

THE ROLE OF SUFFICIENCY AND OF ESTIMATION IN THERMODYNAMICS¹

BY BENOIT MANDELBROT

I. B. M. Thomas J. Watson Research Center

1. Introduction and summary. The purpose of this paper is to point out (and to use) the relations of certain statistical concepts with “statistical” thermodynamics.

(A) It is observed that Gibbs’s “canonical distribution” of energy is precisely what statisticians have later labeled a “distribution of the exponential type”. It follows that a rigorous treatment of the canonical law can be based upon the concept of “sufficiency”, which is thereby related to the physical idea of “thermal equilibrium” and to the “zero-th principle of thermodynamics”. In other words, *the theory of physical fluctuations can be based upon “principles” very similar to those of the “phenomenological”, or “classical, non-statistical” thermodynamics.* Naturally, our results will be less detailed than those of statistical *mechanics*. However, the foundations of the latter theory still raise a host of unanswered problems, and it seems good in the meantime to show that the less powerful phenomenological theory has a wider scope than is commonly thought (see also [15]).

The possibility of a purely phenomenological approach to statistical thermodynamics is not in itself a new idea. A procedure somewhat similar to ours has indeed been long ago suggested in Szilard’s admirable, but very difficult and neglected, paper [18]—not to be confused with his [19]. Of course, Szilard used a quite different vocabulary; but, with hindsight, one may now say that he has co-invented the concept of sufficiency with R. A. Fisher; by showing that, under certain regularity conditions, Gibbs’s canonical law is the only probability distribution with a single scalar sufficient statistic, Szilard also anticipated the results of G. Darrois [2], B. O. Koopman [10] and E. J. G. Pitman [16], but was partly anticipated by Poincaré [17].

(B) The second thesis of the paper is independent of Szilard, and concerns the concept of temperature. For systems with a canonical energy, the temperature is the parameter of the Gibbs distribution; as such it is undefined for isolated systems with a determined energy. However, it is necessary to generalize the concept of temperature to isolated systems. Several definitions have been proposed and, although they all safely converge mutually for the usual very

Received May 23, 1960; Revised March 5, 1962.

¹ This work was started at the University of Geneva, under partial support from the Rockefeller Foundation. It was summarized in [12], and tentatively presented in [13].

Many stimulating comments concerning this work were made by Professor Laszlo Tisza of the Physics department of M. I. T., by Dr. Paul M. Quay, S.J., and by Dr. Marko P. Schutzenberger. The referees of the paper were also unusually helpful.

large systems, the temperature remains mathematically ambiguous for small isolated systems; it also becomes physically meaningless.

We shall show that the temperature for systems-in-isolation should be viewed as a statistical estimate of the parameter of a conjectural canonical distribution, from which the presently isolated system may be presumed to have once been drawn. This interpretation explains the nature of the ambiguity of the concept of temperature; it also meets the actual practice of physicists; finally, some of the *a priori* conditions, which the physicists impose upon their “estimators”, turn out to correspond to the statistical conditions of consistency, unbiasedness, and efficiency. Physicists also use two very interesting variants of consistency and unbiasedness, which we shall study under the names of “self-consistency” and “self-unbiasedness”. The most commonly used temperature, due to Ludwig Boltzmann, turns out to be the maximum likelihood estimator.

In summary, we hope to show that it is a great pity that mathematical and physical statistics should have developed largely independently of each other, while using the same concepts. By combining the rigor of modern statistics with the intuitive vigor of thermodynamics, both should be served well. However, as things stand, the mathematical statistician should not hope to unearth in the literature of physics any result as yet unknown to him.

An important open problem suggested by this paper is the following. When sufficiency and estimation are defined in the most general terms, it seems that one should also be able to generalize the scope of thermodynamics. However, an approach such as that of P. R. Halmos and L. J. Savage [5] could not be applied to thermodynamics without substantial restrictions, as we shall show in Section 7. It remains to study these restrictions in greater detail, before one can assert that a non-void generalization of thermodynamics is possible. The problem is addressed to both mathematicians and physicists.

We shall strive to reduce to the minimum the detailed knowledge of physics required to read this paper. If the reader's appetite for information about thermodynamics has been awakened, he could do no better than to make use of references [11] and [20].

2. Presentation of the statements to be used as axioms. In order to avoid axioms picked out of thin air, we shall start in this section by deriving them as properties, obvious in the relevant frames of reference, of either matter-in-bulk or of canonical distributions.

2.1. Definitions relative to the canonical law. The statistical thermodynamics of equilibrium is concerned with physical systems that can be characterized by a few so-called “macroscopic” quantities, referred to as “variables of state”. Their number is independent of the size of the system; some of them can either be random variables or parameters—depending upon the problem—while others can only play the role of parameters.

Consider the energy of the system, a real strictly positive quantity. U follows Gibbs's law, whenever $dF(u) = d\text{Prob}(U \leq u)$ is of the form:

$$dF(u | \beta) = dG(u) \exp(-\beta u) / Z(\beta).$$

Here, $G(u)$ is a non-decreasing and right-continuous function with $G(0) = 0$.² The single parameter β is called the *inverse-absolute-temperature* and $Z(\beta)$ is called the *partition function* of the distribution and of the system to which it refers. One obviously has the relationships

$$Z(\beta) = \int_0^\infty \exp(-\beta u) dG(u); E(U | \beta) = -\partial \log Z(\beta) / \partial \beta;$$

$$E[U - E(U)]^2 = \partial^2 \log Z(\beta) / \partial \beta^2.$$

The third of these expressions being by definition positive, $E(U | \beta)$ is a decreasing function of β .³

We shall eventually need a reformulation of the Gibbs law in terms of a fictitious real variable G , as follows: A being an interval of the real line, and $u(g)$ being the inverse function of $G(u)$, one can write

$$\text{Prob}(G \in A) = \int_{g \in A} dg \exp[-\beta u(g)] / Z(\beta).$$

2.2. Systems in thermal equilibrium with heat reservoirs. The physical significance of the canonical distribution is that "it rules the energy of a system which is in contact and in thermal equilibrium with a heat reservoir." Let us comment upon the meaning of the various terms used in this statement:

Definition of contact. M physical systems S_m are said to be in thermal contact, if their respective energies are random variables. Degenerate variables, equal to $+\infty$, are not excluded.

