

Greek Letters

α	= wave number
β	= amplification factor
δ	= vapor film thickness
λ	= heat of vaporization
λ'	= $\lambda + 0.5 c_p \Delta T$
μ	= dynamic viscosity
ρ	= density
σ	= surface tension
ω	= angular frequency

Subscripts

LP	= Leidenfrost point (q minimum)
L	= saturated liquid
v	= saturated vapor
max	= maximum

LITERATURE CITED

1. Bromley, L. A., *Chem. Eng. Progr.*, **46**, 221 (1950).
2. Frederking, T. H. K., and J. A. Clark, *Advan. Cryog. Eng.*, **8**, 501 (1963).
3. Hsu, Y. Y., and J. W. Westwater, *A.I.Ch.E. J.*, **4**, 58 (1958).
4. Berenson, P. J., *J. Heat Transfer*, **83C**, 351 (1961).
5. Hosler, E. R., and J. W. Westwater, *A.R.S. J.*, **32**, 553 (1962).
6. Patel, B. M., and K. J. Bell, paper presented at A.I.Ch.E. Los Angeles meeting (August, 1965).
7. Lyon, D. N., *Intern. Advan. Cryog. Eng.*, **10**, 371 (1965).
8. Lamb, H., "Hydrodynamics," p. 455, Dover, New York (1945).
9. Chang, Y. P., *J. Heat Transfer*, **81C**, 1 (1959).
10. Zuber, N., *Trans. Am. Soc. Mech. Engrs.*, **80**, (1958); Ph.D. thesis, Univ. California, Los Angeles (1959).
11. Kistemaker, T., *Physica*, **29**, 351 (1963).
12. Challis, L. J., *Proc. VII, Intern. Conf. Low Temp. Physics*, Toronto, 476 (1960).
13. Ruzicka, J., "Heat Transfer to Boiling Nitrogen", "Problems of Low-Temperature Physics and Thermodynamics," p. 323, Pergamon Press, New York (1959).
14. Merte, H., Jr., and J. A. Clark, *Am. Soc. Mech. Engrs. Paper No. 63-HT-28*.
15. Frederking, T. H. K., *J. Appl. Math. Phys.*, **14**, 207 (1963).
16. Marshall, L., *Rev. Sci. Instruments*, **26**, 614 (1955).
17. Karagounis, A., *Suppl. Bull. Inst. Intern. Froid*, Annex 2 (1956).
18. Madsen, R. A., Ph.D. thesis, p. 78, Purdue Univ., Lafayette, Ind. (1965).
19. Frederking, T. H. K., *Advan. Cryog. Eng.*, **9**, 71 (1964).

Manuscript received March 9, 1965; revision received September 28, 1965; paper accepted October 1, 1965. Paper presented at A.I.Ch.E. San Francisco meeting.

Why Thermodynamics Is a Logical Consequence of Information Theory

MYRON TRIBUS, PAUL T. SHANNON, and ROBERT B. EVANS

Dartmouth College, Hanover, New Hampshire

An observation of the properties of bulk matter necessarily omits many details of the motion of the particles which comprise that matter. If a communication regarding the observation is to be consistent, it must conform to certain rules which follow from the theory of information. An analysis of the problem of deciding upon a consistent encoding for such observations leads, as a unique result, to the concepts and equations of classical thermodynamics. The information-theory analysis of this problem also leads to a better understanding of the basis for the "laws" of thermodynamics.

Consider an observer who wishes to report as faithfully as he can his observations of his environment. If he has any experience at all, he is aware that his observations are seldom accurate or complete. At the very minimum he realizes that his knowledge is tentative, that is, not so

much subject to correction (as when one misinterprets a 7 for a 1), but rather subject to reinterpretation (as when it is found that the 1 is part of a binary, not a decimal system). The observer sends his observations to a receiver and, if the messages sent to the receiver are to carry the

intended meanings, there must be agreement between the observer (sender) and the receiver as to a common code.

It is our intention in this article to demonstrate how the agreed upon code serves to both limit and form the theoretical structure which the observer and receiver may construct to help them to organize their knowledge. As a special illustration, we shall show how classical thermodynamics evolves from such constraints. This way of looking at thermodynamics has become known as the "information theory approach."

