Reprinted from THE PHYSICAL REVIEW, Vol. 108, No. 2, 171–190, October 15, 1957
Printed in U. S. A.

Information Theory and Statistical Mechanics. II

E. T. JAYNES
Department of Physics, Stanford University, California
(Received March 15, 1957)

Treatment of the predictive aspect of statistical mechanics as a form of statistical inference is extended to the density-matrix formalism and applied to a discussion of the relation between irreversibility and information loss. A principle of "statistical complementarity" is pointed out, according to which the empirically verifiable probabilities of statistical mechanics necessarily correspond to incomplete predictions. A preliminary discussion is given of the second law of thermodynamics and of a certain class of irreversible processes, in an approximation equivalent to that of the semiclassical theory of radiation.

It is shown that a density matrix does not in general contain all the information about a system that is relevant for predicting its behavior. In the case of a system perturbed by random fluctuating fields, the density matrix cannot satisfy any differential equation because $\dot{\rho}(t)$ does not depend only on $\rho(t)$, but also on past conditions The rigorous theory involves stochastic equations in the type $\rho(t)=\mathcal{G}(t,0)\rho(0)$, where the operator $\mathcal{G}$ is a functional of conditions during the entire interval $(0\rightarrow t)$. Therefore a general theory of irreversible processes cannot be based on differential rate equations corresponding to time-proportional transition probabilities. However, such equations often represent useful approximations.

INTRODUCTION

IN a previous paper[1] the prediction of equilibrium thermodynamic properties was developed as a form of statistical inference, based on the Shannon[2] concept of entropy as an information measure, and the subjective interpretation of probabilities. The guiding principle is that the probability distribution over microscopic states which has maximum entropy subject to whatever is known, provides the most unbiased representation of our knowledge of the state of the system. The maximum-entropy distribution is the broadest one compatible with the given information; it assigns positive weight to every possibility that is not ruled out by the initial data.

This method of inference is extended in the following sections (numbered consecutively from those of I), to the density-matrix formalism, which makes possible the treatment of time-dependent phenomena. It is then applied to a discussion of the relation of information loss and irreversibility, and to a treatment of relaxation processes in an approximation equivalent to that of the semiclassical theory of radiation. The more rigorous treatment, corresponding to quantum electrodynamics, will be taken up in a later paper.

Our picture of a prediction process is as follows. At the initial time $t=0$ certain measurements are made. In practice, these will always represent far less than the maximum observation which would enable us to determine a definite pure state. Therefore, we must have recourse to maximum-entropy inference in order to represent our degree of knowledge about the system in a way free of arbitrary assumptions with regard to missing information.[3] As time goes on, each state of the maximum-entropy distribution changes due to perturbations that are in general unknown; thus it "spreads out" into several possibilities, and our initial knowledge as to the state of the system is gradually lost. In the "semiclassical" approximation considered here, the final state of affairs is usually one in which the initial information is completely lost, the density matrix relaxing into a multiple of the unit matrix. The prediction of thermal equilibrium, in which the limiting form of the density matrix is that of the Boltzmann distribution with finite temperature, is found only by using a better approximation which takes into account the quantum nature of the surroundings.

It is of the greatest importance to recognize that in all of this semiclassical theory it is possible to maintain the view that the system is at all times in some definite but unknown pure state, which changes because of definite but unknown external forces; the probabilities represent only our ignorance as to the true state. With such an interpretation the expression "irreversible process" represents a semantic confusion; it is not the physical process that is irreversible, but rather our ability to follow it. The second law of thermodynamics then becomes merely the statement that although our information as to the state of a system may be lost in a variety of ways, the only way in which it can be gained is by carrying out further measurements. Essential for this is the fact, analogous to Liouville's theorem, that in semiclassical approximation the laws of physics do not provide any tendency for systems initially in different states to "accumulate" in certain final states in preference to others; i.e., the time-development matrix is unitary.

In opposition to the foregoing views, one may assert

[1] E. T. Jaynes, Phys. Rev. **106**, 620 (1957). Hereinafter referred to as I.

[2] C. E. Shannon, Bell System Tech. J. **27**, 379, 623 (1948). These papers are reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[3] A very interesting quotation from J. W. Gibbs [*Collected Works* (Longmans, Green and Company, New York, 1928), Vol. II, p. 180] suggests the same basic idea. In discussing the interaction of a body and a heat-bath, he says "The series of phases through which the whole system runs in the course of time may not be entirely determined by the energy, but may depend on the initial phase in other respects. In such cases the ensemble obtained by the microcanonical distribution of the whole system, which includes all possible time-ensembles combined in the proportion which seems least arbitrary, will better represent than any one time-ensemble the effect of the bath."

that irreversibility is not merely a loss of human information; it is an experimental fact, well recognized long before the development of statistical mechanics. Furthermore, the relaxation times calculated below are not merely measures of the rate at which we lose information; they are experimentally measurable quantities expressing the rate at which physical systems approach equilibrium. Therefore, the probabilities involved in our calculations must be ascribed some objective meaning independent of human knowledge.

Objections of this type have already been answered in large part in I, particularly Sec. 4. However, we wish to indicate briefly how those arguments apply to the case of time-dependent phenomena. The essential fact is again the "principle of macroscopic uniformity." In the first place, it has been shown that the only quantities for which maximum-entropy inference makes definite predictions are those for which we obtain sharp probability distributions. Since maximum-entropy inference uses the broadest distribution compatible with the initial data, the predictable properties must be characteristic of the great majority of those states to which appreciable weight is assigned. Maximum-entropy inference can never lead us astray, for any quantity which it is incapable of predicting will betray that fact by yielding a broad probability distribution.

We can, however, say much more than this. We take it as self-evident that the features of irreversible processes which are experimentally reproducible are precisely those characteristic of most of the states compatible with the conditions of the experiment. Suppose that maximum-entropy inference based on knowledge of the experimentally imposed conditions makes a definite prediction of some phenomenon, and it is found experimentally that no such phenomenon exists. Then the predicted property is characteristic of most of the states appearing in the subjective maximum-entropy distribution, but it is not characteristic of most of the states physically allowed by the experimental conditions. Consider, on the other hand, the possibility that a phenomenon might be found which is experimentally reproducible but *not* predictable by maximum-entropy inference. This phenomenon must be characteristic of most of the states allowed by the experimental conditions, but it is not characteristic of most of the states in the maximum-entropy distribution. In either case, there must exist new physical states, or new *constraints* on the physically accessible states, not contained in the presently known laws of physics.

In summary, we assert that *if it can be shown that the class of phenomena predictable by maximum-entropy inference differs in any way from the class of experimentally reproducible phenomena, that fact would demonstrate the existence of new laws of physics, not presently known.* Assuming that this occurs, and the new laws of physics are eventually worked out, then maximum-entropy inference based on the new laws will again have this property.

From this we see that adoption of subjective probabilities in no way weakens the theory in its ability to give reliable and useful results. On the contrary, the full power of statistical mechanics cannot be seen until one makes this distinction between its subjective and objective aspects. Once this is done, its mathematical rules become a methodology for a very general type of scientific reasoning.

