INFERENTIAL SCATTERING†

E. T. Jaynes
Dept of Physics, Washington University
St. Louis MO 63130, U.S.A.

Abstract: Some difficult conceptual problems that have plagued Statistical Mechanics from the start, are explained by reference to a very simple experiment, which also explains why the MAXENT formalism gives reliable predictions. Then some of the inner workings of MAXENT are revealed by a general perturbation theorem, showing how a prediction is modified by adding a new constraint. It is illustrated by the example of Rayleigh scattering in acoustics. Here it appears rather like Schwinger's Source Theory in that multiple scattered waves of arbitrarily high order appear already in the first order of the MAXENT perturbation scheme. The result holds in much more general problems of "inferential scattering" in which any statistical inference is modified by new information.

---

† An early version of this was presented under the title 'Generalized Scattering' at the Second Maximum Entropy Workshop, Laramie WY, August 1982, and is in the Proceedings volume, *Maximum–Entropy and Bayesian Methods in Inverse Problems*, C. R. Smith & W. T. Grandy, editors, D. Reidel Publishing Company, Holland (1985); pp. 377–398. The present version was converted to TeX and Times Roman type font for greater legibility, and extensively rewritten for greater clarity, in August 1993

1. HISTORICAL BACKGROUND

In much of Maximum–Entropy Inference (MAXENT) and its ancestor, Gibbsian thermodynamics, we are concerned with a single solution – imposing one set of constraints and examining the resulting distribution and predictions. However, it was shown already by Gibbs (1875) that in thermodynamics new and important facts appear when we consider the relation between two different solutions. We show that the same is true for the MAXENT generalization; indeed, much of conventional physics – and for that matter, conventional wisdom – is contained in a general theorem relating two different MAXENT predictions, before and after adding a new constraint.

Gibbs (1875) gave two relations connecting neighboring thermal equilibrium states. The linear one is the familiar

$$T\delta S - \delta U - P\delta V + \sum \mu_i \delta n_i = 0 \tag{1}$$

where we use the conventional symbols for temperature, entropy, energy, pressure, volume, chemical potentials, and mole numbers respectively. This "Gibbs relation" is in constant use in chemical thermodynamics.

The less familiar quadratic relation (*loc cit*, Eq. 171):

$$\delta T\delta S - \delta P\delta V + \sum \delta\mu_i \delta n_i \geq 0 \tag{2}$$

expresses a basic convexity property, from which Gibbs derived all his conditions for stability. But this convexity may fail at certain critical points, and then we have some kind of "catastrophe" – phase transition, bifurcation, or other instability.[‡] These relations were retained implicitly in Gibbs' final work, *Statistical Mechanics* (1902); but the work was left unfinished and they were not emphasized. Few readers since then have been aware that the Gibbs "canonical ensemble" formalism contains such convexity relations as (2).

In the modern MAXENT formalism these properties are still present; but now they apply to problems of inference in general. Some of these relations have been hinted at, rather cryptically, in Jaynes (1980).

2. INFERENTIAL SCATTERING

The usual scattering theory of physics is also concerned with a relation between two solutions, rather than with a single solution. In physical scattering a wave field is modified by an obstacle that imposes new constraints (typically, new boundary conditions) on the field, and by the "scattered" wave we mean the difference between the two solutions.

By analogy we may define "inferential scattering" in which an inference is modified by new information that imposes new constraints on the entropy maximization. The difference between the two predictions is a kind of "scattering" off the new information. Highly relevant information is information that scatters strongly; that is, makes a large change in subsequent predictions.

But inferential scattering is a more general phenomenon than physical scattering; it need not involve an influence traveling in physical space and time but may, for example, take place in a

[‡] We remark parenthetically that Rene Thom's modern catastrophe theory may be given an alternative mathematical form in terms of convexity of an entropy–like function. This was anticipated by Gibbs, whose first published work (1873) determined conditions for thermodynamic stability as a geometrical convexity property of the entropy, the condition for coexistence of two phases in equilibrium being a local non–convexity which makes it possible for a supporting tangent plane to make contact with the entropy surface at two points. Gibbs' choice of variables has the advantage of avoiding multiple–valued functions; a catastrophe is explained in terms of dimples in a single–valued function instead of folds in a multiple–valued one.

more abstract thermodynamic state space (or in a study of political stability, in a space whose coordinates include the popularity of the leader, the resources of the opposition, and the amount of foreign investment – just to illustrate the range of possible applications).

Generally, new information about any quantity A will change our predictions of any other quantity B that was correlated with A in the original MAXENT distribution. But there is an old adage in statistics: "*Correlation does not imply causation*!"

In particular, when inferential scattering does take place in physical space and time, it need not be "causal" in the physicist's sense of that word. That is, while physical scattering proceeds only forward in time (a perturbation at time t affecting the later state but not the earlier one), inferential scattering runs equally well forward or backward. New information about the present can change our state of knowledge about the past as well as the future; such "backward travelling" inferences are essential for geology. Put differently, we are concerned fundamentally with *logical* relations, which may or not correspond to causal physical connections.

Indeed, in purely classical physics, the "causality" by which new interactions influence the future but not the past, appears in our equations only because of the unsymmetrical information we have put into them. In specifying definite initial conditions without making any allowance for uncertainty about them we are, in effect, claiming exact information about the past – so firmly established that new information about the present cannot change it.

But if we were more honest and admitted some uncertainty about the initial state (representing it by an ensemble of possible past states), then new information about what is happening now would, obviously, change our estimates of what had happened in the past, as well as what will happen in the future. Viewed in this way, we see that backward–traveling terms in the equations of physics are not necessarily paradoxical; indeed, they seem natural and necessary in any statistical theory.*

We believe, as firmly as anyone else, that "You can't change the past". But you can improve your knowledge of the past; that is the goal of virtually everything that is done under the label of "education", and it is a legitimate goal of statistical mechanics.

On the other hand, if our inference is about a causal wave process taking place in space and time with the antecedent state fully specified, and if the new information tells us of a scattering obstacle, we might expect some relation between the change in the MAXENT prediction and conventional physical scattering theory.

Our original aim here was only to investigate this for the particular case of Rayleigh acoustical scattering. However, it developed that we are here running into other deep conceptual problems that have plagued statistical mechanics for two generations; and indeed, lie behind many of the attacks on the Principle of Maximum Entropy itself. These problems must be cleared up first, otherwise what we are about to do will seem incomprehensible or worse to those with conventional statistical training.

Operationally, what we are going to do when given new information is simply to re–maximize the information entropy subject to the new constraints. But this evokes howls of protest from some, who say, "This procedure cannot succeed because you are ignoring the dynamics – merely re–maximizing the entropy is a completely arbitrary procedure and one can have no reliance at all in the results. I would as soon trust the predictions of a crystal ball gazer".

* As another parenthetic remark, our present Quantum Field Theory expresses a kind of mixture of principles of physics and principles of inference; but the latest mathematical formulations of QFT look remarkably like the most general (functional integral) MAXENT formalism. The Feynman propagators with parts that seem to run backward in time, have been puzzling conceptually – but perhaps they may be understood eventually in this way: what is travelling backward is not a physical influence, but only a logical inference.