AN ACCEPTABLE CODE

All the messages sent by the observer, insofar as they pertain to scientific observations, may be represented in a two-part coded system. The first part may be called the *substantive* and the second the *interpretive* part of the message.

The substantive part is an unambiguous *statement*, in English (if that is the observer's native language), in the form of a proposition. The other part of the message is an interpretation of the truth value or credibility of the substantive part of the message. It is the observer's way of interpreting the subject matter without pretending to be omniscient about the subject. The difference between these two aspects is exemplified by the difference between Mr. A, who is uncertain about knowing at which airport his plane will land, and Mr. B, who is uncertain about everything. To be uncertain about everything represents a lesser state of knowledge than to be uncertain *about* something definite. The substantive part of the message specifies something about which to be uncertain.

Both the sender and receiver should be aware that the phrase "believing is seeing" does, in fact, apply to scientific reporting. There is no such thing as an "immaculate perception." The features of reality an observer chooses to notice and report are strongly conditioned by the scientific training received *before* the observation. There is more to seeing than meets the eye. Tycho Brahe and Einstein did not see the same thing as they watched the sun set. (Our trained observer notices an automobile collision in terms of momenta; his wife notices the noise.)

A simple symbolic method of communication for our observer-receiver pair is a notation such as $p(A|X)$, where A represents a proposition, X represents the conceptual framework within which the meaning of A is to be understood, and $p(A|X)$ represents the numerical interpretation of the credibility of A . To a layman it may seem stilted for our observer to say: "The probability of the truth of A , conditional on the truth of X , is p ." As indicated in references 1, 2, and 3, there does not appear to be any alternative method of encoding the observer's knowledge which does not violate the elementary requirements of consistency, unambiguity, and universality.

This point in the development of the demonstration is critical and deserves elaboration. At issue are such difficult matters as *uniqueness*, *completeness*, and *freedom from inconsistency*. We are unable to give demonstrations of these qualities. What we can do, with Cox's and Jaynes' works as a guide, is to demonstrate that any alternative form of communication which is in contradiction with a code based upon the two fundamental equations of probability theory

$$p(AB|X) = p(A|BX) p(B|X) \quad (1)$$

$$p(A|X) + p(\sim A|X) = 1 \quad (2)$$

will be in violation of the requirements of consistency, unambiguity, and universality.

In this context, of course, p refers to a real number which interprets what our observer knows about A and not to any objective property of the real world. That is,

A may pertain to an observable but p pertains to knowledge about A . The problem facing our observer-receiver pair is to agree upon a method for assigning a value to p .

The substantive part of the message always refers to a well-defined question. A well-defined question Q is one for which the set of mutually exclusive *possible* answers is known. What is unknown is *which* of the possible answers is correct. We may imagine the answers to be arbitrarily numbered $A_1, A_2, A_3, \dots, A_n$, so that the question may be put simply as "which of the A_i is correct?" ($i = 1, 2, \dots, n$). The knowledge about the set A_i is summarized in the agreed upon interpretive part of the code which the observer uses to assign values to $p_i = p(A_i|OX)$. A well-trained observer does not assign $p_i = 1$ for any i , that is, he does not attempt to tell us that a specific answer is correct. Instead, he attempts to summarize his *actual* knowledge about the answers to the given question. An assignment of $p_i = 1$ for a particular i may be said to represent *deterministic* knowledge.

Shannon's function (4)

$$S = -\sum_i p_i \ln p_i \quad (3)$$

uniquely measures what the observer does *not* know, when all he can state is a probability distribution for the various A_i .

There are many derivations of Shannon's measure which show why the function S has the meaning ascribed to it. A simple derivation (5) shows that if the logarithm is taken to base 2, S measures the expected number of questions in a taxonomic game (that is, in the style of twenty questions) that the observer would have to ask to clear up all doubt.

Equation (3) supplies our observer with a measure to be used in a variational principle in the assignment of probabilities. The principle of maximum entropy, as given by Jaynes (6), is simply this: Assign that set of probabilities which maximizes S but which agrees with whatever is known to be true. By using this principle, the observer reduces his subjectivity to the minimum possible value.

