carnot, 7/8/1996

THE EVOLUTION OF CARNOT'S PRINCIPLE*

E. T. Jaynes**
St. John's College
Cambridge CB2 1TP, U. K.

Abstract: We trace the development of the technical ideas showing that the Second Law of Thermodynamics became, over a Century ago, a general principle of reasoning, applicable to scientific inference in other fields than thermodynamics. Both the logic and the procedure of our present maximum entropy applications are easily recognized in the methods for predicting equilibrium conditions introduced by Gibbs in 1875. Chemical thermodynamics has been based on them ever since. What is new in this field is not the method, but the recognition of its generality.

1. INTRODUCTION 2

2. CARNOT'S PRINCIPLE 3

3. FIRST METAMORPHOSIS: KELVIN 4

4. SECOND METAMORPHOSIS: CLAUSIUS 6

5. THIRD METAMORPHOSIS: GIBBS 7

6. FOURTH METAMORPHOSIS: BOLTZMANN 8

7. CONCLUSION 11

APPENDIX A: COMMENTS ON KELVIN'S RELATION 14

APPENDIX B: ANTI–CARNOT ENGINES 15

APPENDIX C: REVERSIBILITY 16

REFERENCES 17

*The opening talk at the EMBO Workshop on Maximum–Entropy Methods in x–ray crystallographic and biological macromolecule structure determination, Orsay, France, April 24–28, 1984. Reprinted in Ericksen & Smith (1988), Vol 1, pp. 267–282

**Visiting Fellow, 1983–84. Permanent Address: Department of Physics, Washington University, St. Louis MO 63130, USA.

1. INTRODUCTION

The first reaction of nearly everybody, on hearing of a mysterious principle called "maximum entropy" with a seemingly magical power of extracting more information from incomplete data than they contain, is disbelief.

The second reaction, on sensing that there does seem to be something in it, is puzzlement. How is it possible that a quantity belonging to thermodynamics could escape from that setting and metamorphose itself into a principle of reasoning, able to resolve logical ambiguities in situations that have nothing to do with thermodynamics?

Newcomers to this field usually start by asking, not how to apply the method, or even what numerical results it gives; but "I don't see what this has to do with the entropy of thermodynamics – is there a connection?" Therefore it might be useful, before seeing details of present applications, to explain that connection.

We are taught to think of the First Law of Thermodynamics as a basic law of physics, true of necessity in every case. But attempts to see the Second Law in this way (Kelvin, Clausius, Planck, Boltzmann and many others) never quite succeeded; and Gibbs (1875) recognized that its logic is different. He concluded that "the impossibility of an uncompensated decrease seems reduced to improbability", a remark that Boltzmann quoted 20 years later in the Introduction to his *Gastheorie*.

Clausius saw the second law as a law of physics, but only a qualitative one – a kind of arrow to tell us in which general direction a process will go. Gibbs, while depriving it of that logical certainty, extended its practical application to serve the stronger purpose of quantitative prediction; to fill the logical void left by the great incompleteness of thermodynamic data. Out of all the different macroscopic behaviors permitted by the macroscopic data and the microscopic laws of physics, which should we choose as, not what *must* happen, but only what will most likely happen?

Since Gibbs' *Heterogeneous Equilibrium* (1875–78) the second law has been used in practice, not as a "law of physics", but as a principle of human inference; a criterion for resolving the ambiguities of incomplete data. In this service it does indeed extract more information than could have been obtained from the data alone; not by magic but by combining the evidence of the data with the additional information contained in the entropy function.

In other words, Gibbs' use of the second law to predict equilibrium states was virtually identical in rationale with our present maximum–entropy inference. The experimental confirmation of Gibbs' thermodynamic predictions, and the success of maximum–entropy predictions outside thermodynamics, are just two illustrations of the power of that rationale.

The above summarizes our general philosophical viewpoint; now we must justify it by examining those mysterious technical details, to show that there is not just a similarity of philosophy, but an identity of mathematical method.

But we must be prepared for the same disappointment that James Clerk Maxwell felt in 1878 when he examined that mysterious new invention called a "telephone", and found that it had no ultrafine machinery and ran on principles readily understood by every schoolboy. He reported that his disappointment at its simplicity "was only partially relieved on finding that it actually did work".

We hope that the disappointment of some at the simplicity of the maximum entropy principle, will be partially relieved on finding that it actually does work; and perhaps even fully compensated on seeing the connection with thermodynamics, and the logical unity of these seemingly different fields. This may be hoped for particularly in a lecture given in Paris; for the general viewpoint and the specific principle, from which it all follows, were given by Laplace and Carnot, not far from where we are now. After more than 150 years, we are beginning to understand them.

In the following we survey the reasoning that follows from Carnot's principle, not in the confusion of the actual historical development, but as today's hindsight shows us it could have been done. Of course, if our object were merely to explain maximum–entropy inference, that can be done by much shorter arguments without reference to thermodynamics, as will doubtless be shown by other speakers here. Indeed, we think that in the future thermodynamics itself will be approached by that shorter route and will be seen as only one particular application of maximum–entropy inference. But our present object is to make the logical connection to what is familiar to scientists today; the historical approach to thermodynamics.

Part 2 recalls the background of Carnot's work. Parts 3–6 offer a short course in "Classical Thermodynamics Made Easy". All the results will be familiar from textbooks, but the reasoning is simpler; even in the initial period when it was considered a new physical law, we can now see that Carnot's principle was actually used only to resolve ambiguities. Then in Part 7 the "new" maximum entropy methods will be seen as just one more step in the natural development of the subject.

Since entropy maximizers are sometimes accused of trying to "get something for nothing", we note that the method expresses, and has evolved from, an explicit statement of the opposite; that you *cannot* get something for nothing.

2. CARNOT'S PRINCIPLE

In the revolutionary years 1791–1797 an army engineer named Lazare Carnot found himself suddenly catapulted into the ruling bodies of France. As a member of the Legislative Assembly, the Committee of Public Safety, and the Directoire, he participated in voting for the execution of Louis XVI, the appointment of Napoleon to his Italian command; and in so many other things that his biography (Reinhard, 1950) fills two volumes.

