rangle] \end{aligned} \tag{17}$$

are the Kubo–type space–time covariance functions of transport theory.

A different – although equivalent – mathematical form of the theory, based on the caliber instead of the partition functional, makes the relation to the work of Einstein, Fokker–Planck, and Onsager clearer. The first functional derivatives of the caliber generate the Lagrange multipliers, or "potentials":

$$\lambda_k(x,t) = \frac{\delta\sigma}{\delta A'_k(x,t)} \tag{18}$$

and at a point where σ_A is locally convex, by which we mean that under a slight change in the problem

$$A'_k(x,t) \to A'_k(x,t) + \delta A'_k(x,t) \tag{19}$$

with $\delta A'_k$ not identically zero, we have

$$\delta\lambda\cdot\delta A = \sum_{k=1}^m \int_{R_k} d^3x\,dt\,\delta\lambda_k(x,t)\,\delta A'_k(x,t) < 0, \tag{20}$$

the second functional derivatives generate a new set of space–time functions

$$G_{ij}(xt;x',t') \equiv \frac{\delta^2\sigma}{\delta A'_i(x,t)\,\delta A'_j(x',t')} \tag{21}$$

which are the functional inverses of the covariance functions $K_{ij}(xt;x',t')$.

If the A_k are ordinary thermodynamic variables, our convexity condition becomes positive definiteness of the matrix G, and they reduce to inverse matrices, $G = K^{-1}$. If in addition we are at an equilibrium point, the matrix G defines the quadratic form in Onsager's expansion of the entropy

$$S = S_0 - \frac{1}{2}\sum_{ij} G_{ij}\,\delta A_i\,\delta A_j + \ldots \tag{22}$$

about that point. Onsager thought of the entropy gradient

$$X_i = \frac{\partial S}{\partial A_i} = -\sum_j G_{ij}\, \delta A_j \tag{23}$$

as the "force" that drives a system to equilibrium according to his phenomenological equations

$$\dot{A}_j = \sum_k L_{jk}\, X_k \tag{24}$$

and argued for the reciprocal relations $L = L^T$.

In Onsager's work there was no apparent connection with the earlier work of Einstein on diffusion in ordinary space or of Fokker–Planck on diffusion in momentum space. But now, if we restate our general prediction algorithm in terms of the caliber, a relation appears.

5. THE MAXIMUM CALIBER PRINCIPLE.

Although the mathematical details needed to carry it out can become almost infinitely complicated, the principle itself remains almost trivially simple in content. We shall describe the principle in generality, then give a simple application of it that combines the Fokker–Planck and Onsager ideas.

We are given macroscopic information A which might consist of values of several physical quantities $\{A_1(x,t)\dots A_m(x,t)\}$ such as distribution of stress, magnetization, concentration of various chemical components, etc. in various space–time regions $\{R_1\dots R_m\}$. This defines a caliber $\sigma_A = \log W_A$, which measures the number of time–dependent microstates consistent with the information A. What predictions should we make of the quantities $\{B_1(x,t)\dots B_n(x,t)\}$ (some of which might be the same as some of the A's) in space–time regions $\{U_1\dots U_n\}$ (some of which might be the same as some of the R's)?

Suppose we just make any guess we please about B and consider the caliber σ_{AB} subject to both the A and B constraints. Of course, since we have imposed a new constraint, we shall have $\sigma_{AB} \le \sigma_A$. If $\sigma_{AB} > 0$, then our guess was a possible one; there exist microstates whose time development would reproduce both the A and B macrobehaviors. But the choice of B which renders σ_{AB} a maximum is the "majority vote" macrobehavior that could be realized by Nature in more ways than could any other consistent with A. So it is an obvious generalization of the principle given by Gibbs in 1875, that this guess is our optimal prediction – optimal in the sense that it takes into account all our knowledge of the microphysics and all our macroscopic data; and makes no arbitrary assumptions beyond that. Whether the prediction is correct or not, to make any better one would require more information than we had.