7. REPRESENTATION OF A QUANTUM-MECHANICAL SYSTEM

We now develop a method of representing any state of knowledge of a quantum-mechanical system, leaving aside for the moment any consideration of how this knowledge might have been obtained. Suppose that on the basis of the information available we conclude that the system may be in the "pure state" ψ_1 with probability w_1, or it may be in the state ψ_2 with probability w_2, etc. The various alternative possibilities ψ_i are not necessarily mutually orthogonal, but each may be expanded in terms of a complete orthonormal set of functions u_k:

$$\psi_i = \sum_k u_k a_{ki}. \tag{7.1}$$

This state of knowledge may be visualized in a geometrical fashion by considering a complex function space, whose dimensionality may be finite or infinite, in which the state ψ_i is represented by a point P_i with coordinates a_{ki}, $k=1, 2, \cdots$. At P_i, place a weight w_i; thus the state of knowledge is described by a set (which may be discrete or continuous) of weighted points; such a set will be called an *array*. Since each of the possible wave functions is normalized to unity,

$$(\psi_i,\psi_i) = \int |\psi_i|^2 d\tau = 1,$$

we have

$$\sum_k |a_{ki}|^2 = 1, \tag{7.2}$$

and all points P_i are at unit "distance" from the origin, on the surface of the unit hypersphere.

If each of the possible states ψ_i satisfies the same Schrödinger equation,

$$i\hbar\dot{\psi} = H\psi,$$

then as time goes on the function space as a whole is subjected to a unitary transformation, so that all "distances" and scalar products

$$(\psi_i,\psi_j) = \int \psi_i^* \psi_j d\tau$$

remain invariant, and the entire motion of the array may be visualized as a "rigid rotation" of the hypersphere. An array with this behavior will be called *simple*. A simple array is conceptually somewhat like a microcanonical ensemble; it consists of points lying on a closed surface which are subjected, in consequence of

the equations of motion, to a measure-preserving transformation which continually unfolds as t increases.

The transformation with time may be of a different type; much more interesting is the case where the initial information is of the form: "The system may be in state ψ_i with probability w_i, and in this case the Hamiltonian will be H_i." Then different parts of the array are subjected to different rotations, and separations or interpenetrations occur. Such an array will be called *compound*. It arises, for example, when we have a system consisting of two coupled spins in a strong magnetic field, and we wish to describe our knowledge of the state of one of them.

Consider a measurable quantity represented by a Hermitian operator F; in state ψ_i its expectation value is

$$\langle F\rangle_i=(\psi_i,F\psi_i)=\sum_{kn} a_{ki}a_{ni}{}^*F_{nk}, \tag{7.3}$$

where $F_{nk}=(u_n,Fu_k)$ are the matrix elements of F in the u_k representation. The average of (7.3) over the array is

$$\langle F\rangle=\sum_i w_i\langle F\rangle_i=\mathrm{Tr}(\rho F), \tag{7.4}$$

where

$$\rho_{kn}=\sum_i w_i a_{ki}a_{ni}{}^*=\langle a_k a_n{}^*\rangle \tag{7.5}$$

is the density matrix. The probability $p(f)$ that a measurement of F will yield the particular eigenvalue f, is also expressible as an expectation value; define the projection operator O by $O\psi=(\varphi,\psi)\varphi$, where φ is the corresponding normalized eigenfunction of F: $F\varphi=f\varphi$. Then

$$p(f)=\langle O\rangle=\mathrm{Tr}(\rho O). \tag{7.6}$$

From (7.5) it is seen that in general an infinite number of different arrays, representing different mixtures of pure states, all lead to the same density matrix. The most general discrete array which leads to a given density matrix ρ corresponds to the most general matrix A (not necessarily square) for which

$$\rho=AA^\dagger, \tag{7.7}$$

the dagger denoting the Hermitian conjugate. An array is uniquely determined by A, for from (7.2) and (7.5) we have

$$A_{ki}=a_{ki}w_i{}^{\frac{1}{2}},\quad \sum_k|A_{ki}|^2=w_i.$$

To find another array with the same density matrix, insert a matrix U:

$$\rho=(AU)(U^{-1}A^\dagger).$$

This has the form $BB^\dagger$ with $B=AU$ if and only if U is unitary; thus the group of transformations from one array of n states to another of n states is isomorphic with the group of unitary transformations in n dimensions. These are not, however, transformations of the wave functions ψ_i, but of the *probability-normalized* wave functions

$$\Psi_i=\psi_i w_i{}^{\frac{1}{2}}. \tag{7.8}$$

If we carry out the unitary transformation

$$\Phi_j=\sum_i \Psi_i U_{ij}, \tag{7.9}$$

and write

$$\Phi_j=\varphi_j p_j{}^{\frac{1}{2}},$$

where φ_j is normalized to unity, then the array in which state φ_j has probability p_j leads to the same density matrix as the original array $\{\psi_i,w_i\}$. Evidently an array is determined uniquely by specifying a set $\{\Psi_i\}$ of probability-normalized states.

From an array $\{\Psi_i\}$ of n states we can construct new arrays of $(n+1)$ states. Define $\Psi_{n+1}\equiv 0$; then new transformations of the form (7.9) are possible, in which U is a unitary matrix of dimensionality $(n+1)$. These generate an infinite number of new arrays for which, in general, all $(n+1)$ states Φ_j are different from each other and from zero. The inverse process of *contracting* an array to one of fewer states is possible if any linear combination of the ψ_i vanishes.

An array of n states will be called *minimal* with respect to its density matrix ρ if no array exists which leads to ρ with fewer than n states. The states of an array are linearly independent if and only if the array is minimal.

In general, a given density matrix can be represented in only one way as a mixture of orthogonal states. Since ρ is Hermitian, there always exists a unitary matrix U which diagonalizes it;

$$d=U\rho U^{-1}, \tag{7.10}$$

with $d_{mn}=d_m\delta_{mn}$. If the eigenvalues d_m of ρ are nondegenerate, only one such matrix U exists. The basis functions of the new representation in which ρ is diagonal,

$$v_m=\sum_k u_k U_{km}{}^{-1}, \tag{7.11}$$

are the orthogonal states which, when mixed with probabilities d_m, lead to the given density matrix.

Suppose we have a density matrix ρ and a state φ which is considered a "candidate" for inclusion in a minimal array which will lead to ρ. What is the probability $p_A(\varphi)$ which should be assigned to φ in such an array? To answer this, we first construct the orthogonal array $\{v_m,d_m\}$, and expand

$$\varphi=\sum_m v_m C_m.$$

If this is to be equivalent to one of the columns of (7.9), it is necessary that

$$\frac{1}{p_A}=\sum_m\frac{|C_m|^2}{d_m}. \tag{7.12}$$

This is uniquely determined by the density matrix and the state φ, regardless of which other states φ_j might also appear in the array. The array probability p_A is in general different from the measurement probability (7.6), which is equal to

$$p_M(\varphi)=\sum_m d_m|C_m|^2. \tag{7.13}$$

It is readily shown that $p_M \geqslant p_A$, with equality if and only if φ is an eigenstate of ρ.

The representation in terms of orthogonal states is important in connection with the entropy which measures our knowledge of the system. It might be thought that for an array $\{\psi_i,w_i\}$ we could define an entropy by

$$S_A = -\sum_i w_i \ln w_i. \tag{7.14}$$

This, however, would not be satisfactory because the w_i are not in general the probabilities of mutually exclusive events. According to quantum mechanics, if the state is known to be ψ_i, then the probability of finding it upon measurement to be ψ_j, is $|(\psi_j,\psi_i)|^2$. Thus, the probabilities w_i refer to independent, mutually exclusive events only when the states ψ_i of the array are orthogonal to each other, and only in this case is the expression (7.14) for entropy satisfactory. This array of orthogonal states has another important property; consider the totality of all possible arrays which lead to a given density matrix, and the corresponding expressions (7.14). The array for which (7.14) attains its minimum value is the orthogonal one, which therefore provides, in the sense of information content, the most economical description of the freedom of choice implied by a density matrix (Appendix A).