This is the attitude we have to answer first; we have to explain in clear, physical terms why re–maximizing the information entropy $S_I = -\mathrm{Tr}(\rho \log \rho)$ does, after all, lead us to reliable predictions of reproducible phenomena (which are, we hope, the only ones experimentalists are recording for publication in our scientific journals).

Therefore, we must take a rather long detour to deal with these conceptual problems, which will occupy the next six Sections. Then we develop the general inferential scattering formulas in Sec. 9, and return finally, in Sec. 10, to the MAXENT version of Rayleigh scattering.

3. GENERALIZED FIRST LAW

First, let us note some MAXENT relations reminiscent of the linear Gibbs relations. Some physical quantity A is capable of taking on the values $(A_1, A_2, \cdots A_n)$, where the indices may refer to quantum states but we need not commit ourselves to any particular meaning. It is enough that we can assign corresponding probabilities $(p_1 \cdots p_n)$. Then the expectation of A is

$$\langle A \rangle = \sum_{i=1}^{n} p_i A_i \tag{3}$$

A small change in the problem might involve independent changes in the possible values $\{A_i\}$ and in the assigned probabilities $\{p_i\}$. The change in expectation is then

$$\delta \langle A \rangle = \sum_i [p_i \delta A_i + \delta p_i A_i] \tag{4}$$

But we recognize the first sum as the expected change in A: $\langle \delta A \rangle = \sum p_i \delta A_i$. Therefore we can write (4) in the form

$$\delta \langle A \rangle - \langle \delta A \rangle = \delta Q_A \tag{5}$$

where $\delta Q_A \equiv \sum \delta p_i A_i$.

We call (5) a generalized first law, for the following reason. Suppose $A = E$ is the energy of a system, E_i its value in the i'th quantum state. Then $\langle E \rangle$ is the predicted thermodynamic internal energy function U. On a small change of state (caused, for example, by a change in volume, magnetic field, *etc.*, the work done on the system will be δE_i if it is in the i'th state; thus $\delta W = -\langle \delta E \rangle$ is the predicted work done by the system. Eq. (5) then has the form:

$$\delta U + \delta W = \delta Q_E \tag{6}$$

and since U and W are unambiguously identified, δQ_E is identified as representing heat.

The first law of thermodynamics (or, at least, a relation that is often called the "first law" in textbooks) is seen here as a special case of a general rule: a small change in the predicted value of any quantity – whatever its physical meaning – may be resolved into parts arising from a change in the probability distribution (the "generalized heat") and from a change in the physical quantity (the "generalized work"). And of course, this holds for any small change in the ensemble, however it was specified.

Evidently, the change in information entropy $S_I = \sum p_i \log p_i$ can arise only from the components δQ_A, and not from the δW_A. Thus, facts that were first unearthed by lifetimes of careful experimentation and analysis now correspond to a mathematical relation so trivial that it would pass unnoticed if not pointed out.

But it is just because of its mathematical triviality that we need to stress the physical importance of (5), which has nontrivial implications. In phenomenological thermodynamics the first law relation $dQ = dU + dW$ is, of course, used with complete freedom: given any two of these quantities, the third is determined. It has not always been recognized that we have the same freedom in the statistical theory for any quantity – whether conserved or not – and for any situation, equilibrium or nonequilibrium. In any of these circumstances, there are three basically different means by which our knowledge about a system might change:

(I) Measurements on the system may show a macroscopic change in pressure, magnetization, *etc.* Thus a term of the form $\delta\langle A\rangle$ is known.

(II) We may know from theory that a physical quantity A has changed – for example, according to the equations of motion, or because we have varied an external parameter such as magnetic field – or we may know that the physical nature of A has not changed. Thus a quantity of the form $\langle\delta A\rangle$ is known.

(III) We may know from measurements on another system coupled to the one of interest that a flux of heat, charge, particles, angular momentum, *etc.* has taken place. As we shall see presently, this means that a "source term" of the form δQ_A is known.

The content of (5) is that, given any two of these pieces of information, the third is also known. But the notion of "macroscopic sources" δQ_A has also given rise to conceptual difficulties that have retarded development of statistical mechanics for many years, and we need to clean up some of this unfinished old business before turning to the new.

4. THE BASIC DILEMMA

Suppose we put a pot of water on an electric stove and turn on the burner. How shall we account for the heating of the water in terms of statistical mechanics?

Whether we use quantum theory or classical theory will not matter for the point to be made here. We use the quantum–mechanical notation because it is more concise, but wherever our quantum–mechanical density matrix appears, one may equally well think of it as a classical probability distribution over coordinates and momenta, and our Schrödinger equation of motion (8) as the classical Newtonian equations of motion. What is essential is that both are deterministic and measure–preserving, establishing a 1:1 correspondence between past states and future states.

It is usually taken as axiomatic, in either quantum or classical statistical mechanics, that our probability distributions must evolve in time according to the equations of motion. As a result, as noted in the review article of R. Zwanzig (1965), "thermal driving" was long an awkward topic, workers trying constantly to replace a heat source δQ with some kind of dynamical perturbation (new term in the Hamiltonian) that would have similar effects. In principle, of course, this is quite correct because the real process is indeed dynamical. Let us see what this would entail.

The process takes place by converting the energy carried by two macroscopic coordinates (voltage and current supplied by the Electric Company) into excitation of an increased atomic/molecular motion in an enormous number of microscopic coordinates, and a transfer of that increased motion down a chain of interacting atoms of the burner and pot, to the water molecules.

To describe this process according to statistical mechanics as usually taught, one should introduce all the microscopic coordinates of water, pot, and burner and their interactions, giving a grand total Hamiltonian:

$$H_{tot} = H_{water} + H_{pot} + H_{burner} + H_{interactions} \tag{7}$$

Then the applied voltage and current are put into a given "externally applied" Hamiltonian $H_{ext}(t)$ which is added to H_{tot} whenever the switch is turned on. Then we should solve the Schrödinger equation of motion

$$i\hbar\dot{\rho} = [H_{tot} + H_{ext}(t), \rho(t)] \tag{8}$$

with an initial density matrix $\rho(0)$ given, if the water is initially in thermal equilibrium, by a canonical distribution at the initial water temperature T_i:

$$\rho(0) = Z^{-1} \exp(-H_{tot}/kT_i) \tag{9}$$

where $Z \equiv \text{Tr} \exp(-H_{tot}/kT_i)$ is inserted for normalization, $\text{Tr}(\rho) = 1$.

One would expect, naively, that the solution of (8) after we turn off the switch should tend to a final density matrix that is again canonical:

$$\rho(t) \to \rho_c \equiv Z_f^{-1} \exp(-H_{tot}/kT_f) \tag{10}$$

representing the final higher temperature T_f of the water. Indeed, many authors state this as if they were quite sure that it must be true. For example, Akhiezer & Peletminskii (1981, p. 127) note the time–developed $\rho(t$ and state that as $t \to \infty$, "$\cdots$ the system undergoes a transition into a state of statistical equilibrium, described by the Gibbs statistical operator [our ρ_c], independent of the initial state."