The caliber of a space–time process thus appears as the fundamental quantity that "presides over" the theory of irreversible processes in much the same way that the Lagrangian presides over mechanics. That is, in ordinary mechanics we learn first that a variational principle (minimum potential energy) determines the conditions of stable equilibrium; then we learn how to generalize this to the Lagrangian, whose variational properties generate the equations of motion. In close analogy, Gibbs showed that a variational principle (maximum entropy) determines the states of stable thermal equilibrium; now we have learned how to generalize this to the caliber, whose variational properties generate the "equations of motion" of irreversible processes.

But these equations of motion are not, except in a certain approximation, differential equations. In general they turn out to be nonlinear integral equations. Close to equilibrium they become linear integral equations, which contain the conventional Kubo relations for linear transport phenomena as special cases. Or, in a short–memory approximation, they reduce to generalized Fokker–Planck–Onsager equations showing how the approach to equilibrium is both steered and stabilized by the entropy function.

Thus define $\{\delta A_k, \delta B_k\}$ as the departures from the equilibrium values. Then close to equilibrium we may use the expansion (22); in a shorthand notation

$$\sigma_A = S_o - \frac{1}{2}\delta A \cdot G_{AA} \cdot \delta A + \dots \tag{25}$$

and

$$\sigma_{AB} = S_o - \frac{1}{2}[\delta A \cdot G_{AA} \cdot \delta A + \delta B \cdot G_{BA} \cdot \delta A + \delta A \cdot G_{AB} \cdot \delta B + \delta B \cdot G_{BB} \cdot \delta B]. \tag{26}$$

Now the reciprocities

$$G_{AB} = G_{BA} \tag{27}$$

turn out to hold trivially in this theory. Therefore σ_{AB} is maximized with respect to B for fixed A, if

$$G_{BB} \cdot \delta B = - G_{BA} \cdot \delta A \tag{28}$$

which is a set of simultaneous linear integral equations determining $\{\delta B_1(x,t) \dots \delta B_n(x,t)\}$. If the caliber is convex, the kernel G_{BB} is positive definite and there is a formal inversion

$$\delta B = - K_{BB} \cdot G_{BA} \cdot \delta A. \tag{29}$$

Of course, this is extremely compact notation, to demonstrate how simple the underlying ideas are. But it is probably beyond our mathematical ability to do the indicated calculations explicitly for any really nontrivial problem; that is perhaps a task for the computers of the next Century. Nevertheless, many formal relations can be extracted from (28), which are capable of being tested experimentally even if we are unable to calculate the covariance functions from first principles. There is no reason to be surprised by this, since it is still true of the equilibrium canonical ensemble that it predicts experimentally testable relations like $(\partial P/\partial T)_V = (\partial S/\partial V)_T$ but we can seldom calculate $(\partial P/\partial T)_V$ exactly from first principles.

6. BUBBLE DYNAMICS

Finally, we sketch hurriedly the short–memory approximation to a prediction problem. Our given information consists of the past behavior of a few macroscopic variables:

$$A_k(t), \qquad 1 \le k \le m, \qquad -\infty < t < 0 \tag{30}$$

and the quantities B_k that we want to predict are their future values. Our full prediction equations would in principle make use of the values of the $A_k(t)$ arbitrarily far into the past. But we may be able to make a "short memory" approximation, in which only information for a short period, perhaps of the order of the mean time between collisions – or in some cases even the duration of a collision – needs to be considered.

We must warn against a common fallacy here. When we say "short memory", we do not mean to imply that the physical system itself can "forget" where it has been. Rather, beyond a certain time further information – unless it is very unexpected – does not help us to improve our predictions. Mathematically, this is because it does not reduce σ appreciably. Conceptually, when we know the macrohistory A over a certain time in the past, information about still earlier times becomes less relevant in comparison with the far more cogent recent information. We should really say "short relevance".

But about the only thing that can happen to the $A_k(t)$ in a very short time is a kind of diffusion. A given macrostate $(A_1 \dots A_m)$ can be realized by an enormous number $W(A_1 \dots A_m)$ of microstates, which would lead to slightly different macrostates a short time later.