Definition of thermal equilibrium. M systems S_m are said to be in thermal equilibrium, if the distributions of their energies are independent of the moment of observation.

Properties of the heat reservoirs.

(A) The energy of any heat reservoir is a degenerate random variable equal to $+\infty$

(B) If M systems S_m are in contact and in equilibrium with the same heat reservoir, their energies U_m are independent random variables.

(C) If a system is in contact and in equilibrium with a heat reservoir, its energy is a canonical random variable, or equals $+\infty$.

(D) If M systems S_m are in contact and in equilibrium with the same heat reservoir, the parameters β of their canonical distributions are identical. In particular, it is impossible that two heat reservoirs lead to the same probability distribution for the energy of S_1 and to different distributions for the energy of S_2 .

(E) The concept of temperature can also give a meaning to the statement that "heat flows from the warmer body to the cooler."

² In statistical mechanics, $G(u)$ is the number of microscopic states of the system of energy not greater than u . But in a purely phenomenological approach such as ours, $G(u)$ has no concrete interpretation.

³ The symbol of partial differentiation is meant to insure uniformity of notation with the developments of [15], in which external parameters—such as volume—are introduced.

The above properties can be expected not to be independent. Indeed, we shall find it sufficient to postulate the existence of physical systems defined by Property (B).

2.3. *Energy.* Let the G - and Z -functions of S_m be $G_m(u_m)$ and $Z_m(\beta)$. If the S_m are in contact and in equilibrium with a heat reservoir, one has

$$d \text{ Prob } (U_1 \leq u_1, \dots, U_M \leq u_M) = \prod_{m=1}^M dG_m(u_m) \exp(-\beta u_m) / Z_m(\beta)$$

and

$$d \text{ Prob } \left(\sum_{m=1}^M U_m \leq u \right) = dG(u) \exp(-\beta u) / Z(\beta),$$

where $Z(\beta) = \prod_{m=1}^M Z_m(\beta)$ and $G(u) = G_1(u) * G_2(u) * \dots * G_M(u)$.

Properties of energy.

(A) There exists a kind of contact, called *weak interaction*, such that M partial systems in weak interaction can be considered as making up a compound system, with a compound energy equal to the sum of the partial energies.

(B) If M systems interact weakly with each other and not at all with other systems, the compound energy remains constant in time.

(C) Energy is the unique non-trivial invariant of an isolated S , sum of partial S_m in weak interaction with each other. (That is, other invariants do not vary even when the compound S is put in thermal contact with an environment.)

The above properties are again not independent of each other, and from the properties of temperature and of thermal equilibrium. It will be sufficient to postulate unicity.

2.4. *Finite systems in thermal equilibrium with each other.* We now arrive at the most interesting of the properties of thermal equilibrium. For that, note that the joint distribution of the U_m , which was written in the last section, can be reformulated as follows:

$$\prod_{m=1}^M \frac{dG_m(u_m) \exp(-\beta u_m)}{Z_m(\beta)} = \frac{\prod_{m=1}^M dG_m(u_m)}{dG(u)} \frac{dG(u) \exp(-\beta u)}{Z(\beta)}.$$

Therefore, the conditioned distribution of the U_m , knowing β and the value u of the sum $U = \sum_{m=1}^M U_m$, will have as differential $\prod_{m=1}^M dG_m(u_m) / dG(u)$. If the systems S_m exchange energy through a heat reservoir exclusively, this expression has no physical meaning. Suppose however that, even when one isolates the S_m from the heat reservoir, these systems remain in weak thermal interaction. Then, u is the energy of the system compounded from the S_m and $dG(u) \exp(-\beta u) / Z(\beta)$ is its canonical distribution. What about

$$\prod_{m=1}^M dG(u_m) / dG(u)?$$

Depending upon the approach which one chooses, physics either shows or postulates that this is the differential of the joint energy distribution of the U_m , when the compound system $\sum_{m=1}^M S_m$ is isolated from all outside contact. Hence, one can make the following statements:

Properties of thermal equilibrium under weak interaction.

(A) In thermal equilibrium, under weak interaction, the joint distribution of the energies U_m of the parts S_m of a compound system is independent of whether these parts exchange energy directly or only through intermediates.

(B) The parameters that condition the distribution of any subset of the U_m must be invariant under thermal interaction.

(C) If M communicating systems S_m are in equilibrium when in contact with some environment, they remain in equilibrium when cut off from this environment.

(D) The joint distribution of the U_m after isolation from the environment is identical to their conditional distribution, if the set of S_m is not isolated but u is known.

If one disregards the physical content of the above properties, and recalls the unicity of the energy as an invariant of an isolated system, one sees that the above statements imply something about the canonical distribution.

Criterion of sufficiency: The nature of thermal equilibrium is such that, if a system is withdrawn from contact with a heat reservoir, the energy of that system is a necessary and sufficient (or minimal sufficient) statistic for the temperature of the heat reservoir.

2.5. *The set of possible values of energy, dominatedness and regularity.* Clearly, whichever the function $G(u)$, the set of possible values of a canonical U is independent of the parameter β . Similarly, consider an isolated $S_1 + S_2$, of compound energy u . For u_1 to be a possible value of U_1 , it is necessary that it be a possible value when S_1 is in contact with a heat reservoir. In both cases, the families of energy distributions are *dominated* by some measure.

In fact, the behavior of $G(u)$ is usually very special. In the so-called quantum case, $G(u)$ is reduced to jumps, the set of positions of which has no accumulation point, except perhaps at the ends of the interval of variation of u . In the so-called classical case, one feels free to attribute to $G(u)$ as many derivatives as one may need for any given purpose; at least, $G(u)$ is piecewise very smooth.

2.6. *Generalization of Section 2 to the "grand-canonical" case, where there is more than one invariant.* All that preceeds is immediately generalizable to the $(J + 1)$ -dimensional case, where the systems which interact with each other can exchange J kinds of particles as well as energy.⁴ The canonical law is replaced by the grand canonical, which is such that

$$F(u, n_1, \dots, n_J) = \text{Prob} (U \leq u, N_1 \leq n_1, \dots, N_J \leq n_J)$$

⁴ Reference [4] offers a formal generalization of thermodynamics, in which the invariants other than energy are not necessarily the numbers of particles, and hence are not necessarily integral.

satisfies

$$dF = dG(u, n_1, \dots, n_J) \exp(-\beta u - \sum_{j=1}^J \gamma_j n_j) / Z(\beta, \gamma_1, \dots, \gamma_J).$$

Heat reservoirs are replaced by heat and particle reservoirs, but the rest of the treatment is essentially unchanged.⁵

3. First axiomatic of the canonical distribution. The properties of the canonical law, derived in Section 2, are surely compatible. We shall now extract a subset of these properties, from which the canonical law will be shown to follow and which will therefore constitute a complete system of axioms. We know however that our system is not minimal and we shall propose a better one in Section 4.