But before and after that period, Lazare Carnot published articles (Gillispie, 1971) on his true lifelong interest, the most general statement of the principles of mechanics and inferring from them the impossibility of a perpetual motion machine. His son Sadi then turned that idea neatly around, inferring a new principle from the assumed impossibility of a perpetual motion machine. (It was Lazare's grandson, another Sadi Carnot, the nephew of our Sadi, who later became President of the French Republic).

An important technical problem of the time was the design of steam engines. How much work can be extracted from a kilo of coal? Can the efficiency be improved by different temperatures or pressures, a different working substance than water; or some different mode of operation than pistons and cylinders?

In the absence of any sound understanding of the nature of heat, misconceptions flourished. In 1818 Petit suggested that an air engine would be far more efficient than a steam engine, because no heat of vaporization need be supplied; an erroneous argument that must have inspired much wasted effort. Sadi Carnot (1824) resolved this confusion by enunciating a single, intuitively compelling principle that answered all these questions, and a hundred others not yet asked. He envisaged a generalized *heat* engine (and the qualification "heat" is essential – see Appendix B) which operates by drawing heat q_1 from a source which is at thermal equilibrium at temperature t_1, and delivering useful work W. He saw that, in order to operate continuously, the engine requires also a cold reservoir, $t_2 < t_1$, to which some heat q_2 can be discharged. The temperature scale t was any convenient thermometer calibration.

Now Carnot had the happy idea of a reversible engine; one can turn the shaft backwards, delivering the same work W back to the engine, which then delivers the same heat q_1 back to the high–temperature reservoir. One does not think about this very long before perceiving:

Carnot's Principle: No heat engine E can be more efficient than a reversible one E_r operating between the same temperatures.

For suppose that some engine E, given heat q_1, can deliver a greater amount of work $W > W_r$ than does E_r. Then we need only connect them mechanically so that E runs E_r backwards, delivering the work W_r to it, and thus pumping the heat q_1 back into the source reservoir, ready for re–use. The excess $W - W_r$ can then be used to drive our ships, locomotives, and factories. Once started, this would run forever, delivering an infinite amount of useful work without any further expenditure of fuel. We would have a new kind of perpetual motion machine.

We can imagine the economic impact that invention of such a machine would have today. As Max Planck put it, "we expect to make a most serviceable application" of any physical phenomenon that is found to deviate from Carnot's principle, or from any other principle that can be deduced from it. But an astonishing number of things can be deduced from Carnot's principle, as we shall now see.

For example, it follows at once that all reversible heat engines have the same efficiency, independent of the working substance or mode of operation; *i.e.,* the reversible efficiency is a universal function $e(t_1, t_2)$ of the two temperatures. This answered in one stroke all those questions about improvements in steam engines; already Carnot's principle had resolved a mass of ambiguities.

Carnot's reasoning is outstandingly beautiful, because it deduces so much from so little – and with such a sweeping generality that rises above all tedious details – but at the same time with such a compelling logical force. In this respect, I think that Carnot's principle ranks with Einstein's principle of relativity.

But Carnot solved the problem only implicitly; while he made it clear that one should strive to make an engine more nearly reversible, he did not find the explicit formula for the reversible efficiency that would result.

3. FIRST METAMORPHOSIS: KELVIN

Wm. Thomson (later Lord Kelvin) was collaborating with James Prescott Joule, who in a private laboratory in Manchester was doing the quantitative measurements that established the validity of the First Law. This gave him the essential fact that Carnot had lacked. Recognizing a universal "mechanical equivalent of heat" h, we can express heat $Q = hq$ and work W in the same units; then we have the relation

$$W = Q_1 - Q_2' \tag{1}$$

and deduction of the quantitative consequences of Carnot's principle can begin (the prime will be explained presently).

Consider two reversible engines connected in series; engine A receives the heat Q_1 at temperature t_1, delivers work W_A, and discharges heat Q_2'. Engine B receives the heat Q_2', delivers work W_B, and discharges heat Q_3 to a reservoir at temperature $t_3 < t_2$. Their efficiencies are

$$\begin{aligned} e(t_1, t_2) &= W_A/Q_1, \\ e(t_2, t_3) &= W_B/Q'_2 \end{aligned} \tag{2}$$

But by a simple mechanical linkage, A and B can be combined into a single reversible engine C, which receives heat Q_1 and delivers work $W_C = W_A + W_B$. So we must have also

$$e(t_1, t_3) = (W_A + W_B)/Q_1 \tag{3}$$

From (1), (2), (3) we find that Carnot's principle requires that [using the abbreviation $e_{13} \equiv e(t_1, t_3)$, etc.], the reversible efficiency must satisfy the functional equation

$$e_{13} = e_{12} + e_{23} - e_{12}\, e_{23}, \qquad t_1 > t_2 > t_3. \tag{4}$$

This is a condition of consistency, and it reminds us of a relation of probability theory. The change of variables $x = \log(1-e)$ makes the general solution obvious: the reversible efficiency must have the functional form

$$e_{12} \equiv e_r(t_1, t_2) = 1 - f(t_2)/f(t_1). \tag{5}$$

From (2a) and (3), $f(t)$ is a monotonic increasing function. By Carnot's principle the ratio $f(t_1)/f(t_2)$ must be the same function of t_1, t_2 for all reversible engines.

Now the temperature scale t was basically arbitrary (uniformly spaced marks on a mercury thermometer and on a gas thermometer do not agree), long a troublesome problem. Turning the argument around, Kelvin perceived that Carnot's principle resolves another ambiguity; if the reversible efficiency is a universal function of the temperatures, then it in effect defines a universal temperature scale that is independent of the properties of any particular substance like mercury.

We define the *Kelvin temperature scale* by

$$T(t) = Cf(t) \tag{6}$$

and now only one free choice is left to us; we may choose the arbitrary multiplicative factor C, as a convention, to indicate the size of the units in which we measure temperature.

Carnot's reversible efficiency is then

$$e_r = 1 - T_2/T_1 \tag{7}$$

from which one readily determines the measurements by which one can calibrate his thermometer to read T. All such thermometers, however constructed, and independently calibrated, will then agree in their readings at all points, if the units are chosen so that they agree at one point (at least, if a thermometer is ever found that fails to do this, then we shall have the means to realize Carnot's perpetual motion machine after all).