For the orthogonal array, the w_i in (7.14) are identical with the eigenvalues d_i of the density matrix, so for numerical calculation of entropy given ρ, one would find the eigenvalues and use the formula

$$S = -\sum_i d_i \ln d_i. \tag{7.15}$$

In general discussions it is convenient to express this

$$S = -\mathrm{Tr}(\rho \ln \rho). \tag{7.16}$$

Since this could also be written as $S = -\langle \ln \rho \rangle$, it is the natural extension to quantum mechanics of the Gibbs definition of entropy.

Equation (7.16) assigns zero entropy to any pure state, whether stationary or not. It has been criticized on the grounds that according to the Schrödinger equation of motion it would be constant in time, and thus one could not account for the second law of thermodynamics; this has led some authors[4,5] to propose instead the expression

$$S = -\sum_k \rho_{kk} \ln \rho_{kk}, \tag{7.17}$$

which involves only diagonal elements of ρ in the energy representation, for which a "quantum-mechanical spreading" phenomenon can be demonstrated. It will be shown in detail below how the objections to (7.16) may be answered. With regard to (7.17), we note that it does not assign the same entropy to all pure states; but von Neumann[6] has shown that any pure state may be converted reversibly and adiabatically into any other pure state.

Since, according to (7.4), knowledge of ρ enables one to calculate the expectation value of any Hermitian operator, it is tempting to conclude that the density matrix contains all of our information as to the objective state of the system. Thus, although many different arrays would all lead to the same density matrix, the differences between them would be considered physically meaningless, only their second moments (7.5) corresponding to any physical predictions. The concept of any array as something separate and distinct from a density matrix might then appear superfluous. That this is not the case, however, will be seen in Sec. 13 below, where it is shown that the resolution of a compound array into independent simple arrays may represent useful information which cannot be expressed in terms of the resultant density matrix.

8. SUFFICIENCY AND COMPLETENESS OF THE DENSITY MATRIX

If a density matrix provides a definite probability assignment for each possible outcome of a certain experiment, in a way that makes full use of all of the available relevant information, we shall say that ρ is *sufficient* for that experiment. A density matrix that is sufficient for all conceivable experiments on a system will be called *complete* for that system. Strictly speaking, we should always describe a density matrix as sufficient or complete *relative* to certain initial information.

The assertion that complete density matrices exist involves several assumptions, in particular that all measurable quantities may be represented by Hermitian operators, and that all experimental measurements may be expressed in terms of expectation values. We do not wish to go into these questions, but only to note the following. Even if it be granted that it is always possible in principle to operate with a complete density matrix, it would often be extremely awkward and inconvenient to do so in practice, because it would require us to consider the density matrix and dynamical quantities as operators in a much larger function space than we wish to use.

To see this by a simple example, consider a "molecular beam" experiment in which particles of spin $\frac{1}{2}$ are prepared by apparatus A, then sent into a detection system B which determines whether the spin is up or down with respect to some chosen z axis. Assume, for simplicity, that only one particle at a time is processed in this way. A particle thus has, for our purposes, two possible states u_+ and u_-; our knowledge of the nature of the apparatus A could be incorporated into an array and its corresponding (2×2) density matrix, from which we can calculate the probability of finding the spin aligned in any particular direction. Thus, the (2×2) density matrix adequately represents our state of knowledge as to the outcome of any spin measurement made on a single particle; i.e., it is a sufficient

[4] R. C. Tolman, *The Principles of Statistical Mechanics* (Clarendon Press, Oxford, 1938).

[5] D. ter Haar, *Elements of Statistical Mechanics* (Rinehart and Company, Inc., New York, 1954).

[6] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Dover Publications, New York, 1943), Chap. V.

statistic for any such measurement. The question is, does it also adequately represent our knowledge of the *ensemble* of particles (assuming that the apparatus A is "stationary," so that each particle, considered by itself, would be represented by the same density matrix). More specifically, is it possible for apparatus A to produce a physical situation which can be measured in our detection apparatus, but for which the (2×2) density matrix gives no probability assignment? One such property is easily found; the detecting apparatus tells us not only the fraction of spins aligned along the $+z$ axis, but also the *order* in which spin up and spin down occurred, so that correlations between spin states of successive particles can be observed. Now all possible such correlations can be described only by considering the entire ensemble of N particles as a single quantum-mechanical system with 2^N possible states, and therefore a density matrix which is a sufficient statistic for all conceivable measurements on the spin system must have 2^N rows and columns.[7] This, however, would completely destroy the simplicity of the theory, and in practice we would probably prefer to retain the original (2×2) density matrix for predicting the results of measurements on single particles, while recognizing its insufficiency for other measurements which the same apparatus could perform.

9. SUBJECTIVE AND OBJECTIVE INTERPRETATIONS

The topic of Sec. 8 is closely related to some of the most fundamental questions in physics. According to quantum mechanics, if a system is known to be in state ψ_i, then the probability that measurement of the quantity F will result in the particular eigenvalue f, is $\langle O\rangle_i$, where O is the projection operator of Eq. (7.6). Are we to interpret this probability in the objective or subjective sense; i.e., are the probability statements of quantum mechanics expressions of empirically verifiable laws of physics or merely expressions of our incomplete ability to predict, whether due to a defect in the theory or to incomplete initial information? The current interpretation of quantum mechanics favors the first view, but it is important to note that the whole content of the theory depends critically on just what we mean by "probability." In calling a probability objective, we do not mean that it is necessarily "correct," but only that a conceivable experiment exists by which its correctness or incorrectness could be empirically determined. In calling a probability assignment subjective, we mean that it is not a physical property of any system, but only a means of describing our information about the system; therefore it is meaningless to speak of verifying it empirically.

Is there any operational meaning to the statement that the probabilities of quantum mechanics are objective? If so, we should be able to devise an experiment which will measure these probabilities, for example the probability that a measurement of the quantity F will give the result f. In order to do this, we will need to repeat a measurement of F an indefinitely large number N of times, with systems that have all been prepared in exactly the same way, and record the fraction of cases in which the particular result f was obtained. Which density matrix should we use to predict the result of this experiment? In principle, we should always use the one which contains the greatest amount of information about the ensemble of N systems; i.e., which is complete. The apparatus which prepares them may be producing correlations; thus the ensemble of N systems should be considered as a single large quantum-mechanical system. The probability statements which come from the theory are then of the form, "the probability that system 1 will yield the result f_1, *and* system 2 will yield the value f_2, $\cdots$, is $p(f_1\cdots f_N)$." But then measurement of F in each of the N small systems is not N repetitions of an experiment; it is only a single experiment from the standpoint of the total system. Clearly, no probability assignment can be verified by a single measurement. Note that the question whether correlations were in fact present between different systems is irrelevant to the question of principle involved; even if the distribution factors

$$p(f_1\cdots f_N)=p_1(f_1)p_2(f_2)\cdots p_N(f_N) \qquad (9.1)$$

it remains a joint distribution, not one for a single system. We can, of course, always obtain the single-system probabilities by summation:

$$p_1(f_1)=\sum_{f_2}\cdots\sum_{f_N}p(f_1f_2\cdots f_N), \qquad (9.2)$$

but $p_1(f_1)$ now refers specifically to system 1, and the results of measurements on the other systems are irrelevant to the question whether $p_1(f_1)$ was verified. We cannot avoid the difficulty by repeating all this M times, because for that experiment the complete density matrix would refer to all NM systems, and we would be in exactly the same situation. Thus, the probability statements obtained from a complete density matrix cannot be verified.