3.1. The state-space and conditions of regularity. Suppose that all the possible “macroscopic states” \mathbf{G}_m of the physical system S_m can be described by points of a subset of an h_m -dimensional euclidean space; the possible “states” \mathbf{B} of other systems, that can play the role of an “environment” for S_m , will be points of a subset of a c -dimensional euclidean space. S_m will be said to be in contact with its environment, if \mathbf{G}_m is a random variable with \mathbf{B} as parameter.

We shall only consider two types of random variables: discrete or “smooth.” Further, and even though this may seem strange in an axiomatic, we shall define “smoothness” only to the extent of saying that, at the most, the density of \mathbf{G}_m must satisfy the conditions given by Dynkin [3]; the reason for this procedure is that smoothness will be required only at one stage, in a presumably generalizable theorem, and we would not like our axiomatic to need any deep revision if such a generalization succeeds.

3.2. Axioms associated with the “zero-th principle of thermodynamics.”

Conditioning. A physical conditioning is any set of operations, realizable by means of macroscopic physical operations, which imposes a mathematical relation upon the random “state” of a system.

Axioms of thermal equilibrium.

(A) Let a system be in equilibrium, first under the condition C' , then under the stricter condition C'' . The distribution of its random state under C'' will be statistically independent from its distribution under C' .

(B) The equilibrium distribution under the physical condition C'' may be obtained as follows: begin by the distribution under the less strict condition C' , consider C'' as a mathematical relation, and apply the rules of probability theory relative to conditioning.

It follows from (A) and (B) that the parameters of a system after con-

⁵ Actually, the many variable function G can be reduced to a one-variable function, in several fashions that correspond to the different physical characteristics of the so-called statistics of Maxwell-Boltzmann, Bose-Einstein and Fermi-Dirac. Of course, this reduction involves further axioms; see [14].

ditioning must constitute a necessary and sufficient statistic for its parameters before conditioning.

Subjective and objective interpretations of equilibrium. For Szilard, the state of equilibrium is “really” characterized by a statement analogous to the preceding axiom. However, one can also say that the properties of equilibrium rather refer to the best that an observer “could” or “should” say about a system, if he only knows a small number of sufficient statistics (see [13] and [15b]).

Auxiliary axiom: existence of heat reservoirs. There exists a class of physical systems, to be called the heat-reservoirs, which can simultaneously serve as environments for several distinct S_m and are such that the corresponding states \mathbf{G}_m are statistically independent. Moreover, the carriers of the corresponding probabilities of \mathbf{G}_m are independent of the heat reservoir used as environment.

3.3. *First form of the axiom associated with the first principle of thermodynamics. Axiom of energy.* Consider a composite system S made up of parts S_m . The only time-invariant of the compound system S is real one-dimensional and it can be written as a sum of functions of the states \mathbf{G}_m of the partial S_m ; this invariant is called “energy.”

A theorem of Poincaré and the canonical distribution of energy for replicas of the same system. Consider M systems, which are replicas of each other, in the sense that they have the same probability distribution for energy, whichever their environment. Our axioms require that an expression of the form $\sum_{m=1}^M u_m(\mathbf{G}_m)$ be a sufficient statistic for the estimation of the parameter \mathbf{B} of a heat reservoir from the states \mathbf{G}_m . For that, it is in particular necessary that the maximum likelihood estimate of \mathbf{B} be a function of $\sum_{m=1}^M u_m(\mathbf{G}_m)$ alone. Poincaré [17] has shown that this is possible if and only if there exists a non-decreasing function $G_m(u_m)$ and a real function $\beta(\mathbf{B})$, such that

$$dF(u_m | \mathbf{B}) = dG_m(u_m) \exp [-\beta(\mathbf{B})u_m(\mathbf{G}_m)]/Z(\beta).$$

Conversely, $\sum_{m=1}^M u_m(\mathbf{G}_m)$ is a sufficient statistic for the parameter $\beta(\mathbf{B})$ of this law, irrespectively of the chosen method of estimation.

Comment. We have chosen the previous argument in order to be able to use what we believe to be the oldest theorem deriving the canonical distribution from arguments concerning estimation. By current standards, the reference to maximum likelihood is of course quite unnecessary. Moreover, the proof of [17], which makes many implicit assumptions of regularity, is also obsolete; but the existence of the theorems to be used in Section 4.1 makes it pointless to attempt to update the proof of Poincaré, or the similar proof of p. 197 of [6a] (note that Keynes does not refer to [17]).

4. Second axiomatic of the canonical distribution. In Section 3.3, we postulated the additive character of energy. Actually, as we now proceed to show, this additivity is an automatic consequence of our form of the zero-th principle. The present approach starts with Sections 3.1 and 3.2 and continues as follows.

4.1 *The Dynkin-Jeffreys form of the theorem of Poincaré, Szilard, Darmonis,*

Koopman and Pitman. Let the state and parameter spaces be both euclidean, and the probability distribution of the state be regular in the sense of Section 3.1. The existence of non-trivial sufficient statistic for the parameter then requires that the probability distribution of the state be of the form

$$dF(\mathbf{G} | \mathbf{B}) = dG(\mathbf{G}) \exp \left[- \sum_{j=0}^J \beta_j(\mathbf{B}) u_j(\mathbf{G}) \right] / \zeta(\mathbf{B}).$$

The various functions introduced in this formula have the following meanings. $G(\mathbf{G})$ is either discrete (Jeffreys case) or "smooth" (Dynkin case). $\beta_j(\mathbf{B})$ and $u_j(\mathbf{G})$ are linearly independent. Under these conditions, the sufficient statistic is constituted by the $J + 1$ expressions $\sum_{m=1}^M u_j(\mathbf{G}_m)$.

Proofs will be found in the original references [3] and [6]. The above result should be generalizable to the case where $G(\mathbf{G})$ is a mixture of discrete and smooth parts.