Their failure to prove this assertion is the least of our worries. Even to think about doing the calculation gets us into paradoxes, noted in my review of Akhiezer & Peletminskii (Jaynes, 1982), that are not resolved in even the latest textbooks on statistical mechanics.

Here is the difficulty with (10): It is a theorem that, under the equations of motion $\rho(t)$ undergoes a unitary transformation, which cannot take two different initial density matrices into the same final one. More specifically, in a unitary transformation (a) each individual eigenvalue of $\rho(t)$ is constant in time; (b) the quantity $-\text{Tr}(\rho \log \rho)$, usually interpreted as the thermodynamic entropy, cannot change; and (c) since the eigenvalues of ρ_c in (10) are different from those of $\rho(t)$ as determined by (9), the density matrix at a later time $t > 0$ can never become canonical at a higher temperature!

Then we are hardly surprised by their failure to demonstrate this transition into ρ_c. Yet we know, as about the most familiar experimental fact in thermodynamics, that the water temperature rises by an amount that one can predict correctly without knowing a thing about those microscopic coordinates and interaction forces. Instead of all those billions of microscopic details, we need in practice only two macroscopic numbers: the total energy supplied, and the total heat capacity of the system. Surely if the reason for such a universal fact were really understood, it would be easy to give a simple theoretical proof of it.

So here is the basic dilemma of conventional statistical mechanics: If we deny the validity of the $\rho(t)$, evolved from the dynamics according to (8), we are denying that the system obeys the Schrödinger equation. If we deny the validity of the Gibbsian canonical ρ_c in (10) we are denying experimental facts. Yet it is a theorem that $\rho(t)$ and ρ_c are incompatible, in the sense that they can never become equal.

Each writer must find his own way around this circumstance, and we are not surprised to find that no two writers have done this in the same way. We suspect that, from the death of Gibbs in 1903 to the discoveries made by Wm. C. Mitchell in his 1967 thesis, no person in the world could have given the correct answer.*

---

* It appears to us that of all the writers on the subject, Terrell Hill came closest to seeing the truth here. Others failed utterly to comprehend the relation between the abstract mathematics and the real world. Invariably, the failure was due to the Mind Projection Fallacy; failure to perceive that a probability distribution is not an external reality, only a creation of our own minds as an aid to reasoning. Therefore it does not describe reality, only our own incomplete information about reality. As their writings reveal, Maxwell and Gibbs understood this perfectly well; and so for them there was no paradox.

Surely, it ought to be considered a major scandal that statistical mechanics, as usually taught today, is helpless to describe the most familiar of all thermodynamic experiments. To understand what is happening functionally, we need to examine more closely: on which type of problems have the methods typified by (8) and (10) been successful?

For predicting the behavior of a system initially in thermal equilibrium, when an external perturbation takes it into a nonequilibrium state, we have full confidence in the dynamically evolved $\rho(t)$ generated by (8), (9). For example, it gives all the intricate details of multiple spin echoes (Slichter, 1978). Indeed, our confidence in $\rho(t)$ is so great that discovery of a single case where it can be proved to fail would shake the foundations of physics and merit a dozen Nobel Prizes.

But we have an almost equal confidence in the Gibbsian ρ_c of (10); in every case where one has succeeded in doing both the calculations and the experiments, it has led us to quantitatively correct predictions of equilibrium properties. The intricate details of ortho– and para–hydrogen provide an impressive example of this success.

In short, ρ_c has never failed us for the case of thermal equilibrium; and $\rho(t)$ has never failed us for small departures from thermal equilibrium. We do not expect either to fail us here.

The dilemma appeared only because workers had expected $\rho(t)$ to predict final equilibrium *in the same way* that ρ_c does. That is, we were making an unconscious hidden assumption, rather like that of absolute simultaneity in pre–relativity physics. There is a paradox only if we suppose that a density matrix (*i.e.* a probability distribution) is something "physically real" and "absolute".

But now the dilemma disappears when we recognize the "relativity principle" for probabilities. A density matrix (or, in classical physics, a probability distribution over coordinates and momenta) represents, not a physical situation, but only a certain *state of knowledge* about a range of possible physical situations.

The results $\rho(t)$ and ρ_c are both "correct" for the two different problems which they solve. They represent different states of knowledge about the final condition of the water; but that does not mean that they make different predictions of the *observable macroscopic properties* of the hot water.

Thus in Eq. (8) the constancy of

$$S_I = -k\mathrm{Tr}(\rho \log \rho) \tag{11}$$

ceases to be paradoxical as soon as we recognize that S_I is not in general the same as the phenomenological thermodynamic entropy. It is rather the information entropy, essentially (by Boltzmannn's $S = k \log W$) the logarithm of the number W of "reasonably probable" quantum states in whatever ensemble we may have before us, however defined. In the MAXENT principle, S_I is the thing we maximize to define our initial density matrix.

On the other hand, the phenomenological entropy S_E of the experimenter is by construction a function $S_E(P, T, M, \ldots)$ of the experimentally observed macroscopic quantities $(P, T, M, \ldots)$. The relationship between these entropies has been demonstrated before (Jaynes, 1963, 1965). The experimentally measured entropy S_E of Clausius is only the *upper bound* of the von Neumann–Shannon information entropy $S_I = -\mathrm{Tr}(\rho \log \rho)$ over all density matrices ρ that agree with the constraints. *Only after it has been maximized subject to the constraints of the experimenter's data* does S_I become equal to S_E.

Therefore, as we have stressed before, the constancy of (11) under the equations of motion, far from presenting a difficulty for the second law, is precisely the dynamical property we need to demonstrate that law, in the Clausius adiabatic form $S_E(final) \geq S_E(initial)$.

Given any ensemble ρ, to ask "What information is contained in this ensemble?" is the same as asking, "With respect to which constraints does this ensemble have maximum S_I?" We can answer this at once for both $\rho(t)$ and ρ_c.

For $\rho(t)$ the initial density matrix (9) has maximum S_I for prescribed initial energy $E_{initial}$ of the cold water; calling this maximum S_i, the multiplicity $W_i = \exp(S_i/k)$ is therefore essentially the number of quantum states that have energy near $E_{initial}$ ("near" meaning within the range of thermal fluctuations).

For all practical purposes, one could think of the density matrix $\rho(t)$ as assigning uniform probabilities to these W_i states, zero probability to all others; this corresponds to the "Asymptotic Equipartition Theorem" of Information Theory.

The dynamical evolution (8) induces a unitary transformation of $\rho(t)$ which does not lose any information, and therefore always defines a "high–probability set" containing the same number W_i of states, each being the time development of one of the initial states.

But the dynamically evolved $\rho(t)$ at later times $t > 0$ would indicate, by an increase in the predicted energy $\langle H_{water} \rangle$, that the water is being heated. Although we cannot actually carry out the calculation (8) we believe, with absolute confidence, that this calculation would yield the correct final energy E_f of the hot water (at least, anyone who can disprove this will start a major revolution in physics).