In practice, of course, one will never bother with such considerations, but will find a density matrix which operates only on the space of a single system and incorporates as much information as possible subject to that limitation. The probability $p(f)$ computed from this density matrix is presumably equal to $p_1(f)$ in (9.2). If the result f is obtained approximately $Np(f)$ times, one says that the predictions have been verified, and $p(f)$ is correct in an objective sense. This result is obtained, however, only by renouncing the possibility of predicting any mutual properties of different systems, and the record of the experiment contains some information about those mutual properties.

[7] This is a very conservative statement. It would be more realistic to assume that all the coordinates of apparatus A must also be included in the space upon which this complete density matrix operates.

Thus, we enunciate as a general principle: *Empirical verifiability of a probability assignment, and completeness of the density matrix from which the probabilities were obtained, are incompatible conditions.* Whenever we use a density matrix whose probabilities are verifiable by certain measurements, we necessarily renounce the possibility of predicting the results of other measurements which can be made on the same apparatus.

This principle of "statistical complementarity" is not restricted to quantum mechanics, but holds in any application of probability theory; in a very fundamental sense no experiment can ever be repeated, and the most comprehensive probability assignments are necessarily incapable of verification.

If an operational viewpoint[8-10] is to be upheld consistently, it appears that the probabilities computed from a complete density matrix must be interpreted in the subjective sense. Since this complete density matrix might be a projection operator corresponding to a pure state, one is led very close to the views of Einstein[11] and Bohm[12] as to the interpretation of quantum mechanics.

Entirely different considerations suggest the same conclusion. A density matrix represents a fusion of two different statistical aspects; those inherent in a pure state and those representing our uncertainty as to which pure state is present. If the former probabilities are interpreted in the objective sense, while the latter are clearly subjective, we have a very puzzling situation. Many different arrays, representing different combinations of subjective and objective aspects, all lead to the same density matrix, and thus to the same predictions. However, if the statement, "only certain specific aspects of the probabilities are objective," is to have any operational meaning, we must demand that some experiment be produced which will distinguish between these arrays.

10. MAXIMUM-ENTROPY INFERENCE

The methods of maximum-entropy inference described in I may be generalized immediately to the density-matrix formalism. Suppose we are given the expectation values of the operators $F_1 \cdots F_m$; then the density matrix which represents the most unbiased picture of the state of the system on the basis of this much information is the one which maximizes the entropy subject to these constraints. As before, this is accomplished by finding the density matrix which unconditionally maximizes

$$S-\lambda_1\langle F_1\rangle-\cdots-\lambda_m\langle F_m\rangle, \tag{10.1}$$

in which the λ_i are Lagrangian multipliers. The result may be described in terms of the partition function

$$Z(\lambda_1\cdots\lambda_m)=\mathrm{Tr}[\exp(-\lambda_1F_1-\cdots-\lambda_mF_m)], \tag{10.2}$$

with the λ_k determined by

$$\langle F_k\rangle=-\frac{\partial}{\partial\lambda_k}\ln Z. \tag{10.3}$$

The maximum-entropy density matrix is then

$$\rho=\exp[-\lambda_0 1-\lambda_1F_1-\cdots-\lambda_mF_m] \tag{10.4}$$

which is correctly normalized ($\mathrm{Tr}\rho=1$) by setting

$$\lambda_0=\ln Z, \tag{10.5}$$

and the corresponding entropy becomes

$$S=\lambda_0+\lambda_1\langle F_1\rangle+\cdots+\lambda_m\langle F_m\rangle. \tag{10.6}$$

Use of (10.5) and (10.6) enables us to solve (10.3) for the λ_k:

$$\lambda_k=\partial S/\partial\langle F_k\rangle. \tag{10.7}$$

If the operator F_k contains parameters α_i, we find as before that the maximum-entropy estimates of the derivatives are given by

$$\left\langle\frac{\partial F_k}{\partial\alpha_i}\right\rangle=-\frac{1}{\lambda_k}\frac{\partial}{\partial\alpha_i}\ln Z. \tag{10.8}$$

For an infinitesimal change in the problem, λ_k is the integrating factor for the kth analog of infinitesimal heat;

$$\delta S=\textstyle\sum_k\lambda_k\delta Q_k, \tag{10.9}$$

with

$$\delta Q_k=\delta\langle F_k\rangle-\langle\delta F_k\rangle. \tag{10.10}$$

All of these relations except (10.2) and (10.4) are formally identical with those found in I, the F_k now being interpreted as matrices instead of functions of a discrete variable x.

The definitions of heat bath and thermometer given in I remain applicable, and the discussion of experimental measurement of temperature proceeds as before with the difference that maximization of entropy of the combined system now automatically takes care of the question of phase relations. We have two systems σ_1 and σ_2, with complete orthonormal basis functions $u_n(1)$, $v_k(2)$, respectively. A state ψ_i of the combined system $\sigma=\sigma_1\times\sigma_2$ is then some linear combination

$$\psi_i(1,2)=\sum_{nk}u_n(1)v_k(2)a_{nki}.$$

If ψ_i occurs with probability w_i, the density matrix is

$$(nk|\rho|n'k')=\textstyle\sum_i w_ia_{nki}a_{n'k'i}{}^*=\langle a_{nk}a_{n'k'}{}^*\rangle.$$

[8] P. W. Bridgman, *The Logic of Modern Physics* (The MacMillan Company, New York, 1927).

[9] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1935), second edition, Chap. I.

[10] Hans Reichenbach, *Philosophic Foundations of Quantum Mechanics* (University of California Press, Berkeley, 1946).

[11] *Albert Einstein Philosopher-Scientist*, edited by P. A. Schilpp (Library of Living Philosophers, Inc., Evanston, 1949), pp. 665–684.

[12] D. Bohm, Phys. Rev. **85**, 166, 180 (1952); **89**, 458 (1953).

An operator $\mathcal{G}(1,2)$ has matrix elements

$$(nk|\mathcal{G}|n'k')=\int\int u_n^*(1)v_k^*(2)\mathcal{G}(1,2)u_{n'}(1)v_{k'}(2)d\tau_1 d\tau_2$$

and its expectation value is

$$\langle\mathcal{G}\rangle=\mathrm{Tr}(\rho\mathcal{G})=\sum_{nn'kk'}(nk|\rho|n'k')(n'k'|\mathcal{G}|nk).$$

An operator F_1 which operates only on the coordinates of system 1 is represented in the space of the combined system by a direct product matrix,[13] $\mathfrak{F}_1=F_1\times 1$, with matrix elements

$$(nk|\mathfrak{F}_1|n'k')=(n|F_1|n')\delta_{kk'}.$$

Similarly, for an operator F_2 of system 2, we obtain $\mathfrak{F}_2=1\times F_2$, and

$$(nk|\mathfrak{F}_2|n'k')=\delta_{nn'}(k|F_2|k').$$

Consider, as before, the system of interest σ_1, and a thermometer σ_2. Let their Hamiltonians be H_1, H_2, respectively. In the function space of the combined system σ, these Hamiltonians are represented by

$$\mathfrak{H}_1=H_1\times 1,\quad \mathfrak{H}_2=1\times H_2. \tag{10.11}$$

The available information now consists of a given (measured) value of $\langle H_2\rangle$, and the knowledge that energy may be transferred between σ_1 and σ_2 in such a way that the total amount is conserved. In practice we never have detailed knowledge of the weak-interaction Hamiltonian $\mathfrak{H}_{12}$ of a type that would tell us which transitions may in fact take place and which will not. Therefore we have no rational basis for excluding the possibility of any transition between states of σ with a given total energy, and the most unbiased representation of our state of knowledge must treat all such states as equivalent, in their dependence on energy. Any other procedure would amount to arbitrarily favoring some states at the expense of others, in a way not warranted by any of the available information. Therefore only the total energy may appear in our density matrix, and we have to find that matrix which maximizes

$$S-\lambda\langle\mathfrak{H}_1+\mathfrak{H}_2\rangle, \tag{10.12}$$

subject to the observed value of $\langle H_2\rangle$. The matrix involved in (10.2) and (10.4) now factors into a direct product:

$$\exp[-\lambda(\mathfrak{H}_1+\mathfrak{H}_2)]=(e^{-\lambda H_1})\times(e^{-\lambda H_2}), \tag{10.13}$$

so that the partition function reduces to

$$Z(\lambda)=Z_1(\lambda)Z_2(\lambda), \tag{10.14}$$

with

$$\begin{aligned}Z_1(\lambda)&=\mathrm{Tr}\exp(-\lambda H_1),\\ Z_2(\lambda)&=\mathrm{Tr}\exp(-\lambda H_2).\end{aligned} \tag{10.15}$$

[13] P. R. Halmos, *Finite Dimensional Vector Spaces* (Princeton University Press, Princeton, 1948), Appendix II.