4.2. *Axiom of energy.* The sufficient statistic has now been shown to be necessarily of the additive type. We have earlier postulated that it must be a physical invariant in thermal equilibrium. We now add the postulate that it is one-dimensional.

The canonical distribution of energy for replicas of the same system. The above axiom means that $J = 0$ in the Dynkin-Jeffreys theorem, so that

$$dF(\mathbf{G} | \mathbf{B}) = dG(\mathbf{G}) \exp [-\beta(\mathbf{B})u(\mathbf{G})] / \zeta(\mathbf{B}).$$

The condition, that the probability of the whole state space be one, immediately shows that $\zeta(\mathbf{B})$ depends upon \mathbf{B} through $\beta(\mathbf{B})$ only, so that it can be written as $Z(\beta)$. Moreover, the first and second terms on the right depend upon \mathbf{G} through u only, so that one can integrate over all states of given energy u , to obtain the desired expression

$$dF(u | \beta) = dG(u) \exp (-\beta u) / Z(\beta).$$

Generalization to the case of more than one invariant. If there are $J + 1 \geq 1$ invariants, there are $J + 1$ parameters, each of which is associated with one invariant.

4.3. *The theorem, that the same scale of temperature applies irrespectively of the structure function $G_m(u_m)$.* By appropriate renormalization of the scale of the β , one can prove the usual statement of the zero-th law of thermodynamics, that different systems in contact with the same heat reservoir have the same temperature.

Proof of the universality of the scale of temperature. From the theorem of Poincaré, it follows that there exist two functions $\beta''(\mathbf{B})$ and $\beta'(\mathbf{B})$, not necessarily identical, and such that when two non-identical systems are in contact with the same heat reservoir, the distributions of U' , U'' and $U + U''$ satisfy

$$\begin{aligned} dF'(u' | \beta') &= dG'(u') \exp (-\beta' u') / Z'(\beta'), \\ dF''(u'' | \beta'') &= dG''(u'') \exp (-\beta'' u'') / Z''(\beta'') \\ dF(u' + u'' | \beta) &= dG(u' + u'') \exp [-\beta(u' + u'')] / Z(\beta), \end{aligned}$$

where the function $\beta(\mathbf{B})$ depends upon \mathbf{B} by the intermediate of $\beta'(\mathbf{B})$ and $\beta''(\mathbf{B})$. However, a fresh application of sufficiency requires that

$$dF'(u' | \beta') dF''(u'' | \beta'') / dF(u' + u'' | \beta)$$

be independent of \mathbf{B} . Let u' change by $\Delta u'$ and u'' by $\Delta u''$; then, the logarithm of the last written ratio changes by

$$-\Delta u'(\beta' - \beta) - \Delta u''(\beta'' - \beta).$$

This expression is independent of β'' and of β' (and hence of β), if and only if $\beta' - \beta$ and $\beta'' - \beta$ are constants. Hence, one can make β , β' and β'' equal by multiplying $dG''(u)$ by $\exp[(\beta - \beta'')u]$ and multiplying $dG'(u)$ by $\exp[(\beta - \beta')u]$.

In the case of greatest interest, the regions of convergence of the Laplace integrals $Z(\beta)$, $Z'(\beta)$ and $Z''(\beta)$ are half-planes. In that case, one can choose the origin of the β so that the abscissa of convergence of $Z(\beta)$ is equal to zero for all systems.

Comment. It should be stressed that, the preceding theorem is a consequence of the zero-th and first principles of thermodynamics. It is not necessary to follow Szilard in using the second principle of thermodynamics in its derivation.

4.4. *Comments on Dynkin's proof, on the euclidean character of the state space and on the concept of entropy.* It will be noted that entropy played no role in our approach (see Section 7.3). However, an entropy-like expression plays a central role in Dynkin's proof. Indeed, whichever the discrete or smooth distribution of the state, a sufficient statistic is always constituted by the following function of the parameter:

$$\log p(\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_M | \mathbf{B}) - \log p(\mathbf{G}_1, \dots, \mathbf{G}_M | \mathbf{B}^0),$$

where p is either a discrete probability or a density. If the zero-th law holds in the form in which we postulate it, this statistic becomes

$$\sum_{m=1}^M [\log p(\mathbf{G}_m | \mathbf{B}) - \log p(\mathbf{G}_m | \mathbf{B}^0)].$$

Dynkin's proof is based upon the consideration of the minimal linear space of functions consisting of constants and of the functions $\log p(\mathbf{G} | \mathbf{B}) - \log p(\mathbf{G} | \mathbf{B}^0)$ for all admissible \mathbf{B} . If that space has an infinite dimension, there is no non-trivial sufficient statistic. If the dimension takes the finite value $J + 1$, one obtains the law of Section 4.1.

There is a close similarity between this argument and those used in certain physical treatises, such as the one due to Landau and Lifshitz [11], which is highly respected for its intuitive vigor. In those approaches, one also starts with a state space, which needs not even be separately postulated to be euclidean, because it was from the outset defined using as coordinates the position and momentum parameters of analytical mechanics. The concept of "state," as used by Landau and Lifshitz, has a physical meaning and hence has more properties than ours. In particular, due to Liouville's theorem, the logarithm of the

probability of a cell in the state space is an invariant of motion, additive in the addition of parts S_m following our zero-th law. Hence, this logarithm is a linear form of all other constants of motion, i.e., of energy. This finally leads to the canonical law, as written in the last line of Section 2.1; moreover, the function $G(u)$ is thereby interpreted as being the number of states having an energy at most equal to u .

In our approach, all that can be said *a priori* of $J + 1$, is that it can not exceed the dimension more or less arbitrarily attributed to the euclidean space in which the values of the parameter \mathbf{B} are imbedded.

5. Genuine temperature estimation.

5.2. *Thermometry.* A thermometer is, by definition, a *physical system such that the value of its energy can be ascertained by direct observation*. Hence, from the reading of a thermometer in contact with a heat reservoir, one will determine the temperature of that reservoir by ordinary estimation. If the thermometer is very large, or if one has a large number of independent readings of a small thermometer, the temperature of the heat reservoir may be determined with arbitrary precision.