The receiver of the set of probabilities does not find any value in asking "are these probabilities right?" In the context given here, there is no such thing as a correct or true probability. Probability assignment represents an encoding of knowledge. The only meaningful questions which may be asked are: "do these probabilities yield well-defined predictions?" and "are there any circumstances for which X is true?" If the answer to both these questions is yes, the receiver may well thank the observer for his efforts. If the answer to either of these is no, there is nothing for the observer to do but find additional data (that is, another X') until the answer is yes. There is no point, however, in trying to assign the probabilities in a different way, that is, in a way inconsistent with maximum entropy, for to do so would be equivalent to assuming facts not in evidence. There is no objection to clairvoyance or flashes of inspiration but these should not be represented as observations.

In particular, there is no need to investigate mental replications of the conceived system in the hope that ruminations over the properties of an infinite number of heat baths will give any additional insight. The only thing furnished by these imaginary ensembles is a confusion between the interpretive part of the message (p_i) and the fraction of the ensemble (f_i) for which statement A_i would be true (if X were true and the imaginary ensemble conformed to our knowledge). The disadvantage is that, once we have convinced ourselves that our ensemble justifies (!) our conclusions, we must then reconcile the fact that we actually shall be concerned with averages over *time* on *one* system. The solution to this dilemma is often

sought in the so-called *ergodic* problem, which attempts to show under what circumstances time averages and ensemble averages must be the same.

If a semantic confusion over the words *probability* and *frequency* were the only penalty, there would be little value in protracted discussions on the subject. However, once the distinction is lost, a great deal is lost with it. It is a consequence of the theory that if one *imagines* an ensemble, then the *expected* frequency will be computed in the following way. According to our knowledge of the ensemble, the probability of the state i is assigned a real number we denote by p_i . In n duplications of the system for which our knowledge is valid, the expected value of the number of times n_i that i will be the outcome is given by the theory as $\langle n_i | X \rangle = n p_i$. If we define frequency as the ratio $f_i = n_i/n$, then it is a consequence of the theory that we have $\langle f_i | X \rangle = \langle n_i | X \rangle / n = p_i$.

At this step the advocate of the ensemble-frequency approach may well say, "Aha! You *do* equate frequency and probability." This is not true. There is a difference between f_i and $\langle f_i | X \rangle$ and this difference becomes apparent if we compute the variance of f_i . (One would think that an equation in which both f_i and p_i occur would be convincing evidence that there are two ideas represented by two symbols, but experience indicates that this is usually not the case.)

In our view the role of probability in science can be best appreciated in the context supplied by Jaynes, Cox, and Jeffries (7), Kemeny (8), and a few other writers who consider probability theory to be an aid to inductive logic. This is a point of view explicitly held two centuries ago by Laplace and Bayes. It is implicit in the writings of Gibbs. Unfortunately, other views connected with other names became fashionable in probability theory and dominated the field in the intervening centuries.

When the probabilities have been assigned, they may be eliminated through the use of the expectation operator. Then there will remain only equations which connect observables with one another in the same fashion as a building remains after the scaffolding has been eliminated. The connectives will provide the basis for the *concepts* of a theory. The connectives arise because the mathematics introduces them to preserve the consistency of the eliminated probabilities. These concepts serve to *define* a theory.

It seems a bit presumptuous to some to imply that a theory is *defined*. The history books and elementary treatises in physics and chemistry leave us with the impression that theories are discovered. Thus Newton's phrase, "I feel like a small boy playing on the beach while oceans of truth lay around me to be discovered," certainly gives the impression that one needs only to dig and find rich veins of truths leading to richer deposits labeled theory. But in the last half century or so, scientists have adopted a different view. Few men will defend the statement, "energy is quantized." Most will say, "It certainly is useful to make computations as though energy were quantized," leaving to the future the possibility that another view may emerge as even more fruitful. We no longer pretend to have found truth but merely consistent descriptions which seem to have predictive value.