Similarly, the density matrix (10.4) is the direct product

$$\rho=\left[\frac{\exp(-\lambda H_1)}{Z_1(\lambda)}\right]\times\left[\frac{\exp(-\lambda H_2)}{Z_2(\lambda)}\right]=\rho_1\times\rho_2. \tag{10.16}$$

Because of the absence of correlations between the two systems, it is true once again that the function of the thermometer is merely to tell us the value of the parameter λ in ρ_1, and the properties of the thermometer need not be considered in detail when incorporating temperature measurements into our theory.

An important feature of this theory is that measurement of averages of several noncommuting quantities may be treated simultaneously without interference. Consider, for example, three interacting systems $\sigma=\sigma_1\times\sigma_2\times\sigma_3$, where σ_1 is the system of interest, and σ_2 is a thermometer. Some physical quantity F, represented in the space of σ_1 by the operator F_1, and in σ_3 by F_3, can be transferred between σ_1 and σ_3 in such a way that the total amount is conserved. F_1 could stand for angular momentum, volume, etc., and need not commute with H_1. In addition suppose that a quantity $\langle G_1\rangle$ is measured directly in σ_1, where G_1 does not necessarily commute with either H_1 or F_1. Now the available information consists of the measured values of $\langle G_1\rangle$, $\langle H_2\rangle$, and $\langle F_3\rangle$, plus the conservation laws of F and H. The various operators are now represented in the space σ by direct product matrices as follows:

$$\begin{aligned}&\mathfrak{H}_1=H_1\times 1\times 1,\quad \mathfrak{F}_1=F_1\times 1\times 1,\\ &\mathfrak{H}_2=1\times H_2\times 1,\quad \mathfrak{F}_3=1\times 1\times F_3,\\ &\mathfrak{G}_1=G_1\times 1\times 1,\end{aligned}$$

and the density matrix that provides the most unbiased picture of the state of the total system is the one that maximizes

$$S-\lambda\langle\mathfrak{H}_1+\mathfrak{H}_2\rangle-\mu\langle\mathfrak{F}_1+\mathfrak{F}_3\rangle-\nu\langle\mathfrak{G}_1\rangle. \tag{10.17}$$

We now find the factorization property

$$\begin{aligned}\exp[&-\lambda(\mathfrak{H}_1+\mathfrak{H}_2)-\mu(\mathfrak{F}_1+\mathfrak{F}_3)-\nu\mathfrak{G}_1]\\ &=[e^{-\lambda H_1-\mu F_1-\nu G_1}]\times[e^{-\lambda H_2}]\times[e^{-\mu F_3}],\end{aligned} \tag{10.18}$$

so that once again the partition function and density matrix factor into independent parts for the three systems:

$$Z(\lambda,\mu,\nu)=Z_1(\lambda,\mu,\nu)Z_2(\lambda)Z_3(\mu),\quad \rho=\rho_1\times\rho_2\times\rho_3,$$

and the pieces of information obtained from σ_2, σ_3 are transferred into ρ_1 without interference.

11. INFORMATION LOSS AND IRREVERSIBILITY

In classical statistical mechanics the appearance of irreversibility can always be traced either to the replacement of a fine-grained probability distribution by a coarse-grained one, or to a projection of a joint probability distribution of two systems onto the subspace of one of them. Both processes amount to a loss,

whether voluntary or not, of some of the information which is in principle available. The former is often justified by the very persuasive argument that the mathematics would otherwise be too complicated. But mathematical difficulties, however great, have no bearing on matters of principle, and this way of looking at it causes one to lose sight of a much more important positive reason for discarding information. After sufficient "stirring" has occurred, two different fine-grained distributions will lead to predictions that are macroscopically the same, differing only in microscopic details. Thus, even if we were good enough mathematicians to deal with a fine-grained distribution, its replacement by a coarse-grained one would still be the elegant method of treating the prediction of macroscopic properties, because in this way one eliminates irrelevant details at an early stage of the calculation.

In quantum mechanics, as in classical theory, the increase in entropy characteristic of irreversibility always signifies, and is identical with, a loss of information. It is important to realize that the tendency of entropy to increase is not a consequence of the laws of physics as such, for the motion of points of an array is a unitary transformation prescribed by the Schrödinger equation in a manner just as "deterministic" as is the motion of phase points in classical theory. An entropy increase may occur unavoidably, due to our incomplete knowledge of the forces acting on a system, or it may be an entirely voluntary act on our part. In the latter case, an entropy increase is the means by which we simplify a prediction problem by discarding parts of the available information which are irrelevant, or nearly so, for the particular predictions desired. It is very similar to the statistician's practice of "finding a sufficient statistic." The price we must pay for this simplification is that the possibility of predicting other properties with the resulting equations is thereby lost.

The natural way of classifying theories of irreversible processes is according to the mechanism by which information is lost or discarded. In most of the existing theories we find that this consists of the repetition, at regular intervals, of one of the following procedures. Suppose we wish to find the expectation value of the quantity F; in the representation in which F is diagonal it reduces to

$$\langle F\rangle = \mathrm{Tr}(\rho F) = \sum_n \rho_{nn} F_{nn}. \tag{11.1}$$

Since only the diagonal elements of ρ contribute, $\langle F\rangle$ can be calculated as well by using the density matrix ρ', where

$$\rho_{nk}{}' = \rho_{nn}\delta_{nk}. \tag{11.2}$$

The process of replacing ρ by ρ' will be called *removing coherences*, and is clearly permissible whenever all the quantities which we wish to calculate are diagonal simultaneously. It is readily verified that removal of coherences represents loss of information: $S(\rho')\geq S(\rho)$, with equality if and only if $\rho=\rho'$.

The second procedure by which information may be discarded is an invariant operation, exactly analogous to its classical counterpart. Consider two interacting systems σ_1 and σ_2. As already noted, an operator F_1 which operates only on the variables of σ_1 is represented in the space of the combined system $\sigma=\sigma_1\times\sigma_2$ by the direct product matrix $\mathfrak{F}_1=F_1\times 1$. The expectation value of any such operator reduces to a trace involving only the space of σ_1:

$$\langle F_1\rangle = \mathrm{Tr}(\rho\,\mathfrak{F}_1) = \mathrm{Tr}(\rho_1 F_1), \tag{11.3}$$

where ρ_1 is the "projection" of the complete density matrix ρ onto the subspace σ_1, with matrix elements

$$(n|\rho_1|n') = \sum_k (nk|\rho|n'k). \tag{11.4}$$

Similarly, we can project ρ onto σ_2, with the result

$$(k|\rho_2|k') = \sum_n (nk|\rho|nk')$$

and for any operator F_2 of system 2 we can define $\mathfrak{F}_2=1\times F_2$, whereupon $\langle F_2\rangle=\mathrm{Tr}(\rho\mathfrak{F}_2)=\mathrm{Tr}(\rho_2 F_2)$.