For the sake of convenience, thermometers always carry a scale labeled in units of temperature, and not of energy; this scale is meant to be used with a single reading and it realizes physically the designer's chosen estimator. More precise thermometers will carry two scales, giving respectively the lower and upper confidence limit corresponding to a confidence interval. Still more precise thermometers are not absolute but differential; on the basis of their energy, the physicist will be able to say whether $\beta > \beta^0$ or $\beta < \beta^0$; it is well-known that (Gibbs's distribution being of the exponential type) there exists a uniformly most powerful differential thermometer requiring a single reading.

5.2. *Use of a thermometer to evaluate the energy of an isolated nonthermometric system.* If the isolated system S is not a thermometer, its energy u is fixed but unknown, and one can say nothing in practice about its future behavior or the behavior of its parts—even though in theory they are both fully determined. However, if one accepts a slight perturbation in the energy u , prediction becomes possible. Indeed, let G_t , G_s and G_{t+s} be the G -functions of the thermometer, the system and the sum of the two; u_{t1} being the initial energy of the thermometer, its later energy u_t has the following differential element:

$$dG_t(u_t) dG_s(u + u_{t1} - u_t) / dG_{t+s}(u + u_{t1}).$$

With the help of very many readings of u_t , one will determine exactly the parameter $u + u_{t1}$ and hence the final energy of the system, $u + u_{t1} - u_{t\text{final}}$. If the thermometer is very small, the perturbation $u_{t1} - u_{t\text{final}}$ becomes negligible.

It may however happen that one can only take one reading of the thermometer. Suppose then that the scale is expressed in terms of temperatures corresponding

to the estimator favored by the designer. In that case, the energy $u + u_{t1} - u_{t \text{ final}}$ will be some function of the "apparent temperature" shown by the thermometer.

6. The concept of temperature for an isolated system.

6.1. *Boltzmann's and Gibbs's definitions of a temperature for an isolated system.* For various reasons, which we shall give below, the following expressions have been used as "analogs" of temperature in the case of isolated systems.

Boltzmann's analog $\hat{\beta}_b$. It is the solution of the equation

$$u = E(U | \hat{\beta}_b) = -\partial \log Z(\beta) / \partial \beta |_{\beta=\hat{\beta}_b}.$$

Gibbs's differential analog, $\hat{\beta}_{gd}$. If $G(u)$ has two derivatives, let

$$\hat{\beta}_{gd} = (\partial / \partial u) \log [\partial G(u) / \partial u].$$

This expression is mostly useful when it is an increasing function of u ; then $u = \text{Mode}(U | \hat{\beta}_{gd})$, as one easily ascertains.

Gibbs's integral analog, $\hat{\beta}_{gi}$. If $G(u)$ has one derivative, let

$$\hat{\beta}_{gi} = (\partial / \partial u) \log [G(u)].$$

6.2. *The concept of a self-consistent and self-unbiased temperature for an isolated system.* The most basic and most "hard-boiled" reason for speaking of a temperature for an isolated system is the following. Suppose that the system is compounded of the parts designated as S_m . Then, it is very frequently necessary, in order to predict the outcome of macroscopic experiments, to evaluate expressions of the form $R(U_1, \dots, U_M)$; typically, the function R either depends on a single u_m or is symmetric in the u_m ; usually, one can even write R as $\sum_{m=1}^M R(U_m)$. Unfortunately, such evaluations are usually analytically untractable. But, for most functions R , an adequate approximation is obtained with the following assumptions.

(A) Each individual U_m is canonical, with an inverse-temperature given by one of the expressions of Section 6.1, or some similar expression.

(B) The U_m are statistically independent. Or, at least, if I is small with respect to M , then $U_{m(1)}, \dots, U_{m(i)}, \dots, U_{m(I)}$ are independent, whichever the indices $m(i)$.

That is, let $S' = \sum S_{m(i)}$ and $S'' = S - S'$ have the respective G - and Z -functions $G'(x)$, $G''(x)$, $Z'(\beta)$ and $Z''(\beta)$. Then, for all practical purposes, it is possible to approximate the joint probability

$$dG_{m(1)}[u_{m(1)}] \cdots dG_{m(I)}[u_{m(I)}] dG''[u - \sum u_{m(i)}] / dG(u),$$

by the following product of independent canonical laws:

$$\prod_{i=1}^I dG_{m(i)}[u_{m(i)}] \exp [-\hat{\beta} u_{m(i)}] / Z_{m(i)}(\hat{\beta}).$$

Actually, the independence and the canonicity of the $U_{m(i)}$ both follow from the canonicity of the sum $S' = \sum S_{m(i)}$, since both require the possibility of approximating

$$dG'' \left[u - \sum_{i=1}^I u_{m(i)} \right] / dG(u)$$

by $\prod_{i=1}^I \exp [-\hat{\beta} u_{m(i)}] / Z_{m(i)}(\hat{\beta}) = \exp (-\beta u') / Z'(\hat{\beta}).$

Definition of self-consistency. Whichever its definition, the “error” in the canonical approximation is positive. $\hat{\beta}$ is said to be *self-consistent*, if this error tends to zero as the “size” of the compound system S increases to infinity, for example as the number M of identical systems S_m tends to infinity. (The idea is old, but the term “self-consistent” is new in this context.) The asymptotic argument of self-consistency is the only motivation that certain physicists will accept for the $\hat{\beta}$ of their choice. As the system size increases, the various definitions of Section 6.1 converge mutually; hence either none is self-consistent or all are.

Self-consistency can take various forms. For example, Khinchin [9] proves it with respect to the following weighted distance function between the approximating and the true distributions:

$$\sup_{0 < u < \infty} \{ |F_{\text{true}}(u) - F_{\text{approx}}(u)| \exp(\alpha u) \}, \quad \text{with } \alpha < \hat{\beta}.$$

Earlier works of Khinchin,⁶ as well as most treatises on physics, limit themselves to proving self-consistency with respect to some class of functions $R(U_1, \dots, U_M)$: they prove the mutual convergence, in relative values, of the functions R computed on the basis of the approximating U_m and of the functions R computed on the basis of the approximated U_m .