With this, Carnot's principle starts its metamorphosis. It now says that the efficiency e of a real heat engine must satisfy the inequality $e \leq e_r$; or from (1), (2), (7),

$$e = 1 - Q_2'/Q_1 \leq 1 - T_2/T_1. \tag{8}$$

Writing now $Q_2 = -Q_2'$ so that Q_1, Q_2 are both quantities of heat delivered from a heat reservoir to the engine, it takes the more suggestive form

$$Q_1/T_1 + Q_2/T_2 \leq 0 \tag{9}$$

with equality if and only if the engine is reversible.

A simple generalization is then obvious; we may consider a more complicated heat engine that runs cyclically, making contact successively with n reservoirs at temperatures $(T_1 \cdots T_n)$. Then the first law is $W = \Sigma Q_i$, and Carnot's principle becomes

$$\Sigma Q_i/T_i \leq 0. \tag{10}$$

This is the form that Kelvin used in his work of 1854 on the thermocouple, in which the Seebeck potential and the Peltier, Joule, and Thomson heat effects were all analyzed correctly, leading to the thermoelectric equations still in use.

Equation (10) is pivotal for all further developments, forming the starting point for two very different extensions. Some comments on its meaning and generality are given in Appendix A. In Appendix B we note some important limitations on the applicability of the reversible efficiency formula (7), and speculate about further generalizations.

4. SECOND METAMORPHOSIS: CLAUSIUS

It is curious that, having perceived such an important consequence of Carnot's principle as the temperature scale (6), Kelvin does not seem to have perceived the still more important fact that was now staring him in the face in (10). This was left for Rudolph Clausius. Imagine (10) extended to arbitrarily large n, the sum going into a cyclic integral:

$$\oint dQ/T \leq 0. \tag{11}$$

In the limit of a reversible process, where the equality applies, T is also the temperature of the system. But then (11), holding for any cycle, is the condition that the line integral over any part of a cycle is independent of the path.

Thus was discovered a new function S of the thermodynamic state of the system, defined to within an additive constant by the difference

$$S_A - S_B = \int_B^A dQ/T \tag{12}$$

where we integrate over a reversible path R; *i.e.,* a locus of equilibrium states. As we know, in 1865 Clausius coined the name "entropy" for this quantity.

The integral (12) is over a reversible path, but only part of a cycle. Then complete the cycle by adding to it a return from state A to state B over any path P, reversible or irreversible. With this choice of cycle, (11) becomes

$$\int_A^B dQ/T \leq S_B - S_A \tag{13}$$

with equality if and only if the process $A \rightarrow B$ is reversible (the meaning of the word "reversible" may now be extended beyond what Carnot had in mind, as explained in Appendix C).

Note, from its origin in (10), that in (13) T denotes the temperature of a heat bath with which the system is momentarily in contact. This may or may not be the temperature of the system. Therefore, in all cases, reversible or irreversible, the negative of the left–hand side of (13) is the entropy gained by the heat–reservoirs which constitute, for the system, the "rest of the universe". So Carnot's principle has now become: in the change from one thermal equilibrium state to another, the total entropy of all bodies involved cannot decrease; if it increases, the process is therefore irreversible:

$$S(final) \geq S(initial) \quad . \tag{14}$$

Note also that (14) describes only the net result of a process that begins and ends in thermal equilibrium. Carnot's principle does not permit us to draw any conclusion about time dependence, such as $dS/dt \geq 0$ at intermediate times. Indeed, entropy has been defined only for equilibrium states, in which there is no time dependence.

We shall take (14) as the fundamental Clausius statement of the second law. Other statements have been proposed, but (14) is logically simpler and it has stood the test of time, remaining valid in situations such as negative spin temperatures where some others failed. However, from the standpoint of logic (14) only restates Carnot's principle in a more useful form; it adds nothing to its actual content. In particular, Clausius could still see it as a law of physics.

5. THIRD METAMORPHOSIS: GIBBS

The above statements of the Second Law are the ones traditionally taught to physicists, although they have severe limitations. Equation (14) gives us one piece of information about the general direction in which an irreversible process will go; but it does not tell us how fast it will go, how far, or along what specific path. And it refers only to a closed system (no particles enter or leave).

Gibbs showed how to remove two of those limitations. He generalized the definition of entropy to open systems, as needed for many applications. More important for our purposes, he perceived the correct logical status of Carnot's principle, which enabled him to extend its application to quantitative prediction, thus answering the question: "*How far*?".

Instead of Clausius' weak statement that the total entropy of all bodies involved "tends" to increase, Gibbs made the strong prediction that it *will* increase, up to the maximum value permitted by whatever constraints (conservation of energy, volume, mole numbers, etc.) are imposed by the experimental arrangement and the known laws of physics. Furthermore, the systems for which this is predicted can be more complicated than those envisaged by Clausius; they may consist of many different chemical components, free to distribute themselves over many phases.

Gibbs' variational principle resolved the ambiguity noted above: "Given the initial macroscopic data defining a nonequilibrium state, there are millions of conceivable final equilibrium macrostates to which our system might go, all permitted by the conservation laws. Which shall we choose as the most likely to be realized?"

Although he gave a definite answer to this question, Gibbs noted that his answer was not found by deductive reasoning. Indeed, the problem had no deductive solution because it was ill–posed. There are initial microstates, allowed by the data and the laws of physics, for which the system will not go to the macrostate of maximum entropy. There may be additional constraints, unknown to us, which make it impossible for the system to go to that state; for example new "constants of the motion". So on what grounds could he justify making that choice in preference to all others?

At this point thermodynamics takes on a fundamentally new character. We have to recognize the distinction between two different kinds of reasoning; *deduction* and *inference*. Instead of asking, "What do the laws of physics require the system to do?", which cannot be answered without knowledge of the exact microstate, Gibbs asked a more modest question, which can be answered: "What is the best guess we can make, from the partial information that we have?" Of course, this implies some statement of what we mean by "best".

At first glance, this does not seem a very radical move. It must be clear to every child that virtually all human reasoning, in or out of science, is of necessity inference based on the information at hand, rather than logical deduction – which almost always would require far more information than we possess. Yet to sophisticated scientists this change of thinking has been very difficult conceptually, and it can arouse bitter controversy. To recognize "officially" that we are only doing inference rather than deduction, is very foreign to the attitude that scientists are taught.