In contrast, to determine the canonically assigned final density matrix ρ_c in (10) we need no microscopic details, only the amount of heat δQ delivered to the water. The result has maximum S_I for the prescribed final energy $E_f = E_i + \delta Q$, and its high-probability region of phase space would contain about $W_f = \exp(S_f/k)$ states, the number that are "near" E_f.

The two ensembles $\rho(t)$, ρ_c agree on the value of E_f; in what way are they different?

The difference is that the calculation (8) would tell us much more than the final energy E_f of the water; it would also indicate, out of all the W_f quantum states that have energy near E_f, a small subset of only W_i states. These are the particular states that could have arisen from the exact history (initial temperature and details of heating) by which that final state was reached.

Now all our experience tells us that the reproducible properties of hot water depend only on its present temperature; and not on the details of the particular history by which it got to that temperature. Therefore, while the calculation (8) of $\rho(t)$ is not in any way "wrong" for this problem; it is inefficient. It requires us to calculate some microscopic details that are irrelevant to our purpose.

Let us get some idea of how much more detail is contained in the dynamically evolved $\rho(t)$ than in the canonically assigned ρ_c.

5. THOSE NUMBERS

Every morning I heat about a quart, or $2 \times 453/18 = 50$ moles, of water to the boiling point to make coffee. The molar heat capacity of water is about $9R$, where $R = 6 \times 10^{23} k$ is the gas constant, and k is Boltzmann's constant. So the water absorbs about $Q = 50 \times (373 - 293) \times 9R = 72$ kilocalories of heat, and its entropy increases by about

$$S_f - S_i = 50 \times 9R \log(373/293) = 6.5 \times 10^{25}\, k\,. \tag{12}$$

Therefore, the ratio of the number of states in the two density matrices is about

$$W_f/W_i = \exp[(S_f - S_i)/k] = \exp(6.5 \times 10^{25}). \tag{13}$$

By contrast, the number of microseconds in the age of the universe is only about $10^{24} = \exp(55)$. Had I heated only a cubic millimeter of water through 10^{-3} degrees C, the ratio would still be about $\exp(10^{15})$.

The appearance of such numbers in statistical mechanics was noted long ago by both Boltzmann and Planck, and it was stressed in the textbook of Mayer & Mayer (1940). For reasons we cannot explain, these numbers seldom appear in modern works; yet it is essential to know about them in doing practical calculations.

Thanks to these numbers, an experienced practitioner of the art can get away with approximations that would appear horrendously bad to one not in on the secret. For example, if $W = \exp(10^{25})$, then if we make an error by a *factor* of 10^{1000} in the calculation of W, this leads to an error of only two parts in 10^{22} in the value of $\log W$.

Planck called this phenomenon "The insensitivity of the thermodynamic functions". In our present problem it means that in setting up ρ_c we need not bother with specifying the exact width $\delta E = (\langle E^2 \rangle - \langle E \rangle^2)^{1/2}$ of that range of thermal energy fluctuations within which we are counting the number of states W_f.

In fact, for a one–mole system δE is of the order of $kT\sqrt{n}$, where n is the number of "effective degrees of freedom" of the system, about 10^{24}. But if we took δE as 10^{1000} times too large or too small, it would have an absolutely negligible effect on our calculation of the experimental entropy $S_E = k \log W_f$, and therefore the heat capacity and equation of state, of that hot water. In re–maximizing the information entropy S_I, a single constraint on $\langle E \rangle$ (which already implies about the right δE) will suffice to accomplish all that we could get by using two constraints, specifying also $\langle E^2 \rangle$ and therefore δE.

6. SO WHY DOES MAXENT WORK?

We have just seen, in this water heating episode, that although $\rho(t)$ and ρ_c predict the same energy for the hot water, the ratio W_f/W_i, or

$$\frac{(number\ of\ high\ probability\ states\ in\ \rho_c)}{(number\ in\ \rho(t))}$$

is fantastically large; in other words, $\rho(t)$ contains enormously more information about the state of the hot water than does ρ_c. This makes the entropy re–maximization that leads to ρ_c appear, if anything, even more precarious than crystal–ball gazing.

Yet the experimental fact is that ρ_c works, yielding the correct predictions of observable properties of that hot water by a calculation that, while not exactly trivial, is simpler by many orders of magnitude than that for $\rho(t)$. This is the fact that is not understood in the conventional statistical mechanics of our textbooks. But if we can learn how to understand it, we shall see why MAXENT works in much more general situations.

On closer examination we see that the useful predictions we can make from ρ_c are not greatly different from those we could make if we had the greater information contained in $\rho(t)$. Indeed, if we are interested in predicting only reproducible effects, we expect no difference at all in their predictions. For when we repeat the experiment we do not repeat all the microscopic details that were assumed known in $\rho(t)$.

The pot is never put on the stove twice in exactly the same position to atomic accuracy, and it is never filled twice with exactly the same number of water molecules. Therefore H_{tot} is never the same twice. The switch is never turned on at exactly the same point in the AC cycle, and the Electric Company never supplies exactly the same voltage and current; therefore the unitary transformation of the equations of motion generated by (8) is never the same twice, to anything remotely like molecular accuracy.

In short, there would be an entirely different $\rho(t)$ for every repetition of the experiment. Indeed, in view of the smallness of the high–probability sets W_i compared to W_f we could repeat this water heating every day for millions of times the age of the universe, with almost no chance that any specific quantum state would appear in the W_i set on two different days. On repetitions of the experiment, the tiny sets W_i would be scattered about at random, like stars in the sky, within the MAXENT set W_f.

How then could the effect of the heating Q be reproducible? Evidently, it must be true that all those intricate details contained in $\rho(t)$, that determine a particular set W_i, are *irrelevant* for predicting reproducible effects of the heating δQ. We could as well have used the big set W_f which is the union of all the little sets W_i.

At this point, we finally see why re–maximizing entropy is superior to crystal ball gazing; while it does not take into account all those billions of microscopic details that don't matter and that we never possess anyway, it does take into account all the information that is actually *relevant* for predicting reproducible phenomena.

In the laboratory, a reproducible result can depend only on properties of the microstate that are the same on successive repetitions of the experiment; in the cases we are considering, the only such constant thing is the source strength δQ_A itself. We expect, then, that in the MAXENT theory, or indeed in any rational theory, *information* about that source strength should suffice to predict any reproducible effects that are caused by it. That is, any such effect should be predictable from the ρ_c that incorporates the information about that source strength.

This concludes our rather lengthy sermon; now let's get back to the constructive development of the mathematics that realizes this program in real situations, and see whether it actually works as just supposed.

7. MACROSCOPIC PREDICTIONS

What macroscopic effects may result from operation of a source δQ_A? In general this will cause internal readjustments, in the course of which some other quantity B may be changed. Supposing δQ_A to be so small that the ensemble is only slightly modified, the amount $\delta\langle B\rangle$ of that change is given by the general variational property of neighboring canonical ensembles, given by Gibbs. However, as Mitchell (1967) showed, the answer can be reasoned out heuristically but more generally, without invoking canonical distributions.