Recall that the ordinary diffusion equation for particle density $n(x,t)$

$$\dot{n} = -\nabla \cdot J = D\nabla^2 n \tag{31}$$

expresses a short memory approximation to the correct constitutive equation for particle flux, which has the form

$$J(x,t) = \int d^3x' \int_{-\infty}^{t} dt'\, M(x-x';t-t')\, n(x',t') \tag{32}$$

recognizing that the particles that are here now came on the average from a mean free path away a mean free time ago. The phenomenological diffusion coefficient D is really a space–time integral over a memory function M.

The "irreversibility" in (31) due to the presence of a first time derivative, which some have seen as a deep, mysterious issue, is no more than an expression of the fact that (32) contains information about the past but not the future. Thus (31) represents the best predictions of future density $n(x,t)$ that can be made *from knowledge of its past values*, if the memory time is short compared to the time of observations.

In a similar way, we may introduce a time dependent probability density $P(A_1 \dots A_m;t)$ which represents a "bubble" of probability in the macroscopic state space, which expands due to diffusion and is found after some long arguments to satisfy a phenomenological equation of "bubble dynamics"; slightly oversimplified, it is

$$\frac{\partial P}{\partial t} + (D/k)\nabla \cdot (P\nabla S) = D\nabla^2 P \tag{33}$$

where $S(A) = k \log W(A)$ is the ordinary entropy of thermodynamics; *i.e.*, the entropy of an equilibrium state with the same values of the macrovariables $\{A\}$. More generally, the diffusion coefficient D would be a diffusion tensor.

Equation (33) is like a Fokker–Planck equation in that it has a diffusion effect given by the right–hand side. But it is also like an Onsager equation in that the local entropy gradient is present. It turns out to have beautiful solutions that exhibit both the Fokker–Planck and Onsager behavior. When the entropy is a quadratic function of the A_i, the Green's function is an expansion in Hermite polynomials that can be summed exactly in closed form, so all questions about the solution can be answered. An arbitrary initial distribution $P(A_1 \dots A_m;0)$ relaxes quickly to a gaussian shape, which is retained as it moves to the final equilibrium state:

$$P(A_1 \dots A_m;\infty) \propto \exp(k^{-1}S) = W(A_1 \dots A_m) \tag{34}$$

which is also the final solution of (33) for any entropy function and is of course just the supposition that Einstein made in his discussion of fluctuations; the size of the bubble (34) is the range of fluctuations to be expected.

But because this size is so small, almost any entropy function is accurately a quadratic function in the region of appreciable probability. The result is that, after an initial transient period in which a nongaussian bubble becomes gaussian, to all the accuracy one could want the solution is a gauss bubble moving along a trajectory in A–space and readjusting its size in accordance with the local entropy curvature. In the one–dimensional case, putting $A_1 = x$, the very accurate solution is

$$P(x,t) = \frac{1}{\sqrt{2\pi R(t)}} \exp\left\{-\frac{[x-q(t)]^2}{2R(t)}\right\} \tag{35}$$

where

$$\dot{q} = k^{-1} D S'(q) = L S'(q) \tag{36}$$

$$\dot{R} = 2D[1 + k^{-1} S''(q) R] = -\alpha[R(t) - R_\infty] \tag{37}$$

in which $L \equiv k^{-1} D$, $R_\infty = k|S''(q)|^{-1}$. Equation (36) is of the familiar Onsager form, the system moving along a trajectory defined by the local entropy gradient. In general the Onsager phenomenological coefficients are related to our diffusion tensor by $L_{ij} = k^{-1} D_{ij}$ and the Onsager symmetries $L = L^T$ reduce to the triviality $\langle \delta A_i \, \delta A_j \rangle = \langle \delta A_j \, \delta A_i \rangle$.

But (37) shows a welcome new feature; the curvature of the entropy function stabilizes the bubble if it is convex ($S'' < 0$). If $S''(q) = 0$, and we start from a delta function $R(0) = 0$, (37) gives for the variance

$$\langle \delta x^2 \rangle = R(t) = 2Dt, \tag{38}$$

the Einstein Brownian motion spreading law. With a linear entropy function, bubble dynamics reduces to Einsteinian Brownian motion with superposed steady Onsager drift. But if $S''(q) < 0$, the spreading (38) does not continue forever; the bubble constantly tries to adjust its size to conform to the local entropy curvature.