Norwood Hansen (9) made a revealing study of the manner in which men view the equation

$$F = ma \quad (4)$$

Is it a definition of a Newtonian frame of reference which is implemented by an *invented* "force" of gravity, or is it a statement of human experience? It cannot be both. If it is a *definition*, why shouldn't another definition lead to an equally acceptable dynamics? We shall not pursue the point here. Hansen (9) has done so admirably elsewhere. The point we wish to extract from this discussion is that

a theory prescribes at once both the definitions of its concepts and the circumstances in which they shall be related. In the case of Newtonian mechanics the matter may be summarized by saying, " $F = ma$ is true in a Newtonian frame of reference. A Newtonian frame of reference is one for which $F = ma$."

A similar summary pertains to the theory of thermostatics, a matter we shall now proceed to demonstrate.

THERMOSTATICS

Let us suppose our observer's X tells him the following:

What is called matter is composed of particles which move and interact with one another. The equations of motion of a particle, or system of particles, may, by suitable transformations, be put in such form as to display the *energy* of the motion. Suppose the observer's X also tells him that from quantum theory, this energy may in principle be written for any chosen system:

$$\epsilon = \epsilon(V, X_1, X_2, \dots, X_k; i) \quad (5)$$

where ϵ = energy, V = volume, X_k = force field produced by action of other bodies, and i = quantum number which serves to identify a quantum state. The quantum number i is an index for microscopic descriptions such that to each i there corresponds a composition (that is, a specification of n_{a_i}, n_{b_i}, \dots , the number of particles of type a, b, \dots associated with quantum state i) and a state of motion (that is, specification of how many particles are located at certain velocities, specified within quantum limitations).

Given a system, for which V, X, X_2, \dots, X_k are specified, the question the observer wishes to answer is, "What is the state of the system?" This is a well-defined question to which the answer is a value of i which defines a quantum state. Our observer and receiver both know that the best that they can do is give a value for the *probability* that the system is in a particular state. Furthermore, since the probabilities are derived from the principle of maximum entropy, it is not really necessary to compute the p 's. It is only necessary to give the information from which (by the maximum entropy principle) the p 's may be computed.

If we let $p_i = p(i|V, X_1, X_2, \dots, X_k, X)$ from the properties of the probability function, we have

$$\sum p_i = 1 \quad (6)$$

$$\sum p_i \epsilon_i = \langle \epsilon \rangle \quad (7)$$

$$\sum p_i n_{c,i} = \langle n_c \rangle \quad (\text{all } c) \quad (8)$$

Equations (3), (5), (6), (7), and (8) are the fundamental equations linking macroscopic and microscopic conceptions. They serve to define the macroscopic variables in terms of the microscopic. But the principle of maximum entropy serves to define our knowledge of the microscopic states in terms of the macroscopic variables. Thus the loop is closed. If we choose to use Newtonian or quantum mechanics at a macroscopic level, the principle of correspondence tells us to extend our views into the microscopic level in such a way as to retain consistency. The assignment of mass and energy to particles (and force fields among them) serves to extend these mechanical ideas to the microscopic level. The use of Equations (3), (5), (6), (7), and (8) guarantees that we shall be consistent on the return trip to "macroland." Of course, in all these trips from micro to macro and reverse, it is important to retain the difference between probability and frequency or there will be no logical basis for expecting the results *necessarily* to be valid.

The maximum entropy probability distribution (11) consistent with Equations (6), (7), and (8) is

$$\log p_i = -\Omega - \beta \epsilon_i - \sum_c \alpha_c n_{ci} \quad (9)$$

from which, on eliminating the probabilities, it is found (for details see references 11 or 12) that the entropy, energy, and composition are given by

$$S = k\Omega + k\beta \langle \epsilon \rangle + k \sum_c \alpha_c \langle n_c \rangle \quad (10)$$

$$\Omega = \ln \sum_i e^{-\beta \epsilon_i - \sum_c \alpha_c n_{ci}} \quad (11)$$

$$\langle \epsilon \rangle = -\partial \Omega / \partial \beta \quad (12)$$

$$\langle n_c \rangle = -\partial \Omega / \partial \alpha_c \quad (13)$$

But since

$$\Omega = \Omega(\beta, \alpha_a, \alpha_b, \dots; V, X_1, \dots, X_k) \quad (14)$$

it follows that

$$S = S(\beta, \alpha_a, \dots, \alpha_b; V, X_1, \dots, X_k) \quad (15)$$

$$\langle \epsilon \rangle = E(\beta, \alpha_a, \dots, \alpha_b; V, X_1, \dots, X_k) \quad (16)$$

$$\langle n_c \rangle = N_c(\beta, \alpha_a, \dots, \alpha_b; V, X_1, \dots, X_k) \quad (17)$$