A really careful exposition would have to discuss a number of technical qualifications on the following, but lacking the time and space for it, we ask the reader's indulgence for our aim of expounding only the essential ideas. We believe that anyone who perceives the need for some qualifications here and there, will also see how to supply them for himself.

If in the original ensemble ρ_0 the quantities A and B are positively correlated; *i.e.* they have a positive covariance

$$K_{AB} \equiv \langle AB\rangle - \langle A\rangle \langle B\rangle > 0\,, \tag{14}$$

then in the high–probability set (HPS) of W_0 states picked out by ρ, microstates of higher than average A tend to be also states of higher than average B; and vice versa. Evidently, if we now learn that $Q_A > 0$, the new ensemble with remaximized information entropy will assign higher probability to states of high A; clearly, this will lead us to expect that B has also increased, although it may not be obvious by how much.

But now, consider another quantity C; if it is uncorrelated with A in the initial ensemble, $K_{CA} = 0$, then in the HPS there is no tendency for states of high A to have either higher or lower than average C. Then knowing δQ_A gives us no reason to expect that C has increased rather than decreased; and our prediction of C should be unchanged: $\delta\langle C\rangle = 0$.

This holds for any quantity C that is uncorrelated with A. Therefore, make the choice $C = B - xA$, where x is any fixed number. Then $K_{CA} = K_{BA} - xK_{AA}$, which vanishes if $x = K_{BA}/K_{AA}$, and then $\delta\langle C\rangle = \langle B\rangle - x\langle A\rangle = 0$.

So we have the rule: when a small macroscopic source δQ_A operates and thereby affects any other quantity B internally, the predicted change in B is

$$\delta\langle B\rangle = \frac{K_{BA}}{K_{AA}}\,\delta Q_A \tag{15}$$

This agrees with a more rigorous perturbation treatment of canonical ensembles, but also holds more generally. Indeed, much of the theory of regression in Statistics textbooks is based on a result formally identical with (15).

Now let us indicate the MAXENT distribution more explicitly. We have a number of physical quantities $(A_1 \cdots A_m)$ and associated Lagrange multipliers, or "potentials" $(\lambda_1 \cdots \lambda_m)$. For brevity, write their inner product as

$$\lambda \cdot A = \sum_{k=1}^{m} \lambda_k\, A_k \tag{16}$$

As written, this form includes all those treated by Gibbs. But now these quantities might depend on time and/or position, and the quantity A_k may have a "source region" consisting of some space–time domain R_k.

The covariance of two quantities A, B is a function of whatever parameters are in A, B, so we may have a space-time covariance function $K_{AB}(x,t;x',t')$ which now begins to resemble a Green's function of physics. With such space–time dependences, the partition function and entropy functions of Gibbs become promoted to functionals (Jaynes, 1978, 1980), and the MAXENT formalism strongly resembles that of quantum field theory. But for present simpler purposes we may accomplish nearly the same thing while retaining a Gibbs–like form (16) of our equations, by defining our physical quantities to be localized to small space–time regions.

Our partition function is then

$$Z(\lambda_1 \cdots \lambda_m) = \mathrm{Tr}\,\exp(-\lambda \cdot A) \tag{17}$$

and the MAXENT density matrix is

$$\rho_0 = Z^{-1}\,\exp(-\,\lambda \cdot A) \tag{18}$$

whose entropy is

$$S = (S_I)_{max} = \log Z + \lambda \cdot A \tag{19}$$

and the potentials λ_k are determined from the experimenter's data $(A'_1 \cdots A'_m)$ by the m simultaneous equations

$$A'_k = \langle A_k\rangle = -\frac{\partial}{\partial\lambda_k}\log Z, \qquad 1 \le k \le m\,. \tag{20}$$

These relations merely summarize the standard MAXENT formalism, still another time, but in our present notation.

The two Gibbs relations (1), (2) noted in the Introduction are now generalized to two identities connecting neighboring MAXENT distributions:

$$\delta S = \lambda \cdot [\delta\langle A\rangle - \langle\delta A\rangle] = \lambda \cdot \delta Q \tag{21}$$

$$\delta\lambda \cdot \delta\langle A\rangle \le 0\,. \tag{22}$$

8. MEANING OF THE GIBBS CONVEXITY

To see the meaning of (22), suppose that our original MAXENT ensemble is based on knowledge of only two physical quantities and write $A_1 = A$, $A_2 = B$. Now a "heat-like" source δQ_A operates; *i.e.* the generalized work $\langle \delta A \rangle$ is zero. But B is unconstrained; *i.e.* it is allowed to adjust itself in response to this source. As expounded above, we shall re–maximize the entropy to take account of this new information. But we know only the change in A : $\delta\langle A \rangle = \delta Q_A$ and have to infer that of B from the MAXENT principle. In going to this slightly different MAXENT distribution we expect both potentials λ_a, λ_b to change, so we would have

$$\delta\langle A \rangle = \frac{\partial \langle A \rangle}{\partial \lambda_a}\,\delta\lambda_a + \frac{\partial \langle A \rangle}{\partial \lambda_b}\,\delta\lambda_b. \tag{23a}$$

$$\delta\langle B \rangle = \frac{\partial \langle B \rangle}{\partial \lambda_a}\,\delta\lambda_a + \frac{\partial \langle B \rangle}{\partial \lambda_b}\,\delta\lambda_b. \tag{23b}$$

But by the general MAXENT reciprocity theorem these coefficients are just the covariances:

$$K_{AB} = K_{BA} = -\,\frac{\partial \langle A \rangle}{\partial \lambda_b} = -\,\frac{\partial \langle B \rangle}{\partial \lambda_a} \tag{24}$$

and so (23) is in matrix form

$$\begin{pmatrix} \delta\langle A \rangle \\ \delta\langle B \rangle \end{pmatrix} = \begin{pmatrix} K_{AA} & K_{AB} \\ K_{BA} & K_{BB} \end{pmatrix} \begin{pmatrix} \delta\lambda_a \\ \delta\lambda_b \end{pmatrix} \tag{25}$$

Now if we substitute this into the Gibbs convexity relation (22), it reduces to the statement that, when the inequality holds for all small but non–zero changes, the covariance matrix in (25) is positive definite. Thus (25) can be inverted, and the potentials are uniquely determined by $\langle A \rangle$ and $\langle B \rangle$. This is just the statement that our MAXENT conditions (20) determining the potentials, have a unique solution.

We could imagine more general constraints on MAXENT than specifying expectations $\langle A_k \rangle$. The constraints might themselves take the form of inequalities rather than equalities. But if the constraints confine us to any convex set in the $\langle A_k \rangle$, the solution is still unique.

More generally, the Gibbs convexity relation tells us that when the inequality holds, the eigenvalues of the covariance matrix of any number of quantities are all positive. This makes it clear why Gibbs found – in the phenomenological theory of heterogeneous equilibrium, twenty-five years before his Statistical Mechanics – that the relation (2) was the fundamental key to understanding thermodynamic stability, however many components and phases a thermodynamic system may have.