In the projection onto σ_1, the parts of ρ that are summed out contain information about the state of system σ_2 and about correlations between possible states of σ_1 and σ_2, both of which are irrelevant for predicting the average of F_1.

The operation of *removing correlations* consists of replacing ρ by the direct product $\rho_1\times\rho_2$, with matrix elements

$$(nk|\rho_1\times\rho_2|n'k') = (n|\rho_1|n')(k|\rho_2|k'), \tag{11.5}$$

and the expectation value of any operator composed additively of terms which operate on σ_1 alone or on σ_2 alone, is found as well from $(\rho_1\times\rho_2)$ as from ρ. The removal of correlations also involves a loss of information; the entropy after removal of correlations is additive and never less than the original entropy:

$$S(\rho_1\times\rho_2) = S(\rho_1)+S(\rho_2)\geq S(\rho), \tag{11.6}$$

with equality if and only if $\rho=\rho_1\times\rho_2$.

These remarks generalize in an obvious way to the case of any number of subsystems; to remove correlations from a density matrix ρ operating on the space of three systems $\sigma=\sigma_1\times\sigma_2\times\sigma_3$, project it onto each of the σ_i, and replace ρ by the direct product of the projections:

$$\rho\rightarrow\rho_1\times\rho_2\times\rho_3.$$

If an operator F_2 operates only on the space of σ_2, its matrix representation in the σ space and expectation value are given by

$$\mathfrak{F}=1\times F_2\times 1, \quad \langle F_2\rangle=\mathrm{Tr}(\rho\mathfrak{F})=\mathrm{Tr}(\rho_2 F_2).$$

Most treatments of irreversible processes in the past have been based on the removal of coherences in the energy representation, and the resulting concept of "occupation numbers" N_k, proportional to the diagonal elements ρ_{kk} in this representation. One then introduces a transition probability per unit time λ_{kn}, which usually,

but not always,[14,15] conforms to the assumption of "microscopic reversibility" $\lambda_{kn}=\lambda_{nk}$, and equations of the form

$$dN_k/dt=\sum_m(\lambda_{km}N_m-\lambda_{mk}N_k) \tag{11.7}$$

are the starting point of the theory. The existence of time-proportional transition probabilities is not, however, a general consequence of quantum mechanics, but involves assumptions about the type of perturbing forces responsible for the transitions, and mathematical approximations which represent a loss of information. That information is lost somewhere is seen from the fact the entropy, as calculated from (11.7), is in general an increasing function of the time, while that obtained from rigorous integration of a Schrödinger equation is necessarily constant. The nature of the information-discarding process in (11.7), as well as a clear statement of the type of physical problems to which equations of this form are applicable, can be appreciated only by starting from a more fundamental viewpoint.

12. SUBJECTIVE H THEOREM

In the remainder of this paper, we consider a certain approximation, which might be called the "semiclassical theory of irreversible processes," since it is related to a complete theory in the same way that the semiclassical theory of radiation[16] is related to quantum electrodynamics. The system of interest σ is treated as a quantum-mechanical system, but outside influences are treated classically, their effect on σ being represented by perturbing terms in the Hamiltonian which are considered definite if unknown functions of the time. It is of interest to see which aspects of irreversible processes are found in this approximation, and which ones depend essentially on the quantum nature of the surroundings.

Let the Hamiltonian of the system be

$$H=H_0+V(t), \tag{12.1}$$

where H_0 is stationary and defines the "energy levels" of the system, and $V(t)$ represents the perturbing effect of the environment. Suppose that at time t' we are given information which leads (by maximum-entropy inference, if needed) to the density matrix $\rho(t')$. At other times, the effect of the Hamiltonian (12.1) is to carry out a unitary transformation

$$\begin{aligned}\rho(t)&=U(t,t')\rho(t')U(t\,t')^{-1}\\&=U(t,t')\rho(t')U(t',t),\end{aligned} \tag{12.2}$$

where the time-development matrix $U(t,t')$ is determined from the Schrödinger equation (with $\hbar=1$)

$$i\frac{\partial}{\partial t}U(t,t')=H(t)U(t,t'), \tag{12.3}$$

with $U(t,t)=1$. The entropy

$$S(t)=-\mathrm{Tr}[\rho(t)\ln\rho(t)] \tag{12.4}$$

is unchanged by a unitary transformation, and therefore remains constant regardless of the magnitude or time variations of $V(t)$. Consider, however, the circumstance that $V(t)$ may not be known with certainty; during the time interval $(t'\rightarrow t)$ it may have been the operator $V^{(1)}(t)$ with probability P_1, or it may have been $V^{(2)}(t)$ with probability P_2, $\cdots$, etc. Then our state of knowledge of the system must be represented by a compound array, which is a fusion of several simple arrays corresponding to the different $V^{(\alpha)}(t)$, and which are subject to different rotations. At time t, the density matrix will be the average of the matrices that would result from each of the possible interactions:

$$\rho(t)=\sum_\alpha P_\alpha U^{(\alpha)}(t,t')\rho(t')U^{(\alpha)}(t',t), \tag{12.5}$$

and the transformation $\rho(t')\rightarrow\rho(t)$ is no longer unitary. We might also have a continuous distribution of unknown interactions, and therefore an integration over α, or more generally there might be several parameters $(\alpha_1\cdots\alpha_n)$ in $V(t)$, with probability distribution $P(\alpha_1\cdots\alpha_n)d\alpha_1\cdots d\alpha_n$. We will understand the notation in (12.5) to include such possibilities. Our uncertainty as to $V(t)$ will be reflected in increased uncertainty, as measured by the entropy, in our knowledge of the state of system σ. It is shown in Appendix A that, in case α is discrete, there is an upper limit to this increase, given by the following inequality:

$$S(t')\leq S(t)\leq S(t')+S(P_\alpha), \tag{12.6}$$

where

$$S(P_\alpha)\equiv-\sum_\alpha P_\alpha\ln P_\alpha. \tag{12.7}$$

Equation (12.6) has an evident intuitive content; the entropy of a system is a measure of our uncertainty as to its true state, and by applying an unknown signal to it, this uncertainty will increase, but not by more than our uncertainty as to the signal. The maximum increase in entropy can occur only in the following rather exceptional circumstances. The totality of all possible states of the system forms a function space $\mathcal{S}$. Suppose that our initial state of knowledge is that the system is in a certain subspace $\mathcal{S}_0$ of $\mathcal{S}$. If the perturbation $V^{(\alpha)}(t)$ is applied, this is transformed into some other subspace

$$\mathcal{S}_\alpha=U^{(\alpha)}\mathcal{S}_0,$$

and the maximum increase of entropy can occur only if the different subspaces $\mathcal{S}_\alpha$ are disjoint; i.e., every state in $\mathcal{S}_\alpha$ must be orthogonal to every state in $\mathcal{S}_\beta$ if $\alpha\neq\beta$. From this we see two reasons why the increase is usually less than the maximum possible amount; (a) it may be that even though $V^{(\alpha)}(t)$ and $V^{(\beta)}(t)$ are different functions, they nevertheless produce the same, or nearly the same, net transformation U in time $(t-t')$, so that our knowledge of the final state does not suffer from the uncertainty in the perturbation,

[14] J. S. Thomsen, Phys. Rev. **91**, 1263 (1953).
[15] R. T. Cox, *Statistical Mechanics of Irreversible Change* (Johns Hopkins Press, Baltimore, 1955).
[16] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949).