The conventional attitude is exhibited by those who would object to Gibbs' answer on the grounds that there may be unknown constraints that prevent the system from getting to the state of maximum entropy; and so Gibbs' answer might be wrong. But the same kind of objection would apply whatever answer he gave. If such an objection were sustained, Gibbs would be prohibited from giving any answer at all. Science does not advance on that kind of timidity; let us note how much more realistic and constructive is the opposite attitude. To one who raised that objection, Gibbs might reply as follows:

"Of course, my answer might be wrong. You seem to think that would be a calamity that we must avoid; but you are like a chess player who thinks only one move ahead. If you will think ahead two moves, you will see that, on the contrary, getting a wrong answer would be even more valuable than getting a right one. As you note, at present we do not know whether there may exist

unknown constraints that would prevent the system from getting to the maximum entropy state. But I choose to ignore that warning, go ahead with my calculation, and then ask an experimentalist to compare my prediction with observation. What conclusions will we be able to draw from his verdict?

"Suppose my prediction turns out to be right. That does not prove that no unknown constraints exist; but it does prove that there are none which prevent the system from getting to the macrostate of maximum entropy. So the calculation has served a useful predictive purpose, and its success gives us more confidence in future predictions.

"But suppose my prediction turns out to be wrong; the experiment repeatedly gives a different result. Then we have learned far more; we know that there *is* some new (*i.e.*, previously unknown) constraint affecting the macroscopic behavior, and the nature of the error gives us a clue as to what that new constraint is. We would have a start toward learning a fundamental new physical fact, perhaps a new law of physics. I do not see this as a calamity; how else can we advance to a new state of knowledge about physical law, but by having the courage to go ahead with the best inferences we can make from our present state of knowledge?"

The words we have just put into Gibbs' mouth are not fanciful. Gibbs' classical statistical mechanics did make incorrect predictions of specific heats and vapor pressures. These were the first clues, indicating the new constraints of discrete energy levels, pointing to the quantum theory. Nobody would have realized that such things were relevant to the question, had Gibbs lacked the courage to make an inference because he might be wrong.

After development of the Schrödinger equation, the Gibbs formalism based on maximizing the new quantum expression for entropy has yielded so many thousands of quantitatively correct equilibrium predictions that there seems to be almost no chance that it will ever fail in that problem. Whenever it did seem to fail – as in the case of ortho and para hydrogen – it was seen quickly that it was only performing its second function, revealing an unexpected constraint.

Today, we are only in the initial stages of extensions to predict the details of nonequilibrium behavior; these put our entropy expressions to a more severe test. We can by no means rule out the possibility that nonequilibrium statistical mechanics may lead to incorrect predictions, which would then point the way to the next higher level of understanding of physical law, beyond our present quantum theory. We may be seeing the incipient beginnings of this in the lore of "strange attractors".

We think that this scenario will be repeated many times in the future outside thermodynamics, particularly as the method moves into biology. Most maximum entropy inferences will be correct, serving a useful predictive purpose. But some of the predictions will be wrong; those instances, far from being calamities, will open the doors to new basic knowledge.

Another of the curiosities of this field is that, having done so much with entropy and demonstrated such a deep understanding of the logic underlying the second law, giving thermodynamics an entirely different character, Gibbs said almost nothing about what entropy really means. He showed, far more than anyone else, how much we can accomplish by maximizing entropy. Yet we cannot learn from Gibbs: "*What are we actually doing when we maximize entropy*?" For this we must turn to Boltzmann.

6. FOURTH METAMORPHOSIS: BOLTZMANN

Entropy first appeared, unanticipated and without warning, merely as a mathematical construct in equation (12). Even after its fundamental nature and usefulness were recognized and exploited, the question: "*What is it*?" continued to mystify and confuse. It appears that the answer was first revealed to Ludwig Boltzmann, when he calculated the phase volume of an ideal gas of N atoms

in volume V, for which the energy lies in $(E, E+dE)$:

$$W = \int_R d^3x_1 \cdots d^3x_N \, d^3p_1 \cdots d^3p_N = C\, V^N \, E^{3N/2-1}\, dE \tag{15}$$

where the region R of integration is those points for which all coordinates are within a volume V, and the momenta satisfy

$$E < \Sigma p^2/2m < E + dE. \tag{16}$$

The constant C is independent of V and E.

Now from elementary thermodynamics it was known that the entropy of any system which obeys the equation of state $PV = RT$ with a heat capacity $C_v = const.$, has the form $S(V,T) = C_v \log T + R \log V + const.$, where the *const.* is independent of T and V. But with the heat capacity for Boltzmann's monoatomic gas, $C_v = (3/2)R$ and the resulting internal energy function $E = (3/2)RT$, it was evident that $\log W$ has the same volume and energy dependence as the entropy of that gas, calculated from (12). That is, to within an additive constant independent of T and V, it was true that

$$S = k \log W. \tag{17}$$

This is such a strikingly simple relation that one can hardly avoid jumping to the conclusion that it must be true in general; *i.e.,* the entropy of any macroscopic thermodynamic state A is a measure of the phase volume W_A occupied by all microstates compatible with A.

It is convenient verbally to say that S measures the "number of ways" in which the macrostate A can be realized. This is justified in quantum theory, where we learn that a classical phase volume W does correspond to a number of global quantum states $n = W/h^{3N}$. So if we agree, as a convention, that we shall measure classical phase volume in units of h^{3N}, then this manner of speaking will be appropriate in either classical or quantum theory.

We feel quickly that the conjectured generalization of (17) must be correct, because of the light that this throws on our problem. Suddenly, the mysteries evaporate; the meaning of Carnot's principle, the reason for the second law, and the justification for Gibbs' variational principle, all become obvious. Let us survey quickly the many things that we can learn from this remarkable discovery.

Given a "choice" between going into two macrostates A and B, if $S_A < S_B$, a system will appear to show an overwhelmingly strong preference for B, not because it prefers any particular microstate in B, but only because there are so many more of them. As noted in Appendix C, an entropy difference $(S_B - S_A)$ corresponding to one microcalorie at room temperature indicates a ratio $W_B/W_A > \exp(10^{15})$. Thus violations are so improbable that Carnot's principle, or the equivalent Clausius statement (14), appear in the laboratory as absolutely rigid "stone wall" constraints suggesting a law of physics rather than a matter of probability.