The concept of self-unbiasedness. When one deals with self-consistency only relatively to a class of functions R , one may usefully introduce the following

⁶ It seems appropriate at this point to comment upon the evolution of Khinchin’s thinking. Between his best-known [7] and his later writings, the proof of the self-consistency of $\hat{\beta}_b$ became ever more elementary and stronger. For example, [7] concerns the case of a system S made up of subsystems S_m , such that the variety of different S_m ’s increases without bound as the size of S increases. Such generality is seldom needed in practice. Moreover, the best local central limit theorem available around 1943 was quite complicated. Therefore, [7] remains circuitous, even though it only yields self-consistency with respect to the expected values of functions of the form $\sum R(u_m)$. From the practical viewpoint, it is obviously better to eliminate all extraneous difficulties from the first stage of the theory, by limiting it to identical subsystems S_m . This was indeed done by Khinchin in his [9], which also treats more general cases, and to which we refer the reader for proofs.

Note that our considerations have no counterpart in any of Khinchin’s publications known to us. Although he was evidently aware that the Gibbs distribution satisfies the properties derived in Section 2, Khinchin does not use—and perhaps did not know—the fact that no law but the canonical satisfies those properties. The canonical law is simply introduced as a formal auxiliary device. Therefore, our Section 4 may also be used as a preliminary to Khinchin method.

concept: A definition of $\hat{\beta}$ is said to be "self-unbiased" with respect to a function R , if the expected value of R does not change when one replaces the actual distribution of the U_m by a product of independent canonical distributions of parameter $\hat{\beta}$.

Although the term "self-unbiased" seems new, the idea has been widely used, in particular in the cases where R is analytically related to the functions that define the three "analogs" of Section 6.1. Let us explicitly show what we mean (we can replace ∂ by d all through).

Gibbs's differential analog. Let $R = (d/du_m) \log [dG_m(u_m)/du_m]$. Designating its actual and canonical means by E_m and E_c , we have:

$$\begin{aligned} E_c &= \int_0^\infty (d/du_m) \log [dG_m(u_m)/du_m] dG_m(u_m) \exp(-\hat{\beta}u_m)/Z_m(\hat{\beta}) \\ &= \int_0^\infty (d/du_m)[dG_m(u_m)] \exp(-\hat{\beta}u_m)/Z_m(\hat{\beta}) = \hat{\beta}, \end{aligned}$$

a function of $\hat{\beta}$ alone;

$$\begin{aligned} E_m &= \int_0^u (d/du_m) \log [dG_m(u_m)/du_m] dG_m(u_m) dG_m''(u - u_m)/dG(u) \\ &= \int_0^u (d/du_m) dG_m(u_m) dG''(u - u_m)/dG(u) = [(d/du) dG(u)]/dG(u) \\ &= (d/\dot{d}u) \log [dG(u)/du]. \end{aligned}$$

One notes that E_m takes the same value for all S_m and that this function R is self-unbiased if $\hat{\beta} = \hat{\beta}_{gd}$.

Gibbs's integral analog. Let $1/R = (d/du_m) \log [G_m(u_m)]$. Easy calculations yield $E_c = 1/\hat{\beta}$ and $1/E_m = (d/du) \log [G(u)]$. Hence this R is self-unbiased if $\hat{\beta} = \hat{\beta}_{gi}$.

Boltzmann's analog. Let $R = u_m$. In general, this R is *not* self-unbiased when one chooses $\hat{\beta}_b$. Indeed,

$$E_c = -d \log Z_m(\hat{\beta}_b)/d\hat{\beta}_b$$

depends upon $Z_m(\beta)$ and upon $Z(\beta)$; that is: it depends upon the values of the functions $G_m(x)$ and $G(x)$ for *all* values of x . On the contrary,

$$E_m = \int_0^u u_m dG_m(u_m) dG_m''(u - u_m)/dG(u)$$

depends only upon the values of $G_m(x)$ and $G(x)$ for x smaller than u .

However, self-unbiasedness is satisfied in the single most important special case, when all the functions $G_m(x)$ are identical. In that case, one finds that $E_c = E_m = u/M$.

The last of the above examples shows that self-unbiased definitions of $\hat{\beta}$ cannot in general depend upon the whole range of values of $G(u)$.

Comment. It is clear that the concept of $\hat{\beta}$ would be of little value, if one insisted on self-unbiasedness, since each R would yield its own $\hat{\beta}$. The idea of self-consistency is far more important. Indeed, authors such as H. A. Lorentz and E. Schrödinger have used almost lyrical terms to hail the "unsensitivity" of thermodynamics, with respect to the method chosen to present it is the consequence of this self-consistency. Statisticians would be very happy if large-sample arguments could be a source of similar pride for them!

6.3 *Second motivation for a temperature-in-isolation. Use of unbiased estimation to extend the scope of the parameter of the canonical distribution.* We now come to physicists who wish to extend the idea of thermal equilibrium to systems which are currently isolated from each other, but might be put into contact. Those physicists use arguments which are clearly of an inductive character.

To begin with, let us consider an isolated system of energy u and let us put it into contact with a heat reservoir of appropriately chosen temperature $1/\hat{\beta}$. It is clear that the *expected* exchanges of energy can be made to vanish by choosing $\hat{\beta}_b$. Similarly, the *most probable* exchanges of energy vanish if one chooses $\hat{\beta}_{gd}$.

Further, let us put in contact two systems, of G -functions $G'(u)$ and $G''(u)$, which are now isolated and have the energies u' and u'' . The condition that the flow of energy be most probably zero happens to be conveniently expressed by the equality of the $1/\hat{\beta}_{gd}$ temperatures of the two systems.

Proof. Given $u' + u''$, the most probable value of the energy of S' in equilibrium is obtained by maximizing

$$dG'(x) dG''(u' + u'' - x)/dG(u' + u''),$$

For the maximum to be attained when $x = u'$, it is necessary and sufficient that

$$\left. \frac{\partial}{\partial x} \log \frac{dG'(x)}{dx} \right|_{x=u'} = \left. \frac{\partial}{\partial x} \log \frac{dG''(x)}{dx} \right|_{x=u''},$$

which expresses the equality of the temperatures $\hat{\beta}_{gd}$.

The $\hat{\beta}_b$ unfortunately have no similar property, for the same reason that the related function R is not self-unbiased. However, in the special case where S' and S'' are made up of r' and r'' replicas of some atom, equality of the $\hat{\beta}_b$ is again a necessary and sufficient condition for the expected flow of energy to vanish. Indeed, the equilibrium partition of the total energy $u' + u''$ among S' and S'' is obviously such that

$$\text{expected conditioned energy of } S' = [r'/(r' + r'')](u' + u''),$$

$$\text{expected conditioned energy of } S'' = [r''/(r' + r'')](u' + u'').$$

In order that these energies be respectively equal to the initial energies u' and u'' , it is required that $u'/r' = u''/r''$; that is, one must have

$$-d \log z(\beta)/d\beta|_{\beta=\hat{\beta}_b} = -d \log z(\beta)/d\beta|_{\beta=\hat{\beta}_b},$$

which expresses the equality of the temperatures $1/\hat{\beta}_b$.