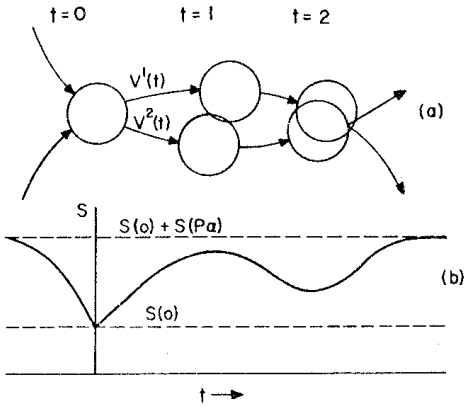

FIG. 1. Illustration of the subjective H theorem. (a) The array. (b) The resulting entropy curve.

and (b) our initial uncertainty may be so great that no such disjoint subspaces exist regardless of the nature of the $V^{(\alpha)}(t)$. The extreme case is that of complete initial ignorance; $\rho(t')$ is a multiple of the unit matrix. Then, no matter what is done to the system we cannot acquire any additional uncertainty, and the entropy does not change at all.

Equation (12.6) corresponds closely to relations that have been used to demonstrate the second law of thermodynamics in the past, and it will be called the "subjective H theorem." The inequalities hold for all times, positive or negative; given the density matrix at time $t'=0$, our uncertainty as to the perturbing signal $V(t)$ affects our knowledge of the past state of the system just as much as it does the future state. We cannot conclude from (12.6) that "entropy always increases." It may fluctuate up and down in any way as long as it remains within the prescribed bounds. On the other hand, it is true *without exception* that the entropy can at no time be less than its value at the instant t' for which the density matrix was given.

Figure 1 represents an attempt to illustrate several of the foregoing remarks by picturing the array. The diagram represents a portion of the surface of the unit hypersphere upon which all points of the array lie.[17] The interior of a circle represents a certain subspace $\mathcal{S}_i(t)$ which moves in accordance with the Schrödinger equation. Separated circles represent disjoint subspaces, while if two circles overlap, the subspaces have a certain linear manifold of states in common. The information given to us at time $t'=0$ locates the system somewhere in the subspace $\mathcal{S}_0$. The two possible interactions $V^{(1)}(t)$, $V^{(2)}(t)$ would induce rigid rotations of the hypersphere which would carry $\mathcal{S}_0$ along two different trajectories as shown. The lower part of the diagram represents the resulting entropy curve $S(t)$. If the subspaces $\mathcal{S}_1$, $\mathcal{S}_2$ coincide at some time t_1, then $S(t_1)=S(0)$. At times when they are completely separated, we have $S(t)=S(0)+S(P_\alpha)$, and in case of partial overlapping the entropy assumes intermediate values.

[17] The representation is necessarily very crude, since a continuous 1:1 mapping of a region of high dimensionality onto a region of lower dimensionality is topologically impossible. Nevertheless such diagrams represent enough of the truth to be very helpful, and there seems to be little danger of drawing fundamentally incorrect conclusions from them.

13. INFORMATION GAME

A typical process by which the subjective H theorem can lead to a continual increase of entropy, and which illustrates the essential nature of irreversibility, may be described in terms of a game. We have a sequence of observers $\mathcal{O}_1$, $\mathcal{O}_2$, $\mathcal{O}_3$, $\cdots$, who play as follows. At the beginning of the game they are given the possible Hamiltonians $H_\alpha=H_0+V^{(\alpha)}(t)$ and the corresponding probabilities P_α. At time t_1, observer $\mathcal{O}_1$ is given a density matrix $\rho_1(t_1)$. He computes from (12.5) the density matrix $\rho_1(t)$ which represents his state of knowledge at all other times on this basis, and the corresponding entropy curve $S_1(t)$. He then tells observer $\mathcal{O}_2$ the value which the density matrix $\rho_1(t_2)$ assumes at time t_2, and gives no other information.

$\mathcal{O}_2$ now computes a density matrix $\rho_2(t)$ which represents *his* state of knowledge at all times, on the basis of the information given him, and a corresponding entropy curve $S_2(t)$. He will, of course, have $\rho_2(t_2)=\rho_1(t_2)$, but in general there will be no other time at which these density matrices are equal. The reason for this is seen in Fig. 2, in which we assume that there are only two possible perturbations $V^{(1)}$, $V^{(2)}$. The information given to $\mathcal{O}_1$ locates the system somewhere in the subspace $\mathcal{S}_0$ at time t_1. At a different time t_2, this will be separated into two subspaces $\mathcal{S}_1(t_2)$ and $\mathcal{S}_2(t_2)$, corresponding to the two possible perturbations. For simplicity of the diagram, we assume that they are disjoint. At any other time t_3, the array of $\mathcal{O}_1$ is still represented by two possible subspaces $\mathcal{S}_1(t_3)$, $\mathcal{S}_2(t_3)$. Observer $\mathcal{O}_2$, however, is not in as advantageous a position as $\mathcal{O}_1$; although he is given the same density matrix at time t_2, and therefore can locate the subspaces $\mathcal{S}_1(t_2)$ and $\mathcal{S}_2(t_2)$, he does *not* know that $\mathcal{S}_1(t_2)$ is associated only with the perturbation $V^{(1)}$, $\mathcal{S}_2(t_2)$ only with $V^{(2)}$. Therefore, he can only assume that either perturbation may be associated with either subspace, and the array representing the state of knowledge of $\mathcal{O}_2$ for general times consists of four subspaces.

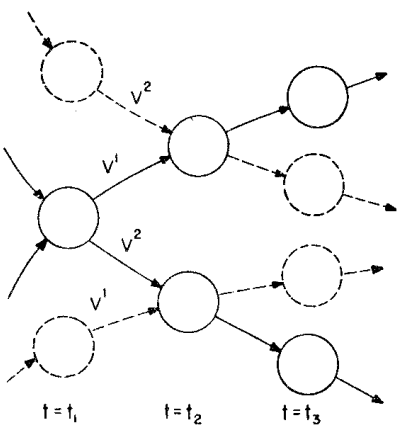

FIG. 2. The information game. The array of observer 1 at times t_1, t_2, t_3 is represented by solid circles. The array of observer 2 includes also the portion in dashed lines.

The game continues; $\mathcal{O}_2$ tells $\mathcal{O}_3$ what the density matrix $\rho_2(t_3)$ is, and $\mathcal{O}_3$ calculates *his* density matrix $\rho_3(t)$ (which, at general times other than t_3, must be represented by eight possible subspaces), and the entropy curve $S_3(t)$, $\cdots$, and so on.

The subjective H theorem applied to the nth observer gives

$$S_n(t_n)\leq S_n(t)\leq S_n(t_n)+S(P_\alpha), \tag{13.1}$$

while from the rules of the game,

$$S_{n-1}(t_n)=S_n(t_n). \tag{13.2}$$

Therefore, we have

$$S_1(t_1)\leq S_2(t_2)\leq S_3(t_3)\leq\cdots. \tag{13.3}$$

Note that no such inequality as $t_1\leq t_2\leq t_3\leq\cdots$ need be assumed, since the subjective H theorem works as well backwards as forwards; *the order of increasing entropy is the order in which information was transferred, and has nothing to do with any temporal order.*

An important conclusion from this game is that a density matrix does not in general contain all of the information about a system which is relevant for predicting its behavior; even though $\mathcal{O}_1$ and $\mathcal{O}_2$ had the same knowledge about possible perturbations, and represented the system by the same density matrix at time t_2, they were nevertheless in very different positions as regards the ability to predict its behavior at other times. The information which was lost when $\mathcal{O}_1$ communicated with $\mathcal{O}_2$ consisted of correlations between possible perturbing forces and the different simple arrays which are contained in the total compound array. The effect of this information loss on an observer's knowledge of the system was not immediate, but required time to "develop." Thus, it is not only the entire density matrix, but also the particular resolution (12.5) into parts arising from different simple arrays, that is relevant for the prediction problem.