Let us see the light that this casts on Gibbs' method, by examining a simple application. We have two systems of one degree of freedom (*i.e.,* their energy and temperature can vary when in contact with other systems). Then their entropy functions are

$$S_1(E_1) = k \log W_1(E_1), \qquad S_2(E_2) = k \log W_2(E_2), \tag{18}$$

The systems start out in thermal equilibrium with arbitrary initial energies E_{1i}, E_{2i}. Then they are placed in contact so they can exchange energy in such a way that the total amount is conserved:

$$E = E_1 + E_2 = const., \qquad E_1 > 0, \quad E_2 > 0. \tag{19}$$

Required: to predict the final energies E_{1f}, E_{2f} that they will reach when they come into equilibrium with each other.

This is manifestly an ill–posed problem; for the final energies must depend on the initial microstates which are unknown; and all values compatible with (19) are possible without violating any known laws of physics. We are thus obliged to use inference rather than deduction. Gibbs' algorithm was: predict that energy distribution that maximizes the total entropy $S_1 + S_2$ subject to the constraint (19). At first this seems arbitrary; but now if (17) is correct we can see why this guess is "best". We are maximizing the product

$$M(E_1) = W_1(E_1)\, W_2(E - E_1) \tag{20}$$

with respect to E_1; but that product is just the multiplicity, or number of ways in which the energy distribution (E_1, E_2) can be realized. So in the light of (17) Gibbs' rule now says, merely: "Predict that energy distribution that can happen in the greatest number of ways, subject to the information you have". An eminently sensible criterion!

Experimentally, one says that equilibrium is reached when the systems have equal temperature. Differentiating (20), we find that the maximum is reached when $d\log W_1/dE_1 = d\log W_2/dE_2$. But the general thermodynamic relation $T^{-1} = dS/dE$ that follows from (12) becomes, in the light of (17)

$$\frac{1}{kT} = \frac{d\log W}{dE}\ . \tag{21}$$

So the general interpretation of entropy by (17) not only predicts equal temperature as the condition for equilibrium; it gives a simple explanation of why this is true.

The above explains why Gibbs' method gives, in a sense, the best guess one could have made in view of our great ignorance as to the microstate; but does not explain why it is so uniformly successful. If the multiplicity (20) had a broad maximum, or many local maxima, one would not expect Gibbs' rule to be very reliable in practice. This raises the question: How sharp is the maximum in the multiplicity (20)? Note that differentiating (21) once more gives the heat capacity:

$$\frac{d^2\log W}{dE^2} = -\frac{1}{kT^2C_v}. \tag{22}$$

But, as (15) shows for an ideal gas and is true in general, C_v may be interpreted as $C_v = nk/2$, where n is the effective number of degrees of freedom of the system (in quantum theory, the number excited at the temperature T), of the order of Avogadro's number for a macroscopic system. Therefore, expanding $\log M(E_1)$ about its peak at E' we have

$$M(E_1) = M(E')\,\exp\left\{-\frac{(E_1 - E')^2}{2\sigma^2}\right\} \tag{23}$$

with the RMS deviation

$$\sigma = kT\left[\frac{n_1 n_2}{(n_1 + n_2)}\right]^{1/2}. \tag{24}$$

which is of the order of $kTn^{1/2} = E'/n^{1/2} = 10^{-12}E'$. Therefore, not only is E' the value of E_1 that can happen in the greatest number of ways for given total energy E; the vast majority of all *possible* microstates with total energy E have E_1 very close to E'. Less than 1 in 10^8 of all possible states have E_1 outside the interval $(E' \pm 6\sigma)$, far too narrow to measure experimentally. From (17), then, we understand also why Gibbs' method succeeds.

But there is still more to be learned from Boltzmann's discovery (17). Imagine n_2 to become very large; then we may expand using (21):

$$\log W_2(E - E_1) = \log W_2(E) - \frac{E_1}{kT} + \cdots \tag{25}$$

and from (22) the next term is negligible. But then the fraction of the multiplicity (23) in the interval $(E_1, E_1 + dE_1)$ becomes

$$f(E_1)dE_1 = Z^{-1}\, W_1(E_1)\, \exp(-E_1/kT)\, dE_1 \tag{26}$$

which is the distribution of Gibbs' "Canonical Ensemble", the basis of his later work on Statistical Mechanics. The normalization constant

$$Z(\beta) = \int W_1(E) \exp^{-\beta E_1}\, dE \qquad , \qquad \beta = 1/kT \tag{27}$$

is Gibbs' partition function, and if we refine the inference procedure by taking as our prediction the mean value over the distribution (26) instead of the peak E', our prediction reduces to

$$\langle E_1 \rangle = -\frac{d \log Z}{d\beta}, \tag{28}$$

the basic predictive rule of statistical mechanics.

All these relations generalize effortlessly to systems with more macroscopic degrees of freedom (volume, magnetization, angular momentum, mole numbers, *etc.*) corresponding to Gibbs' grand canonical ensemble and its generalizations. So the interpretation (17) of entropy has given us the key to essentially everything that has happened since in the field of equilibrium thermodynamics and statistical mechanics. This was recognized, and exploited in their own fundamental research, by both Planck and Einstein.

In our opinion, (17) represents by far the greatest of all of Boltzmann's achievements, just because of its fundamental, timeless character. One hundred years from now, his transport equation will be a nearly forgotten detail of the history of science; but a thousand years from now, the relation $S = k \log W$ will still be the foundation stone of this subject. A more appropriate inscription for his gravestone can hardly be imagined.

7. CONCLUSION

We have followed the evolution of Carnot's principle, via Kelvin's perception that it defines a universal temperature scale, Clausius' discovery that it implied the existence of the entropy function, Gibbs' perception of its logical status, and Boltzmann's interpretation of entropy in terms of phase volume, into the general formalism of statistical mechanics. But now, we can see how utterly simple it all is, and that the reasoning had nothing fundamentally to do with thermodynamics.

From our present vantage point, everything we have done could have been found as a trivial consequence of (20). But, as a principle of reasoning, that had been given in far greater generality by James Bernoulli (1713) and Laplace (1812). Had their works been better understood and applied, we might have passed directly from first principles of inference to the canonical ensemble, with the second law as a straightforward predicted consequence rather than a puzzling empirical fact. The genius of Carnot was to have seen the one case where that fact is not puzzling, but intuitively compelling.