For the purpose of communicating information about energy and composition, for systems about which there is nothing more to say, it is obviously sufficient to announce values for $V, X_1, X_2, \dots, X_k, \alpha_a, \alpha_b, \dots$, and β . This result has been suggested as a necessary fourth principle of classical thermodynamics and has been called the *state principle* (12). A simpler form of the state principle results from considering a closed system (that is, one for which $n_{ci} = n_c$ for all i and every c) for which Equations (9) through (16) reduce to

$$\log p_i = -\psi - \beta \epsilon_i \quad (9a)$$

where

$$\psi \equiv \Omega + \sum_c \alpha_c n_c$$

and

$$S = k\psi + k\beta \langle \epsilon \rangle \quad (10a)$$

$$\psi = \ln \sum_i e^{-\beta \epsilon_i} \quad (11a)$$

$$\langle \epsilon \rangle = -\partial \psi / \partial \beta \quad (12a)$$

$$\psi = \psi(\beta, V, X_1, X_2, \dots, X_k) \quad (14a)$$

$$S = S(\beta, V, X_1, X_2, \dots, X_k) \quad (15a)$$

$$\langle \epsilon \rangle = \epsilon(\beta, V, X_1, X_2, \dots, X_k) \quad (16a)$$

For closed systems, in the absence of effects due to capillary, magnetism, etc. (that is, invariance of X_1, X_2, \dots, X_k), the statistical or equilibrium state has two degrees of freedom (in agreement with the state principle of Klein and Koenig).

$$\psi = \psi(\beta, V) \quad (18)$$

$$S = S(\beta, V) \quad (19)$$

$$\langle \epsilon \rangle = \epsilon(\beta, V) \quad (20)$$

Whether or not the state principle deserves the status of a separate principle is a matter to be argued by those who wish to display classical thermodynamics as a separate phenomenological discipline. (The various approaches now being taken to the development of macroscopic phenomenological theory will be treated by the authors in a separate paper.) The conclusion that simple systems have two degrees of freedom is a result of the fact that if we state at the outset that we shall be interested only in systems which have an energy and are in some quantum state, and we rule out in advance any questions concerning magnetism, capillarity, composition, etc., and if we confine our experiments to situations for which these statements are in fact all that matter, then the probabilis-

tic treatment tells us that having posed but two questions we shall answer them with but two parameters. The only caveat to be observed is to be sure that the systems to which the theory is applied are at equilibrium. One of the well-known tests for equilibrium is, of course, the phase rule, which, as Klein and Koenig have pointed out, is derivable from the state principle.

For a closed system we ask two questions:

$$\sum p_i = 1 \quad (\text{In what quantum state is the system?}) \quad (21)$$

$$\sum p_i \epsilon_i = \langle \epsilon \rangle \quad (\text{What is its energy?}) \quad (22)$$

It does not follow, however, that the statistical results cannot be used to answer more questions. For example, we may be interested in *processes* between equilibrium states.

We may classify processes between equilibrium states into two groups: those which at every stage of the process are at equilibrium and those which are at equilibrium only at the end states. For a small change along an equilibrium path (a path for which we know something about what is going on) we have:

$$\sum dp_i = 0 \quad (23)$$

$$\sum p_i d\epsilon_i + \sum \epsilon_i dp_i = d\langle \epsilon \rangle \quad (24)$$

$$\sum \ln p_i dp_i + \sum dp_i = dS/k \quad (25)$$

$$\ln p_i = -\psi - \beta \epsilon_i \quad (26)$$

from which we find

$$dS/k\beta = d\langle \epsilon \rangle - \sum p_i d\epsilon_i \quad (27)$$

From the definition of energy, work, and force in mechanics, we have for a simple closed system (that is, in the absence of effects due to capillarity, magnetism, etc.)

$$\langle P \rangle = \sum p_i (\partial \epsilon_i / \partial V) \quad (28)$$

so that

$$dS/k\beta = d\langle \epsilon \rangle + \langle P \rangle dV \quad (29)$$

This equation serves to link together the macroscopic concepts of entropy, energy, pressure, temperature, and volume. As before, the supporting logical structure furnished by the probabilities has been eliminated. It remains only for the equation to define its own domain of validity.