The above properties of $\hat{\beta}_b$ and $\hat{\beta}_{gd}$ make many physicists say that one or the

other of these expressions represents the best evaluation of the temperature for an isolated system. Needless to say, this argument amounts to the reconstruction (or the construction) of a heat reservoir, with which the currently isolated system was (or might have been) originally in equilibrium. The temperature of that reservoir is *estimated*, in the usual statistical sense of the word, using a procedure that is mean—or mode—unbiased with respect to the function $E(U | \beta)$ of the parameter. β_b is also a maximum-likelihood estimator.

In the further development of the theory (see Section 7.3), each estimate of temperature will be associated with an estimate of entropy. All these estimates will converge, in relative values, for very large systems in equilibrium, but will remain ambiguous for small systems. For systems which are not in equilibrium, the various entropies may differ even for large physical bodies; hence, the extension of thermodynamics beyond equilibrium may well be undeterminate.

6.4. *The role of the usual concept of statistical consistency.* Suppose now that the system S has actually been prepared by contact with some heat reservoir. Then, the ordinary consistency of the estimators $\hat{\beta}$ can be combined with their self-consistency. Together, they show that, by making the system S sufficiently large, one may reduce at will the probability that the isolation of S from a heat reservoir perturb the value of any function $R(U_1, \dots, U_M)$ significantly. Hence, one proves that the heat reservoirs, the existence of which had been postulated and axiomatized in earlier sections, can also be constructed effectively by adding to our system S_m a large number of other systems, say of replicas of S_m , and by isolating the whole from outside thermal contact.

6.5. *The role of the lower bound to the variance of $\hat{\beta}$; the concept of temperature fluctuations.* Let us now examine the role played in thermodynamics by efficiency, the third of the major requirements which statisticians like to impose upon parameter estimates. This will involve the theory of the fluctuations of physical quantities other than energy, density, volume, etc., which already occur in mechanics. The question is very fundamental, but it has remained to this day very obscure and quite controversial. Without entering into any of the controversies, we would like to say that we believe at least part of the difficulties to be due to an inadequate recognition of the statistical conditioning relevant to each case, so that progress in the understanding of the foundations may well have practical implications.

The present exposition being centered upon systems with one random variable and one parameter, the only non-trivial fluctuation concerns the temperature of an isolated system. Landau and Lifshitz describe it in these words: "When applied to an isolated body, the Gibbs distribution completely describes its statistical properties, except that it gives very small, but nevertheless non-zero fluctuations of the total energy of the body, which cannot exist in reality. Conversely, if one assumes the energy to be a given quantity, then one cannot ascribe a completely definite temperature to the body and one must assume that the latter undergoes fluctuations which evidently determine the accuracy with which the temperature of an isolated body can be specified" [11, p. 355]. From

the physical viewpoint, this is clear enough. But, in order to really understand what the authors mean, one must look up the result of their computation. One finds that their formula for the square of the fluctuation of $\hat{\beta}$ is precisely $[d^2 \log Z(\beta)/d\beta^2]^{-1}$, that is, the Fisher-information lower limit to the variance of $\hat{\beta}$ considered as an estimate. In the case of the estimate $-\partial \log Z(\hat{\beta}_b)/\partial \hat{\beta}_b$ for the parameter $-\partial \log Z(\beta)/\partial \beta$, this limit is achieved irrespectively of the size of the system; in the case of $\hat{\beta}_g$, the lower limit is achieved only asymptotically.

We believe that, if one adopts the variance as definition of temperature fluctuation, there will be no reason for any reluctance in the use of this concept. Actually, our definition also has a very intuitive Kinetic meaning. Indeed, in a perfect gas, the mean kinetic energy of the molecules of a perfect gas is proportional to the temperature. If one knows the mean of the energy over a finite sample only, one will consider that it is compatible with any value of temperature contained in some interval.

7. Generalizations and other remarks.

7.1. *On the evaluation of $G(u)$.* Our theory includes a so-far unspecified function $G(u)$, which should be determined through measurements of such macroscopic quantities as heat capacities. Hence, our approach can be kept "phenomenological" all through, and realizes a dream of the 19th century "energetists": to describe matter-in-bulk without reference to atoms. It is a pity that all energetists have passed away long ago.

7.2. *A reference to statistical mechanics.* In statistical mechanics, one does not postulate the existence of heat reservoirs. The canonical law of Gibbs is derived, as a self-consistent approximation, from assumptions which we also like to call the zero-th and first principles (see [15]):

(A) zero-th principle: all *microscopic* states compatible with the constraints have *identical* probabilities;

(B) first principle: the constraints are linked to a *single* invariant, the energy, which is a *sum-function* of the form $\sum u(\mathbf{G}_m)$.

Of course, in order to link statistical and analytical mechanics, it is necessary to derive these axioms from more elementary physical considerations. But, as long as this aim has not been achieved, the axioms of this non-phenomenological theory can usefully be numbered to make them parallel to those of the phenomenological approach.

In our approach, the fact that energy is a sum-function is replaced by the existence of heat reservoirs, and hence is transferred from the "first principle" to the "zero-th."

7.3. *The need for a "second principle of thermodynamics."* The considerations of the present paper will be continued in [15], and will in particular include a treatment of the concept of entropy. We shall see that this continuation requires a further fundamental assumption, which can always be associated with the "second principle of thermodynamics."

This restricts considerably the scope of physically admissible distributions

of the exponential type; in particular, most random variables of the exponential type *cannot* be divided into “work” and “heat”, and the link between entropy-information and heat has no counterpart in the general case.

7.4. *Possibility of a generalization of the scope of sufficiency, as used in thermodynamics.* The regularity and dimensionality considerations of Sections 2.5 and 3.1 raise an open problem. On the one hand, as of today, the converse derivation of the canonical law from sufficiency can only be carried under the conditions of E. B. Dynkin [3] or of H. Jeffreys [6], which are precisely those of the classical and quantum cases; hence, the limitations of mathematics and of physics are perfectly matched.