In our comments on Gibbs' work we have noted two fundamentally different attitudes, epitomized by the words *deduction* and *inference*. But the inference attitude that Gibbs introduced

into thermodynamics is just what Bernoulli had recognized and expounded, much earlier. With an accuracy and honesty almost unique in probability theory, Bernoulli called his work, simply, *Ars Conjectandi,* the "Art of Conjecture". Seeing correctly that a major intellectual problem of both science and everyday life is the necessity of reasoning somehow from incomplete information, he asked whether there are any general principles, of consistency and rationality, which would help us in this.

A single desideratum of consistency; that propositions about which we are in the same state of ignorance should be given equal weight in our reasoning, implies all else. This desideratum led Bernoulli to write down the general equations of probability theory.

In the hands of Bernoulli those equations were seen, not merely as rules for calculating frequencies; but the consistent rules for conducting inference; a probability distribution is used as a means of describing our state of incomplete knowledge, and the equations show how probabilities of different propositions must be related for consistency. Laplace adopted this viewpoint and applied it with great success to a mass of problems of scientific inference. In the course of this, he developed the analytical theory relevant to applications to a level that is not often surpassed today. The history is recounted in more detail in Jaynes (1983).

Now by Bernoulli's principle the multiplicity function $M(E_1)$ in (20) expresses just the relative probabilities that we should assign to different values of E_1 to represent a state of knowledge about E_1. "*Whose* state of knowledge?" everyone asks. *Answer*: a person whose relevant information consists of the phase volume functions W_1, W_2, the total energy E; and nothing else. These equations represent the "best" inferences that can be made by a person in that state of knowledge. Aware of the microstates and their relative numbers but, having no grounds for preferring any particular microstate consistent with our knowledge to any other, an honest description of what we know requires us to assign equal probabilities to them, resulting in (20).

A person with greater knowledge would have a smaller set of possible states, and would be able to make better predictions of some things. But he would seldom do better in prediction of reproducible phenomena, because those are the same for virtually all microstates in an enormously large class C; and therefore also in virtually any subset of C. Indeed, as Gibbs showed, in almost every case the knowledge supposed above is already sufficient to predict equilibrium states correctly. Still greater knowledge (such as, perhaps, that the real system stays in some complicated fractal subset of C) might be very interesting and important for future purposes; but it would not have helped for the predictions that Gibbs was making.

Knowledge of the "data" E alone would not enable us to choose among the different values of E_1 allowed by (19); the additional information contained in the entropy functions, nevertheless leads us to make one definite choice as far more likely than any other, *on the information supposed.*

On the other hand, we need not know the entropy functions in advance; nothing prohibits us from trying out some guess, as a working hypothesis, to see what predictions it would make. Good predictions give us confidence that our guess was right; bad predictions can, from the particular way they fail, give us clues to a better guess. We expect that many applications will take this form.

The sharpness of the thermodynamic predictions, and the resulting "stone wall" character of the second law, arise from the extremely high dimensionality of the space of microstates, with the result that $W(E)$ is an enormously rapidly increasing function, of the order of E^N, where N is Avogadro's number.

The same reasoning will apply to many situations outside thermodynamics, in which we can enumerate a set C of conceivable situations (corresponding to the global quantum states in thermodynamics), and have some data that, although incomplete, restricts them to some subset of C. In thermodynamics we are generally trying to predict only a few quantities, seldom more than three

or four; but in image reconstruction it is not unheard of to estimate over a million pixel intensities. This is not a difference of principle, although it is a major difference to a computer programmer.

Our numerical values will not in general be as extravagant as in thermodynamics, but may still be enormous by ordinary standards. The preferred choice may be indicated over others by a factor of only a million, instead of the fantastic numbers of statistical mechanics. But the general conclusions and usefulness of the reasoning will be the same; there will still appear to be a "second law" favoring situations of high entropy, because they can be realized in more ways.

In the above we have tried to show the basic unity of thermodynamics and the "new" maximum entropy methods, by expounding a general philosophy and rationale that pertains equally to both. But it may not yet be clear where the mathematical connection lies. The mathematical appearance of our solution depends on how we choose the basic "hypothesis space" that corresponds to enumerating the global quantum states of thermodynamics.

As a highly oversimplified example, if in thermodynamics our system consisted of N molecules with r possible energy levels each, there would be r^N. possible global states, each one defined by specifying the state of each molecule. If we wanted to predict the number n_k of molecules in the k'th energy level, subject to some incomplete data D that partially restricts the possibilities, we would want to maximize the number of ways a certain set $\{n_1 \cdots n_r\}$ of occupation numbers could be realized:

$$W(n_1 \cdots n_r) \tag{29}$$

But when these numbers become large we may use the Stirling approximation, and the entropy corresponding to W is given by

$$(1/N)\log W \rightarrow -\Sigma f_i \log f_i \tag{30}$$

where $f_i = n_i/N$ is the fraction of molecules in the i'th energy level. Maximization of the expression (30) gives the familiar mathematical form of the principle [the right–hand side of (30) is often called the "Shannon entropy" nowadays, although of course it was given by Boltzmann long before Shannon was born].

Now there are many other applications where enumeration of the possibilities leads us to write the same combinatorial factor (29); and therefore the solution will proceed in the same mathematical way. For example, we may think of an image created by strewing N little "elements of luminance" over r pixels, the k'th one receiving $n_k = Nf(k)$ of them. The set of proportions $\{f(k)\}$ constitutes the "true" image, and to estimate the $f(k)$ from incomplete data we would maximize (30) subject to the constraints of the data. This would give us the reconstructed image that is "most likely" on the supposed information; *i.e.,* it is the one that can be realized in more ways than any other that agrees with the data. The situation is logically equivalent to that of thermodynamics.

But suppose we are dubious as to whether our image has been formed by anything like this strewing process. Is there now any reason for using the same maximum entropy algorithm? It appears that there may be several.

One is the argument that we put, rather presumptuously, into Gibbs' mouth above: if there is anything seriously wrong with this hypothesis space, then reconstructions based on it ought to show systematic deviations from the true scene. If such deviations are found, then they give us a clue to a better hypothesis space; if not, there is no need for a different one. Therefore, if those doubts are not specific enough to suggest a definite alternative hypothesis space, we have nothing to lose and something to gain by continuing to use the above one.