9. MITCHELL'S RELATIONS

But there is a still more interesting result contained in the above relations. Now notice that if $\delta\lambda_b = 0$, (25) reduces to

$$\begin{aligned} \delta\langle A \rangle &= -K_{AA}\,\delta\lambda_a = \delta\,Q_A \\ \delta\langle B \rangle &= -K_{BA}\,\delta\lambda_a \end{aligned} \tag{26}$$

or

$$\delta\langle B \rangle = (K_{BA}/K_{AA})\,\delta Q_A\,, \tag{27}$$

which is identical with the prediction rule (15) that we reasoned out in an entirely different way!

To make a long story short, the situation uncovered by Mitchell (1967) shows that when a source δQ_A operates and an unconstrained quantity B readjusts itself as a result, to predict the amount of that readjustment there are three principles:

(I) Quantities C uncorrelated with A are unchanged.
(II) Potentials λ_b of unconstrained quantities B are unchanged.
(III) The information entropy is re–maximized.

Mitchell discovered the remarkable fact that these three conditions are mathematically equivalent. It is remarkable because they seem so different to our untutored intuition. Almost everybody finds (I) so intuitive that he will accept it at once, without demanding any formal proof. But to many, (II) and (III) are so far from being intuitive that they will scarcely believe them even after seeing the proof. This shows how much our intuition can be educated by studying the MAXENT formalism and thinking hard about why and how it works.

Mitchell's next relation introduces us to inferential scattering. Introduce a third variable C, so that on a small change in the MAXENT distribution

$$-\begin{pmatrix} \delta\langle A\rangle \\ \delta\langle B\rangle \\ \delta\langle C\rangle \end{pmatrix} = \begin{pmatrix} K_{AA} & K_{AB} & K_{AC} \\ K_{BA} & K_{BB} & K_{BC} \\ K_{CA} & K_{CB} & K_{CC} \end{pmatrix} \begin{pmatrix} \delta\lambda_a \\ \delta\lambda_b \\ \delta\lambda_c \end{pmatrix} . \tag{28}$$

Now as soon as we recognize that these infinitesimal changes are related linearly, all the intuitive understanding we may have of other linear systems can be applied immediately. For example, if we think of $\{\delta\langle A\rangle,\, \delta\lambda_a\}$ as analogous to current and voltage at the a'th port of an electrical network, then (20) describes quantitatively the observable "black box" properties of a 3–port network, in which the covariances K_{ij} are the elements of the admittance matrix (inverse of the impedance matrix). In this analogy, the source δQ_A corresponds to a current injected at the a'th port.

If the source δQ_A operates and both B and C are left free to readjust to this (*i.e.*, in the electrical analogy, ports B and C are short–circuited, offering no resistance to a current flow), we have by the above principles

$$\begin{aligned} \delta\langle B\rangle &= (K_{BA}/K_{AA})\,\delta Q_A \\ \delta\langle C\rangle &= (K_{CA}/K_{AA})\,\delta Q_A \end{aligned} \tag{29}$$

amounting to two independent applications of our rule. But now let us impose a new constraint, that $\delta\langle C\rangle = 0$ (port c is open–circuited). Writing out the bottom line of (28), we can solve for $\delta\lambda_c$:

$$K_{CC}\,\delta\lambda_c = -K_{CA}\,\delta\lambda_a - K_{CB}\,\delta\lambda_b \tag{30}$$

and substituting this into (28) we find that the changes in $\langle A\rangle$ and $\langle B\rangle$ are still related by an equation like (25), but with a new "renormalized" covariance matrix:

$$\begin{pmatrix} \delta\langle A\rangle \\ \delta\langle B\rangle \end{pmatrix} = \begin{pmatrix} K'_{AA} & K'_{AB} \\ K'_{BA} & K'_{BB} \end{pmatrix} \begin{pmatrix} \delta\lambda_a \\ \delta\lambda_b \end{pmatrix} \tag{31}$$

with the new matrix elements

$$\begin{aligned} K'_{AA} &= K_{AA} - K_{AC}K_{CC}^{-1}K_{CA} \\ K'_{AB} &= K_{AB} - K_{AC}K_{CC}^{-1}K_{CB} \end{aligned} \tag{32}$$

and so on. By our principle, $\delta\lambda_b$ will still be zero if B is left free to readjust, so the predicted change in B due to the source δQ_A is now

$$\delta\langle B\rangle = \frac{K'_{BA}}{K'_{AA}}\,\delta Q_A\,. \tag{33}$$

The difference between (33) and (27) represents inferential scattering; the logical connection between A and B is altered by the new constraint $\delta\langle C\rangle = 0$.

We shall examine the meaning of every term in this difference. Define the correlation coefficient of A and C:

$$R_{AC} \equiv \frac{K_{AC}}{(K_{AA}K_{CC})^{1/2}}\,. \tag{34}$$

Then we can write (33) as

$$\delta\langle B\rangle = \left[\frac{K_{BA}}{K_{AA}} - \frac{K_{BC}}{K_{CC}}\,\frac{K_{CA}}{K_{AA}}\right]\delta Q'_A \tag{35}$$

where

$$\delta Q'_A \equiv \frac{\delta Q_A}{(1-R^2_{AC})} \tag{36}$$

is a "renormalized source strength", whose significance will appear presently. On the right–hand side of (35) we have two terms, the first representing the effect we would have without the constraint on C if the renormalized source strength had operated. The last term in (35) can be interpreted by rewriting it as

$$\frac{K_{CB}}{K_{CC}}\,\text{“}\delta Q_C\text{”}\,. \tag{37}$$

This is the response to a fictitious source strength

$$\text{“}\delta Q_C\text{”} \equiv -\frac{K_{CA}}{K_{AA}}\,\delta Q'_A \tag{38}$$

which we recognize as minus the change $\langle C\rangle$ that would be produced in (29) by the renormalized source $\delta Q'_A$ if C were unconstrained.

The new constraint $\delta\langle C\rangle = 0$ has therefore modified our predicted relation between A and B in two ways:

(I) The source strength δQ_A is renormalized; intuitively, if A and C are correlated (positively or negatively), then holding $\langle C\rangle$ fixed makes the system "stiffer" against an attempt to change A, and a given actual change δQ_A has a greater effect on B because of this. As an analogy, if the input impedance to an electrical network is increased by blocking an internal current path, then to inject a given current into it will in general result in increased voltages at other points.

(II) A new scattering term appears which, as 'seen from' B, appears to come from a fictitious source at C. Our electrical network analogy still holds; a point C where the current was blocked, becomes a new voltage source whose effects propagate to other points of the network.

But the relations just found are of far more general meaning than those of the network. A, B, C may stand for any physical quantities, not necessarily localized in space or time.