For example, if the volume is held constant (or if $\int \langle P \rangle dV$ can be reduced to negligible proportions), we have

$$dS = k\beta d\langle \epsilon \rangle \quad (V = \text{const.}) \quad (30)$$

There is nothing in the theory which tells us we *ought* to invent the concept of heat. Indeed, some have suggested that it would be a good thing if the concept could be banished from thermodynamics. However, if we wish to extend the conversation to include processes *not* at equilibrium, it is convenient to enlarge our vocabulary. There are many kinds of nonequilibrium processes and it is necessary to choose a limited class of these processes to be assured that it is known which are the experimental conditions for which the theory may be applied. The simplest nonequilibrium interaction is one between a system and other bodies which are *themselves* in equilibrium. Thus, if we are interested in body A, we shall not be deterred if body A is out of equilibrium. However we insist that it be in contact only with bodies B, C, D, etc., and that B, C, D, ... are at equilibrium. We may then make use of our knowledge about B, C, D, ... to infer some things about A, making use of whatever we know about the nature of the interactions of A with B, C, D, ...

Suppose that body A can interact only with B and B only with A. Suppose, too, that the total volume $V_A + V_B$ is constant. From the principle of conservation of energy (not to be confused with the first law of thermodynamics, see reference 11, p. 136), we have $\epsilon_A + \epsilon_B = \text{const.}$;

hence, from the linearity of the expectation operator we have (neglecting energy of interaction, that is, surface effects)

$$\langle \epsilon_A \rangle + \langle \epsilon_B \rangle = \text{const.} \quad (31)$$

$$dS_B = k\beta_B d\langle \epsilon_B \rangle + k\beta_B \langle P_B \rangle dV_B \quad (32)$$

$$d\langle \epsilon_A \rangle + d\langle \epsilon_B \rangle = \quad (33)$$

$$d\langle V_A \rangle + d\langle V_B \rangle = 0 \quad (34)$$

Suppose, for the moment, that system A is at equilibrium. Then

$$dS_A = k\beta_A d\langle \epsilon_A \rangle + k\beta_A \langle P_A \rangle dV_A \quad (35)$$

and

$$dS_{\text{total}} = dS_A + dS_B = k(\beta_A - \beta_B) d\langle \epsilon_A \rangle + (k\beta_A \langle P_A \rangle - k\beta_B \langle P_B \rangle) dV_A \quad (36)$$

If, in addition, the system of A and B taken together is at equilibrium, then dS_{total} must be a maximum (for this is what we mean by equilibrium); hence by the line of reasoning which goes back to Gibbs, for the allowed variations, equilibrium is characterized by

$$\beta_A = \beta_B, \langle P_A \rangle = \langle P_B \rangle$$

Suppose, however, system B is in equilibrium at all times, but system A is not necessarily at equilibrium. In this case one may not write Equation (35), but one may nevertheless substitute Equation (32) into Equation (33) to obtain

$$d\langle \epsilon_A \rangle = -dS_B/k\beta_B + \langle P_B \rangle dV_B \quad (37)$$

and, from Equation (34), (dropping the subscript A for system A)

$$d\langle \epsilon \rangle = -dS_B/k\beta_B - \langle P_B \rangle dV \quad (38)$$

A calorimeter may be defined as any body for which it is possible to measure $dS_B/k\beta_B$. A pure heat effect may be defined for any process for which $dV = 0$, so that from Equation (38)

$$d\langle \epsilon \rangle = dS_B/k\beta_B = dQ \quad (39)$$

Similarly, a pure work effect may be defined for the case where no heat effect $dS_B/k\beta_B$ occurs, so that from Equation (38)

$$d\langle \epsilon \rangle = -\langle P_B \rangle dV = -dW \quad (40)$$

Substitution of Equations (39) and (40) into (38) gives the conventional form of the first law:

$$d\langle \epsilon \rangle = dQ - dW$$

The results just given serve to demonstrate how naturally the maximum entropy formalism defines the quantities which make up the theory of thermodynamics for the above prescriptions, for the macroscopic concepts of heat and work are precisely those which have been found necessary to have a coherent theory of thermodynamics. The information theory approach, however, has introduced one more idea which was not generally considered before: whenever we tautologically define two mutually exclusive interaction mechanisms (on any basis) by the equation

$$d\langle \epsilon \rangle = dQ' - dW' \quad (41)$$

then, in view of Equation (16) it will certainly be found that

$$\int d\langle \epsilon \rangle = \int (dQ' - dW') = 0 \quad (42)$$

since the idea of a cyclic integral guarantees a return to the same statistical state of information. It has often been suggested that the first law has been saved by inventing new forms of work. The information theory approach shows that the critical point is the way it is decided the cycle has been closed. New forms of work will always be found.

By substituting Equation (41) into (29), we have

$$dS = k\beta dQ + k\beta[\langle P \rangle dV - dW] \quad (43)$$

The last term is the uncompensated transformation of Clausius and serves to describe entropy change, not as an inequality, but as an equality due to the sum of the effects of heat addition and mechanical irreversibility. If dQ , $\langle P \rangle$, and dW are properly defined at a macroscopic level, Equation (43) may be viewed as a *definition* for entropy change. Indeed, this is the historic, classical approach.

CONCLUSIONS

In this exposition we have not demonstrated such well-known results as the fact that $\beta = 1/kT$, that the Lagrangian multipliers $\alpha_a, \alpha_b, \alpha_c, \dots$ correspond to the negative values of $\mu_a/RT, \mu_b/RT, \mu_c/RT, \dots$ and are the more natural potentials for chemical transfer since they occur in irreversible processes, and that the third law has a very simple interpretation in information theory. Our purpose here has been merely to show how the basic ideas of classical thermodynamics arise quite naturally out of the information theory approach.

It should be apparent from this development that just as Newtonian mechanics serves to define its own frame of reference, so does classical thermodynamics: "The laws of thermodynamics are valid if a system is at equilibrium. A system is at equilibrium if the laws of thermodynamics work."

The clarification of classical thermodynamics, although useful of itself, is not the sole benefit that can be derived from the information theory treatment. The deletion of unnecessary axioms, the elucidation of the connections among branches of science, the development of new ways of looking at subjects are all positive contributions to science. Not only is it possible to gain greater insight into thermodynamics via information theory, it is also possible to work in the inverse direction and use thermodynamic methods via information theory to give light on other fields. These contributions will be the subject of other papers.

ACKNOWLEDGMENT

Work on this paper was supported by a grant from the National Science Foundation whose assistance is gratefully acknowledged.

LITERATURE CITED

1. Tribus, Myron, and R. B. Evans, *Appl. Mech. Rev.*, **16**, No. 10, 765-769 (October, 1963).
2. Cox, Richard T., "The Algebra of Probable Inference," Johns Hopkins Press, Baltimore, Md. (1961).
3. Tribus, Myron, in "Heat Transfer, Thermodynamics and Education," Boelter Anniversary volume, H. A. Johnson, ed., McGraw-Hill, New York (1964).
4. Shannon, C. E., *Bell System Tech. J.*, **27**, 379, 623 (1948).
5. Cox, R. T., *ibid.*, 35.
6. Jaynes, E. T., *Phys. Rev.*, **106**, 620, **108**, 171 (1957).
7. Jeffreys, Harold, "Theory of Probability," Oxford Press, England (1961).
8. Kemeny, J., *J. Symbolic Logic*, **203**, 263-273 (September, 1955).
9. Hanson, Norwood R., "Patterns of Discovery," Cambridge Univ. Press, England (1958).
10. Tribus, Myron, *J. Appl. Mech.*, **1** (March, 1961).
11. ———, "Thermostatistics and Thermodynamics," D. van Nostrand, Princeton, N. J. (1961).
12. Kline, S. J., and F. O. Koenig, *J. Appl. Mech.*, 29-34 (1957); see also **79**, 625-630 (1957).

Manuscript received February 1, 1965; paper accepted October 1, 1965. Paper presented at A.I.Ch.E. San Francisco meeting.