But the maximum–entropy procedure may be supported by more than one rationale. There are other arguments, based on logical consistency or on information theory, rather than multiplicity, which also point uniquely to the maximum entropy rule.

Finally, we may invoke entropy not as an element in a probabilistic argument, but as a value judgment; out of all reconstructions that agree with the data, the maximum entropy one is preferred, simply because it is cleaner, free of artifacts, and therefore more informative and safer to use. This comes about because entropy is not only a measure of multiplicity; the expression (30) is also a measure of smoothness.

Any change in the direction of equalizing any two elements f_i and f_j, increases the entropy (30). When we maximize (30), f_i and f_j will therefore be equalized unless this is prevented by some constraints coming from the data; the variational principle that generates it guarantees that the maximum entropy reconstruction cannot show any detail for which there is no evidence in the data. To the best of our knowledge, no other algorithm that has been proposed for this problem has this safety feature.

Indeed a procedure is something that exists in its own right, and is not tied down to any particular theoretical rationale. In the last analysis, the methods of science are determined by their pragmatic success; a procedure that yields useful results will be used even if it has no theoretical basis at all. If it does not yield useful results, it will not be used however strong the theoretical arguments for it. So, now let us hear from others how the method is working on some real, important, and highly nontrivial problems.

APPENDIX A. COMMENTS ON KELVIN'S RELATION

Some have thought it inelegant to base a scientific theory on such vulgar things as heat engines; but in Kelvin's form (10) of Carnot's principle:

$$\Sigma Q_i/T_i \leq 0 \qquad (A1)$$

the vulgarity and inelegance are gone. (A1) refers to any cyclic process involving coupling a system of interest to n heat reservoirs, whether or not any work is actually done. The engines are only, so to speak, held in abeyance. We infer a number of general inequalities which make no reference to any engine; but if ever we find a physical phenomenon that violates one of these inequalities, then some marvelous engines can be built.

Equation (A1) is more general in another respect. With engines in mind we have been making the tacit assumption that contacts with the different heat reservoirs are to be made successively in time. But (A1) makes no explicit reference to time either, and a moment's thought persuades us that this assumption was not necessary for our argument. Kelvin applied (A1) to the continuous flow of electric current around his thermocouple, which was in continuous contact with the heat reservoirs.

In using (A1), Kelvin launched another new field by introducing a new assumption; that we may apply Carnot's principle in the form (A1) with the equality sign, to the reversible Peltier and Thomson heat effects, even though irreversible heat conduction is also present.

But a heat current presumably "drags" some electric current with it; i.e. conduction electrons ought to be carried along a little by a stream of phonons, like sand grains carried downstream by the current of a river. Conversely, a stream of moving electrons ought to drag along some phonons, generating a heat current. The reversible effects cannot be physically independent of the irreversible ones; and this should have introduced an error into Kelvin's analysis.

The fact that Kelvin's equations were nevertheless verified means that these two errors must have cancelled each other. This was recognized later as implying a proportionality relation between the two dragging coefficients; the first example of an "Onsager Reciprocal Relation". Thus both the Clausius–Gibbs entropy developments in equilibrium theory, and Onsager's irreversible thermodynamics (1931), had their origins in Kelvin's use of (A1).

APPENDIX B: ANTI–CARNOT ENGINES

It is important that we understand some limitations on the applicability of the reversible efficiency formula (7):

$$e_r = 1 - T_2/T_1. \qquad (B1)$$

This applies, not to every type of energy converter but only to *heat* engines – *i.e.,* engines which operate by extracting heat from one reservoir which is at thermal equilibrium at some temperature T_1 and delivering heat to a similar reservoir at a lower temperature T_2. But there is no reason why (B1) should apply to engines that deliver work by a different mode of operation.

Indeed, the world's most universally available source of work – the animal muscle – presents us with a seemingly flagrant violation of that formula. Our muscles deliver useful work when there is no cold reservoir at hand (on a hot day the ambient temperature is at or above body temperature) and a naive application of (B1) would lead us to predict zero, or even negative, efficiency. But according to Alberts *et al.* (1983), under favorable conditions the observed efficiency of a muscle can be as high as 70%, although a Carnot engine would require an upper temperature of about $(273 + 37)/.30 = 1000$K to achieve this.

The answer, of course, is that a muscle is not a heat engine. It draws its energy, not from any heat reservoir, but from the activated molecules produced by a chemical reaction. This is why we stressed the word "heat" when we introduced Carnot's principle.

Only when we first allow that activation energy to degrade itself entirely into heat – and then extract only that heat for our engine – does Kelvin's formula apply. If we can learn how to capture the activation energy before it has a chance to degrade, as our muscles have already learned how to do, then we shall be able to achieve higher efficiency than (B1) in an engine. Such an anti–Carnot engine will, of course, not violate the second law. Rather, to achieve it will require a very clear understanding of what the second law really says.

What efficiency might one hope for in such an engine? There is no reason to doubt that, with proper understanding of these matters, the performance of our muscles could be at least equalled in vitro. Now, whatever the theoretical maximum efficiency, it can always be written in the Kelvin form (B1) if we wish to do so; the question then becomes: "What are the effective upper and lower temperatures?"

As a partial answer we imagine that our engine will, like our muscles, eventually discharge some heat to the outside world; then let us take T_2 as the ambient temperature; for our muscles, body temperature. What is the effective upper temperature? It appears to us that this was answered already by Gibbs; it is the highest temperature to which the activated molecules could deliver heat, which would make the maximum efficiency close to 100%.

To see this more specifically, note that at body temperature, $273 + 37 = 310$K, the average thermal energy $(kT/2)$ per degree of freedom is about 0.013 ev. It appears that the source of energy for muscle contraction is hydrolysis of adenosine triphosphate (ATP), which has a reaction energy $\Delta H = 9.9$ kcal/mole $= 0.43$ ev/molecule. The reaction might leave a product molecule in an excited state with that activation energy. If this is concentrated in only $N = 2$ degrees of freedom (one vibration mode), it thus represents a tiny "hot spot" with an effective temperature $T_{eff} = 2E_a/N_k$, of the order of 16 times body temperature.