10. ACOUSTICS – DIRECT PROPAGATION

Finally, we are ready for the promised specific case. In an acoustical problem, take our initial ensemble as the conventional canonical ρ_0 representing the air in thermal equilibrium at some temperature T_0. In the following, expectations are over this ensemble: $\langle X \rangle = \mathrm{Tr}(\rho_0 X)$, and we examine the effect of modifying it by new information. Let $n(x,t)$ be the particle density (number of molecules per unit volume) and choose A to be the number N of particles in a small volume V_A about the point x' at time t', while B is the air pressure at a different space–time point (x,t):

$$A = n(x',t')V_A = N \tag{39}$$

$$B = P(x,t) \tag{40}$$

We make the nonessential but simplifying assumption that the region V_A is small compared to a mean free path, so that for all practical purposes the fluctuations in N are those of the ideal gas law, as used by Einstein long ago:

$$K_{AA} = \langle A^2 \rangle - \langle A \rangle^2 = \langle \delta N^2 \rangle = N_0 = n_0 V_A \tag{41}$$

where n_0 is the equilibrium particle density. The region V_A is to act as an acoustical source during a short time interval about t', in which

$$\delta Q_A = \delta n\, V_A = \text{number of particles injected.} \tag{42}$$

In conventional acoustics a source $s(x,t)$ is usually defined instead in terms of volume of fluid injected; so they are related by

$$\delta s = \delta Q_A / n_0 \,. \tag{43}$$

With these preliminaries, our general prediction rule $\delta\langle B \rangle = (K_{BA}/K_{AA})\delta Q_A$ becomes: the predicted sound pressure is

$$\begin{aligned} \delta\langle P(x,t)\rangle &= \left[\frac{[\langle P(x,t)n(x',t')\rangle - \langle P\rangle\langle n\rangle]V_A}{\langle n\rangle V_A}\right][\langle n\rangle \delta s] \\ &= [\langle P(x,t)n(x',t')\rangle - P_0 n_0]\,\delta s \end{aligned} \tag{44}$$

where $P_0 = \langle P \rangle$, $n_0 = \langle n \rangle$ are the equilibrium pressure and particle density, supposed independent of x and t.

Comparing this with the conventional acoustic Green's function solution for a prescribed source distribution $s(x',t')$:

$$P(x,t) = \int dt' \int d^3x' G(x,t;x',t')\, s(x',t') \tag{45}$$

we see that the MAXENT prediction of the acoustic Green's function is

$$G(x,t;x',t') = \langle \delta P(x,t) n(x',t')\rangle = (1/kT)\langle \delta P(x,t)\delta P(x',t')\rangle \tag{46}$$

where now we are writing (in a notation perhaps slightly inconsistent with our previous usage), $\delta P = P - P_0$, the departure from equilibrium pressure.

The evident symmetry in (46) is recognized as just the Helmholtz–Rayleigh reciprocity theorem (Rayleigh, 1877, §294). All the known reciprocity principles seem to appear automatically in

MAXENT as simple mathematical identities of the general formalism, without our ever having to make any special effort to get them. Extra physical assumptions such as time–reversal symmetry or the decay law of spontaneous fluctuations, are never needed.

In principle, we could calculate the pressure–pressure covariance function in (46) directly; but this is a complicated problem in many–body theory which would itself require a separate long article. Our point is made more quickly if we just note that we already know the Green's function G from ordinary acoustical theory. A velocity potential $\phi(x,t)$ generates the velocity and density fields through $v = \nabla\phi, \delta n = -n_0\dot{\phi}/c^2$, where c is the velocity of sound and v the mass velocity of the fluid. The point source solution of the acoustical wave equation is spherically symmetric:

$$\phi(r,t) = -\frac{\dot{s}(t - r/c)}{4\pi r}\,, \tag{47}$$

and if the source operates as a short pulse, our δQ_A is

$$\delta Q_A = n_0\,\delta s = n_0 \int \dot{s}(t')\,dt' = \int \dot{Q}_A\,dt'\,. \tag{48}$$

At this point it is easier mathematically – and also more general – to go into the frequency domain by taking time Fourier transforms of (47). Using (48) this gives

$$\phi(r,\omega) = i\omega\,\frac{\exp(i\omega t/c)}{4\pi r}\,Q_A(\omega) \tag{49}$$

Therefore, using the above relations, we predict density and pressure variations at B given by (now we write r_{AB} for the distance $|x - x'|$):

$$\delta n(x,\omega) = \frac{n_0\omega^2}{4\pi r_{AB}}\,\exp(i\omega r_{AB}/c)\,Q_A(\omega) \tag{50}$$

$$\delta P(x,\omega) = \frac{i\omega}{4\pi r_{AB}}\,\exp(i\omega r_{AB}/c)\,Q_A(\omega) \tag{51}$$

This completes our derivation of the direct propagation term, corresponding to (27) and (neglecting for the moment the renormalization of the source strength) the first term $(K_{BA}/K_{AA})\,\delta Q_A$ in (35). We now try to relate the inferential scattering indicated by the last term of (35) to Rayleigh scattering.

11. THE RAYLEIGH SCATTERING TERM

Let us introduce that third quantity C, as representing, like A, the number of particles in a small volume V_C, again supposed small compared to a mean free path (and therefore small compared to the wavelength $\lambda = 2\pi c/\omega$):

$$C = n(x'', t'')\,V_C \tag{52}$$

To impose the constraint $\delta\langle C\rangle = 0$ is, in effect, to replace the boundary of V_C by a rigid wall that allows no particles to cross it. This is just the problem that Rayleigh (1877) solved as a boundary–value problem of mathematical physics, and we now try to relate it to our statistical result (35).

Comparing the two terms in (35) enables us to define the scattering cross-section. The energy radiated from the source A is $U(\delta Q_A)^2$, where U is a factor that we could easily, but need not, calculate, because the same factor appears in the energy scattered from C:

$$\sigma\left[\frac{U\delta Q_A^2}{4\pi r_{AC}^2}\right] \tag{53}$$

which defines the cross–section σ, and they cancel out. This is related to the fictitious source strength "δQ_C" by noting that the energy density arriving at B from A is $U(\delta Q_A)^2/4\pi r_{AB}^2$, so the scattered flux at B from C is

$$\left[\frac{U(\text{“}\delta Q_C\text{”})^2}{4\pi r_{BC}^2}\right] = \left[\frac{U(K_{CA}/K_{AA})^2\delta Q_A^2}{4\pi r_{AC}^2}\right] = \sigma\left[\frac{U\delta Q_A^2}{4\pi r_{AC}^2}\right]\frac{1}{4\pi r_{BC}^2}\,. \tag{54}$$

Therefore the predicted scattering cross–section is given in terms of our covariance functions by

$$\sigma = 4\pi r_{AC}^2\left|\frac{K_{CA}}{K_{AA}}\right|^2 \tag{55}$$

But the required ratio is, from our previous equations,

$$\frac{K_{CA}}{K_{AA}} = \left[\frac{n_0\omega^2 V_C}{4\pi r_{AC}\,c^2}\right]\exp(i\omega r_{AC}/c)\,. \tag{56}$$

and so the scattering cross–section predicted by MAXENT is

$$\sigma = \frac{\omega^4\,V_C^2}{4\pi c^4} \propto \frac{V_C^2}{\lambda^4} \tag{57}$$

which is just Rayleigh's formula† down to the last factor of π, with the λ^{-4}. dependence which he used to explain the blue color of the sky in the analogous electromagnetic scattering problem.