Thus we can get a very good intuitive understanding of an irreversible process simply from (36), (37) without further mathematics. In fact, the theory leads us to the following picture of macrostate change, which seems general enough to apply to macroeconomics as well (even though we still lack an "economic Liouville theorem" to help us define the right microvariables).

In physics or economics, even though a neighboring macrostate of higher entropy is available, the system does not necessarily move to it. A pile of sand could convert some of its potential energy into heat, thus increasing its entropy, by levelling itself; but it does not actually do this unless there is an earthquake. Any system might just stagnate where it is, unless it is shaken up a little by what an Englishman might call a "dither" of some sort.

As Einstein saw in physics and as we conjecture to be the case also in economics, the dither that prevents stagnation and drives us up the entropy hill is a kind of turbulence injected into the macrovariables by fluctuations in the underlying microvariables. By this means, the macrostate is constantly "exploring the possibilities" of neighboring states. But in this exploration the system is always more likely to move to one of higher than lower entropy, simply because there are more of them (greater multiplicity).

In thermodynamics, the dither is provided by what are usually called "thermal fluctuations". In the present view of the problem we see these fluctuations, not merely as a small diffusion manifestation of the microphysics, but as the active driving force that makes an irreversible process go. Thus "fluctuation–dissipation theorems" and their Hilbert transforms, "fluctuation–response theorems" are not merely curious accidental relations; they express directly the physical mechanism of the irreversible process.

Generally, the dither not only puts random uncertainty into macrovariables, it drives a systematic movement of the macrostate at a drift velocity proportional to

$$(\text{entropy gradient}) \times (\text{mean–square fluctuation})$$

Thus we see that stagnation can have two quite different causes: loss of entropy gradient, and loss of dither. Without an entropy gradient, the sense of direction is lost and the system drifts aimlessly in a way not determined by any macrovariables. In thermodynamic systems at very low temperatures, the dither is nearly lost, and systems may approach equilibrium so slowly that they appear to be frozen in nonequilibrium states.

There is a close analogy in the mechanism of biological evolution. There the dither that drives the process is spontaneous random mutations, as a result of which every species is constantly exploring the possibilities of a slightly different design. Darwin hypothesized that a "good" mutation has a better chance than a "bad" one of surviving, and thus introducing a new, slightly altered form of the species. Of course, this is necessarily true to the extent that a sufficiently bad mutation cannot survive at all. However, species might also diverge merely as a matter of chance, the resulting different subspecies having no particular survival advantage. Ecology may represent a case where the steering from the entropy gradient is very weak, but the dither still operates.

In economics, perceiving the dither but not the entropy factor, Keynes did not find the phenomenon that our model considers one of the fundamental causes of macroeconomic change. This is conceivably the reason why microeconomic theories do not account well for the facts of macroeconomic change.

We find that deterministic, random, and unstable behavior are all exhibited by this model, as follows. A strongly convex (i.e., downward curving, $S'' < 0$) entropy function is strongly stabilizing, leading to a small bubble; i.e., nearly deterministic behavior. When the curvature of the entropy function decreases, stabilizing forces are weaker and the bubble expands, representing more "random" behavior (by which we really mean "less predictable from macrovariables").

When the entropy curvature is zero (i.e., entropy is a locally linear function), the restoring forces are zero, and the bubble spreads following Einstein's law of Brownian motion (dimensions growing as the square root of the time). As noted, it seems plausible that this may be the case for biological evolution.

At a point where convexity of the entropy fails altogether (i.e., it becomes upward curving) we have instability, the bubble stretching out and usually splitting into two smaller bubbles that go their separate ways. In thermodynamics, such a bifurcation signifies a phase transition, the two bubbles representing the development and eventual equilibrium of the two phases. In economics it signifies perhaps a political revolution, the two bubbles representing the different possible societies growing out of the revolution. So, if these economic conjectures prove to be right, enlightened Governments of the distant future will keep an eye not only on the local tilt of the entropy function, but also on its local curvature.