However, once equilibrium has been identified with sufficiency, it becomes unclear what to do with regularity and dimensionality, and even with the independence postulated in Section 2.2.

First of all, Dynkin's and Jeffreys' derivations are of a global character. But Dynkin has emphasized the local character of sufficiency, and E. W. Barankin and M. Katz [1] have used this idea in the case where the U_m are not independent. They have shown that the minimal number of invariants, which generalize energy, needs not be independent of the value of the parameter, which generalizes the temperature. Surely, the dependence between U_m is not excluded in physics; quite to the contrary, it is essential in order to generalize thermodynamics.

But one can go further. It seems that, from the viewpoint of the general idea of thermal equilibrium, “such ideas as euclidean space, dimensionality, partial differentiation and the distinction between the continuous and the discrete are all extraneous”. P. R. Halmos and L. J. Savage [5], from whom we borrowed the quotation, then proceed to introduce what amounts to abstract-space generalizations of the concepts of energy and of temperature, and to define sufficiency in that case. Let \mathbf{G} and \mathbf{B} designate the state and the parameter of a generalized canonical system. Halmos and Savage gave one possible meaning to the following statement, which—as written—is made up of undefined terms:

“Consider the parameters that condition the distribution of the states \mathbf{G}_m of a set of systems S_m before and after this set is isolated from contact with some “environment.” “The nature of thermal equilibrium” is such that a necessary and sufficient statistic for the parameters-before-isolation is provided by the parameters-after-isolation.”

Unfortunately, the Halmos-Savage interpretation of the above statement is too general to be of use in thermodynamics. The reason for it will be addressed to the reader familiar with [5], in order to avoid a heavy investment in notation. Briefly, in the H-S case, the conditional distribution of the generalized states \mathbf{G}_m , given the generalized state \mathbf{G} , is defined with the help of the Radon-Nikodym theorem, so that it needs not have the properties of an ordinary probability distribution. However, in physics, it must have all these properties. The difficulty vanishes if one has a euclidean generalization of the state and of the parameter. We know nothing, however, of whether thermodynamics can be generalized on these lines in a way that is not physically void or mathematically trivial.

REFERENCES

- [1] BARANKIN, EDWARD W. and KATZ, MELVIN, JR., (1959). Sufficient statistics of minimal dimension. *Sankhya* **21** 217-246 and **23** 401-404.
- [2] DARMOIS, GEORGES (1935). Sur les lois de probabilité à estimation exhaustive. *C. R. Acad. Sci. Paris* **200** 1265-1266.
- [3] DYNKIN, E. B. (1951). Necessary and sufficient statistics for a family of probability distributions (in Russian). *Uspehi Mat. Nauk N.S.* **6** 68-90; in English: *Selected Transl. Math. Stat. Prob.* (I.M.S.) (1961), 23-40.
- [4] GRAD, HAROLD (1952). Statistical mechanics, thermodynamics and fluid mechanics of systems with an arbitrary number of integrals. *Comm. Pure Appl. Math.* **5** 455-494.
- [5] HALMOS, PAUL R. and SAVAGE, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Statist.* **20** 225-241.
- [6] JEFFREYS, HAROLD (1960). An extension of the Pitman-Koopman theorem. *Proc. Cambridge Philos. Soc.* **56** 393-395; see also: *Theory of Probability*, 3rd Ed., (1962), 168-170.
- [6a] KEYNES, JOHN MAYNARD (1921). *A Treatise on Probability*. Macmillan, New York.
- [7] KHINCHIN, A. I. (1943). *Mathematical Foundations of Statistical mechanics* (in Russian). State Publishing House, Moscow; in English: Dover, New York, 1949.
- [8] KHINCHIN, A. I. (1951). *Mathematical Foundations of Quantum Statistics* (in Russian). State Publishing House, Moscow; in English: Graylock Press, Albany, N. Y., 1960.
- [9] KHINCHIN, A. I. (1955). Symmetric functions on multi-dimensional surfaces (in Russian). *To the memory of Alexander Alexandrovich Andronov*. Moscow; in English: Supplement 6 to [8].
- [10] KOOPMAN, B. O. (1936). On distributions admitting sufficient statistics. *Trans. Amer. Math. Soc.* **39** 399-409.
- [11] LANDAU, L. D. and LIFSHITZ, E. M. (1958). *Statistical Physics*. Addison-Wesley, Reading, Mass.
- [12] MANDELBROT, BENOIT (1956). Exhaustivité de l'énergie totale d'un système en équilibre, pour l'estimation de sa température. *C. R. Acad. Sci. Paris*. **243** (1956), pp. 1835-1838.
- [13] MANDELBROT, BENOIT (1956). An outline of a purely phenomenological theory of statistical thermodynamics: Canonical ensembles. *Inst. Radio Eng. IRE Trans. Information Theory*. **IT-2** 190-203.
- [14] MANDELBROT, BENOIT (1959). Les ensembles grand-canoniques de Gibbs; justification de leur unicité, basée sur la divisibilité infinie de leur énergie aléatoire. *C. R. Acad. Sci. Paris* **249** 1464-1466.
- [15] MANDELBROT, BENOIT (1962). Derivation of statistical thermodynamics from purely phenomenological principles. I.B.M. Research, Yorktown, N. Y.
- [15a] MANDELBROT, BENOIT (1962). A critical note on information theory and statistical mechanics. I.B.M. Research, Yorktown, N. Y.
- [15b] MANDELBROT, BENOIT (196-). Statistical inference in thermodynamics. To appear.
- [16] PITMAN, E. J. G. (1936). Sufficient statistics and intrinsic accuracy. *Proc. Cambridge Philos. Soc.* **32** 567-579.
- [17] POINCARÉ, H. (1896). *Calcul des Probabilités*. Gauthier-Villars, Paris.
- [18] SZILARD, LEO (1925). Über die Ausdehnung der phänomenologischen Thermodynamik auf die Schwankungserscheinungen. *Z. Phys.* **32** 753-788.
- [19] SZILARD, LEO (1929). Über die Entropieverminderung in einem thermodynamischen System bei Eingriff intelligenter Wesen. *Z. Phys.* **53** 840-856.
- [20] WILSON, E. A. (1957). *Thermodynamics and Statistical Mechanics*. Cambridge Univ. Press.