This little test of the MAXENT relations illustrates that covariance functions in a maximum–entropy distribution have a direct physical meaning, equivalent to conventional causal propagators if the situation is one involving physical causation. But those covariance functions are far more general; they represent the best predictions we are able to make from the information we have, whether or not physical causation is involved. With more effort we could have removed our assumption about the smallness of V_C and derived more elaborate (t–matrix) scattering formulas of more general validity.

One bit of unfinished business remains: up till now we have ignored that prime on Q_A in (35). But that is hiding the most interesting part of our story.

12. MEANING OF THE RENORMALIZED SOURCE

We noted before that, intuitively, source renormalization is something like increased "stiffness" of the kind we are familiar with in mechanics or electrical network theory, where imposing a constraint on one motion or current increases the resistance to other motions or currents (by blocking paths where currents might otherwise have flowed).

But in inferential scattering this increased "stiffness" may take an unexpected form. Let us expand the renormalization factor in (36):

$$(1 - R_{AC}^2)^{-1} = 1 + R_{AC}^2 + R_{AC}^4 + \cdots \tag{58}$$

† Rayleigh (1877), §296, Eq. (13). Supposing a rigid sphere, $\Delta m/m = 1$, and integrating the square of Rayleigh's scattering amplitude over all angles, we obtain just our Eq. (57).

and substitute the result into (35). We shall need a more compact notation, so define the "propagators"

$$X_{BA} \equiv K_{BA}/K_{AA}\,. \tag{59}$$

Then our full MAXENT prediction (35) expands into

$$\delta\langle B\rangle = [X_{BA} - X_{BC}X_{CA} + X_{BA}X_{AC}X_{CA} - X_{BC}X_{CA}X_{AC}X_{CA} + \cdots]\,\delta Q_A \tag{60}$$

Each of these terms has a simple meaning. The first is just the standard regression result (25) that held before the constraint $\delta\langle C\rangle = 0$ was imposed. The second, as we have just seen, represents the Rayleigh scattering of the constraint. But that is only the first order term in the full effect of the new constraint.

A moment's contemplation of the third term will reveal its meaning: it is the amplitude of a double scattered wave that has propagated from A to C, scattered off C back to A, then scattered off A on to B. We might represent this by the double scattering process

$$(A \longrightarrow C \longrightarrow A \longrightarrow B)\,.$$

Likewise the fourth term is the effect, as seen at B, of the triple scattering process

$$(A \longrightarrow C \longrightarrow A \longrightarrow C \longrightarrow B),$$

and so on!

So what the source renormalization has done for us, in this particular case, is that it has put in every possible multiple scattering effect in addition to the direct propagation and Rayleigh terms. At first glance, it may seem surprising that arbitrarily high order scatterings are given already by what is only the *first order* of MAXENT perturbation theory. But we can understand it as follows.

This phenomenon was noted before in Heims & Jaynes (1962), where we applied MAXENT to calculation of gyromechanical and gyromagnetic effects. All terms of the famous susceptibility formula of van Vleck, which he derived by second order energy level perturbation theory, appeared in the first order of our calculation. The reason was that the expansion parameter was different in the two calculations.

Schwinger (1969, p. 36) called attention to this same phenomenon in his source theory for quantum fields. One may consider phenomena that are first order in the action function. But the action function itself may be expanded by iteration into an infinite series, in which successive terms are recognized as representing a noninteracting systsm, the primitive interactions, $e-e$ scattering, pair annihilation, and so on. Ordinarily one would consider that the experimental charge and mass of the electron are modified by its interactions with the electromagnetic field, so in principle they could be determined only after summing an infinite perturbation series. But this is not the case here; as Schwinger puts it:

> "It should be emphasized that the iterated solution is a classification of processes in terms of increasing degree of complexity. It is not a perturbation expansion. The physical electron mass m, and the physical electron charge e, which are identified originally under specific physical circumstances, will never change their significance when the class of phenomena under examination is enlarged.$\cdots$ Later terms in this series do not contain modifications of earlier ones."

Schwinger's source concept enables him to define the symbols e, m as the experimental charge and mass from the start, with great simplification of the logic and great pragmatic advantages in calculation.

Our thermal source concept enables us to do something very similar (in fact, we think that they may be seen ultimately as two different examples of the same basic theory). In a conventional physics calculation where one expands in powers of the interaction forces, n'th order multiple scattering would appear only in the $2n$'th order of the perturbation. But we, like Schwinger, are expanding in powers of the source strength, and the MAXENT formalism gives in first order the exact part of the response that is linear in the source strength, however high order it may be in the interaction forces.

REFERENCES

A. I. Akhiezer & S. V. Peletminskii (1981), *Methods of Statistical Physics*, Pergamon Press, New York.

J. W. Gibbs (1873, 1875) These works are reprinted in *The Scientific Papers of J. Willard Gibbs*, Dover Publishing Co., Inc., New York (1961).

J. W. Gibbs (1902), *Elementary Principles in Statistical Mechanics*, Reprinted by Dover Publishing Co., Inc. New York (1961).

S. Heims & E. T. Jaynes (1962), "Theory of Gyromagnetic Effects and some Related Magnetic Phenomena", Revs. Mod. Phys. **34** , pp. 143–165.

E. T. Jaynes (1963) "Information Theory and Statistical Mechanics", in *Statistical Physics*, ed K. W. Ford, Benjamin, New York, pp. 181–218.

E. T. Jaynes (1965) "Gibbs vs Boltzmann Entropies", Am. J. Phys. **33** , pp. 391–398.

E. T. Jaynes (1980) "The Minimum Entropy Production Principle", Ann. Rev. Phys. Chem. Vol. 31, pp. 579–601.

E. T. Jaynes (1982) Review of Akhiezer & Peletminskii, Physics Today, **35**, 57.

J. E. Mayer & M. G. Mayer (1940), *Statistical Mechanics*, van Nostrand, New York.

Wm. C. Mitchell (1967), "Thermal Driving", Ph.D. Thesis, Washington University, St. Louis MO.

Lord Rayleigh (1877), *Theory of Sound*, 2 vols. reprinted as one by Dover Publications, Inc., (1945).

J. Schwinger (1969), *Particles and Sources*, Gordon & Breach, New York. This work is greatly expanded in two later volumes (Schwinger, 1970, 1973), but the little 1969 book, without such a mass of detail, remains the easiest one for a beginner to learn the basic ideas.

J. Schwinger (1970), *Particles, Sources, and Fields*, Volume II (1973), Addison–Wesley Publishing Co., Reading MA.

C. P. Slichter (1978), *Principles of Magnetic Resonance*, Springer–Verlag, Berlin .

R. Zwanzig (1965), "Time–Correlation Functions and Transport Coefficients in Statistical Mechanics ", Ann. Rev. Phys. Chem., Vol 16, pp. 67–102.