If we can convert that little bubble of concentrated energy into useful work before it has a chance to thermalize by spreading out over many vibrational degrees of freedom, we should in principle be able to achieve a conversion efficiency approaching 15/16. Thus there is no reason to think that muscle performance is in violation of the second law.

APPENDIX C: REVERSIBILITY

Thermodynamics is notoriously a field which encourages confusion and illogic by a terminology which may use a common technical term with several different meanings, and fails to distinguish between them. We have noted before (Jaynes, 1980) some of the many different, mutually inconsistent meanings that have been attached to the word "entropy". An equally serious confusion arises from the fact that the word "reversible" is used with different meanings; and few writers since Gibbs and Planck have taken sufficient note of this.

Let A and B stand for two different macrostates, defined by specifying (*i.e.,* controlling or observing) a few macroscopic quantities like temperature, volume, pressure, magnetization, such that the change $A \to B$ can be carried out in the laboratory. What do we mean by saying that it is reversible? In the literature, we find three different meanings:

> **(1) Mechanical Reversibility.** Reversing all molecular velocities in B, the equations of motion carry the system back along exactly its previous path to A. In the end this would restore each individual molecule to its original position.

But this is manifestly not what Carnot had in mind. In his reversible engines he is considering instead:

> **(2) Carnot Reversibility.** The macroscopic physical process can be made to proceed in the opposite direction $B \to A$, restoring the original macrostate.

This is an enormously weaker condition than mechanical reversibility. But it was noted by Clausius, Gibbs, and Planck that thermodynamic reversibility is a still weaker condition (one that Carnot could have used without invalidating his argument):

> **(3) Thermodynamic Reversibility.** Even if the backward process $B \to A$ cannot be made to take place reversibly (for example, because of supercooling at a phase transition), if by any means such as $B \to C \to D \to A$ the original macrostate can be recovered without external change, then all entropies are unchanged and the process $A \to B$ is thermodynamically reversible.

From this we see that the common phrase " – the paradox of how to reconcile the irreversibility of the second law with the reversibility of the equations of motion – " does not define any real problem at all; it is a nonsense utterance, using the term "reversible" in the totally different meanings (1) and (3) in the same sentence.

These observations are hardly new. The distinction between mechanical and thermodynamic reversibility was stressed by Gibbs (1875) in his discussion of gas diffusion. Confusion of thermodynamic reversibility with Carnot reversibility was called by Planck (1949), " – an error against which I have fought untiringly all my life, but which seems impossible to eradicate."

Despite the efforts of Gibbs and Planck, these distinctions have been nearly lost today. We have found no recognition of them in current thermodynamics textbooks, or in the current literature of statistical mechanics. Indeed, both Maxwell and Gibbs are still under attack in the current literature, from authors who still do not comprehend their messages. See, for example, Atkins (1986). In our opinion, recent efforts to "explain irreversibility" by tampering with the equations of motion or the definition of entropy, address themselves to a non–problem, for reasons that Gibbs explained cogently over 100 years ago.

It should not have required the labors of Carnot, Kelvin, and Clausius to convince us that one cannot reverse all molecular velocities with the technology (pistons, stoves, magnets, etc.) available to experimenters. Thermodynamics is concerned with macrostates, for the pragmatic reason that

those are the things the experimenter can control and observe. And thermodynamics, like (we hope) all of physics, is concerned with reproducible phenomena.

By experimental means of macroscopic coarseness one can generate a class of initial states from which a macroscopic process $A \to B$ takes place reproducibly; but in general the reversed process $B \to A$ cannot be achieved reproducibly by macroscopic means. That the microscopic equations of motion may nevertheless be "reversible" in the mechanical sense, is quite irrelevant to what the experimenter can actually do.

If the experimenter's apparatus is able to put his system only in some uncontrolled point in W_A, then because of Liouville's theorem (conservation of phase volume) the process $A \to B$ cannot be reproducible unless $W_B \geq W_A$, or $S_B \geq S_A$. If the inequality holds, then the reverse process is, as Gibbs noted, not impossible, but only improbable; *i.e.,* not reproducible. The probability of success is something like $p = W_A/W_B = \exp(-(S_B - S_A)/k)$. If the entropy difference corresponds to one microcalorie at room temperature, $p < \exp(-10^{15})$. We do not see why any more than this is needed to understand and explain the observed phenomenological facts associated with the second law.

REFERENCES

B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts & J. D. Watson, *Molecular Biology of the Cell*, Garland Publishing Co., New York; pp. 550–609 (1983).

P. W. Atkins (1986), "Entropy in Relation to Complete Knowledge", Contemp. Phys. **27**, 257.

Sadi Carnot, *Reflexions sur la puissance motrice du feu*, Bachelier, Paris, (1824).

G. J. Erickson & C. R. Smith, editors, (1988), *Maximum–Entropy and Bayesian Methods in Science and Engineering*, 2 vols, Kluwer Academic Publishers, Dordrecht–Holland.

J. Willard Gibbs, "On the Equilibrium of Heterogeneous Substances", Trans. Conn. Acad. Sci (1875–78). Reprinted in *The Scientific Papers of J. Willard Gibbs*, Vol. 1; Dover Publications, Inc., N. Y. (1961).

C. C. Gillispie, *Lazare Carnot, Savant* Princeton University Press (1971). A technical analysis of his work, and its relation to that of his son Sadi. Original manuscripts.

E. T. Jaynes, "The Minimum Entropy Production Principle", in Annual Review of Physical Chemistry, Vol. 31, 579–601 (1980). Reprinted in E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, R. Rosenkrantz, Ed., D. Reidel Publishing Co., Dordrecht–Holland (1983)

A. L. Lehninger, *Bioenergetics*, W. A. Benjamin, N. Y. (1965)

A. L. Lehninger, *Biochemistry, the Molecular Basis of Cell Structure and Function*, Worth Publishers, Inc., 444 Park Ave. South, New York, N. Y. (1975).

L. Onsager, Phys. Rev. **37**, 405; **38**, 2265 (1931)

M. Planck, *Scientific Autobiography*, Philosophical Library, N. Y. (1949); pp.17–18.

M. Reinhard, *Le Grand Carnot*, 2 vols., Paris, 1950–52.