Gibbs explained the phase transitions that occur in multidimensional thermodynamic spaces as "catastrophes" arising from the various kinds of local loss of convexity that can occur in an entropy function in a multidimensional space. We do not know whether all the types classified by René Thom are realizable in thermodynamic systems.

However, as the name implies, bubble dynamics has more content than catastrophe theory, which describes only the equilibrium states and not the dynamics telling us along what path and at what velocity the system gets to those states. Bubble dynamics can, for example, describe the fine details of time development at a bifurcation point, determining what fraction of the bubble will move to the left or right (i.e., given the size, shape, and position of a bubble approaching a bifurcation point, what is the probability that the system will move ultimately to the left or right?).

7. CONCLUSION

We should correct a possible misconception that the reader may have gained. Most recent discussions of macrophenomena outside of physical chemistry concentrate entirely on the dynamics (microscopic equations of motion or an assumed dynamical model at a higher level, deterministic or stochastic) and ignore the entropy factors of macrostates altogether. Indeed, we expect that such efforts will succeed fairly well if the macrostates of interest do not differ greatly in entropy. But there are puzzling cases, as noted in the Introduction, where macrobehavior seems hard to understand in terms of any reasonable dynamics alone. In these cases, the entropy factors may be the missing ingredient; as we learned from Gibbs, prediction of chemical equilibrium could not have succeeded at all until the macroscopic entropy was recognized.

It may have appeared from the above that we have gone to the opposite extreme, and ignored the dynamics. But this was only for expository purposes: to emphasize what is new in our approach, the effect of entropy on macroscopic predictions. In realistic problems the dynamics will reappear automatically in our general equations, since the Heisenberg operators needed to define the covariance functions K and inverse covariance functions G contain the full dynamics. This will remain true in the classical limit. In a mechanical problem, the bubble dynamics equations (36), (37) will acquire new terms representing inertial effects that we have left out here.

8. REFERENCES

[1] H. C. Callen, *Thermodynamics*, J. Wiley & Sons, Inc. New York (1960).

[2] C. Truesdell, *Rational Thermodynamics*, McGraw–Hill Book Co., New York (1969); Second enlarged edition, 1985.

[3] J. Willard Gibbs, "Heterogeneous Equilibrium", Trans. Conn. Acad. Sci. (1875–1878). Reprinted in *The Scientific Papers of J. Willard Gibbs*, Longmans, Green & Co., New York (1906) and by Dover Publications, Inc., New York (1961).

[4] G. N. Lewis and M. Randall, *Thermodynamics and the Free Energy of Chemical Substances*, McGraw-Hill Book Co., New York (1923).

[5] H. Jeffreys, *Theory of Probability*, Oxford University Press (1939); also numerous later editions.

[6] R. T. Cox, *The Algebra of Probable Inference*, Johns Hopkins University Press (1961).

[7] E. T. Jaynes, "Predictive Statistical Mechanics", in *Frontiers of Nonequilibrium Statistical Mechanics*, G. T. Moore & M. O. Scully, Editors, Plenum Press, N. Y. (1986). pp. 33–55. Reprinted in [12]. Contains a long list of references to Maximum Entropy applications up to 1984.

[8] E. T. Jaynes, "Gibbs vs. Boltzmann Entropies", Am. Jour. Phys. **133**, 391–398 (1965). Reprinted in [12].

[9] A. Feinstein, *Foundations of Information Theory*, McGraw–Hill Book Co., N. Y. (1958); Ch. 6.

[10] E. T. Jaynes, "The Minimum Entropy Production Principle", in *Annual Review of Physical Chemistry*, S. Rabinovitch, Editor, Annual Reviews, Inc., Palo Alto, California (1980). Reprinted in [12].

[11] E. T. Jaynes, "Where do we Stand on Maximum Entropy?", in *The Maximum Entropy Formalism*, R. D. Levine & M. Tribus, Editors, MIT Press, Cambridge, Mass. (1978). Reprinted in [12].

[12] E. T. Jaynes, *Papers on Probability, Statistics, and Statistical Physics* R. D. Rosenkrantz, Ed., D. Reidel Publishing Co., Dordrecht, Holland (1983). Reprints of 13 papers dated 1957–1980, with commentary.