For these and other reasons, an array must be considered as a more fundamental and meaningful concept than the density matrix; even though many different arrays lead to the same density matrix, they are not equivalent in all respects. In problems where the entropy varies with time, the array which at each instant represents the density matrix as a mixture of orthogonal states is difficult to obtain, and without any particular significance. The one which is resolved into simple arrays, each representing the unfolding of a possible unitary transformation, provides a clearer picture of what is happening, and may contain more information relevant to predictions.

The density matrices $\rho_n(t)$ determined by the successive observers in the information game may be represented in a compact way as follows. Consider first the case where there is only a single possible perturbation, and therefore ρ undergoes a unitary transformation

$$\rho(t)=U(t,t')\rho(t')U^{-1}(t,t'). \tag{13.4}$$

This could also be written in another kind of matrix notation as

$$\rho_{nn'}(t)=\sum_{kk'}(nn'|G(t,t')|kk')\rho_{kk'}(t'), \tag{13.5}$$

or,

$$\rho(t)=G(t,t')\rho(t'), \tag{13.6}$$

where

$$(nn'|G(t,t')|kk')=U_{nk}(t,t')U_{n'k'}{}^*(t,t') \tag{13.7}$$

is the direct product matrix

$$G=U\times U^*. \tag{13.8}$$

In (13.4) ρ is considered as an $(N\times N)$ matrix, while in (13.6) it is a vector with N^2 components, and G is an $(N^2\times N^2)$ matrix. It is readily verified that G has the group property

$$G(t,t')G(t',t'')=G(t,t'') \tag{13.9}$$

in consequence of the same property possessed by U.

The advantage of writing the transformation law in the form (13.6) is that, in the case where there are several possible perturbations $V^{(\alpha)}(t)$, the transformation with time (12.5) cannot be written as a similarity transformation with any "averaged U matrix," but it is expressible by a G matrix averaged over the distribution P_α:

$$\rho(t)=\mathcal{G}(t,t')\rho(t'), \tag{13.10}$$

where

$$\mathcal{G}(t,t')=\sum_\alpha P_\alpha G^{(\alpha)}(t,t'). \tag{13.11}$$

The essential feature of the irreversibility found in the information game is that $\mathcal{G}(t,t')$ does *not* possess the group property (13.9):

$$\mathcal{G}(t,t')\mathcal{G}(t',t'')\neq\mathcal{G}(t,t''), \tag{13.12}$$

for on one side we have the product of two averages, on the other the average of a product. If (13.12) were an equality valid for all times, it would imply that $\mathcal{G}$ has an inverse $\mathcal{G}^{-1}(t,t')=\mathcal{G}(t',t)$, whereupon (13.10) could be solved for $\rho(t')$,

$$\rho(t')=\mathcal{G}(t',t)\rho(t). \tag{13.13}$$

But then, the subjective H theorem would give

$$S(t)\geq S(t'), \text{ from (13.10)};$$
$$S(t')\geq S(t), \text{ from (13.13)}.$$

In the general case $\mathcal{G}(t,t')$ may be singular.

The density matrices of the successive observers are now given by

$$\begin{aligned}\rho_1(t)&=\mathcal{G}(t,t_1)\rho_1(t_1),\\ \rho_2(t)&=\mathcal{G}(t,t_2)\mathcal{G}(t_2,t_1)\rho_1(t_1),\\ \rho_3(t)&=\mathcal{G}(t,t_3)\mathcal{G}(t_3,t_2)\mathcal{G}(t_2,t_1)\rho_1(t_1),\end{aligned} \tag{13.14}$$

in which the information game is exhibited as a Markov

chain,[18,19] the ordering index giving the sequence of information transfer rather than a time sequence.

14. STEP-RELAXATION PROCESS

In the preceding section, the information game was interpreted in the "passive" sense; i.e., we assumed that a certain one of the perturbations $V^{(\alpha)}(t)$ was the one in fact present, and this same one persisted for all time. The different observers then represent different ways of looking at what is in reality only one physical situation, their increasing uncertainty as to the true state being due only to the incomplete transmission of information from one observer to the next.

The game may equally well be interpreted in the "active" sense, in which there is only one observer, but at each of the times $t_1, t_2, t_3, \cdots$, the perturbation is interrupted and a new choice of one of the $V^{(\alpha)}(t)$ made in accordance with the probability distribution P_α. Although it is not required by the equations, it is perhaps best at this point, merely to avoid certain teleological distractions, to assume that

$$t_1 \leq t_2 \leq t_3 \leq \cdots. \tag{14.1}$$

At each of these times the observer loses exactly the same information that was lost in the communication process of the passive interpretation, and his knowledge of the state of the system progressively deteriorates according to the same Eqs. (13.14) as before. The density matrix which represents the best physical predictions he is able to make is then

$$\rho(t) = \begin{cases} \rho_1(t), & t_1 \leq t \leq t_2 \\ \rho_2(t), & t_2 \leq t \leq t_3 \\ \cdots & \\ \rho_n(t), & t_n \leq t \leq t_{n+1}. \end{cases} \tag{14.2}$$

This is a continuous function of time, since

$$\rho_n(t_n) = \rho_{n-1}(t_n).$$

In the following we consider only the case where ρ operates on a function space σ of finite dimensionality N. The maximum possible entropy of such a system is

$$S_{\max} = \ln N, \tag{14.3}$$

which is attained if and only if ρ is a multiple of the unit matrix:

$$\rho_{nk} = N^{-1}\delta_{nk}. \tag{14.4}$$

From this fact and (13.3), it follows that the sequence of values $S(t_n)$ must converge to some definite final entropy:

$$\lim_{n\to\infty} S(t_n) = S_\infty \leq S_{\max}. \tag{14.5}$$

To investigate the limiting form of the density matrix as $t \to \infty$, some spectral properties of the transformation matrices are needed. Let $\mathcal{G}$ stand for any one of the $(N^2 \times N^2)$ step transformations $\mathcal{G}(t_n, t_{n-1})$ operating in the direct product space $\sigma \times \sigma = \sigma^2$, and x, y be any vectors of N^2 components upon which $\mathcal{G}$ can operate. Instead of denoting the components of x, y by a single index running from 1 to N^2, we use two indices each running from 1 to N, so that x, y may also be interpreted as $(N \times N)$ matrices operating in the space σ. We define inner products in the usual way by

$$(x,y) = \sum_{n,k=1}^{N} x_{nk}^* y_{nk} = \mathrm{Tr}(x^\dagger y). \tag{14.6}$$

Since $\mathcal{G}$ is not a normal matrix (i.e., it does not commute with its Hermitian conjugate), we may not assume the orthogonality, or even the existence of a complete set, of its eigenvectors. However, every square matrix has at least one eigenvector belonging to each eigenvalue, so that as x varies over all possible directions, the set of numbers

$$g(x) \equiv (x, \mathcal{G}x)/(x,x)$$

includes all the eigenvalues of $\mathcal{G}$. Writing

$$x_\alpha = U^{(\alpha)} x U^{(\alpha)-1}$$

it is readily shown that $(x_\alpha, x_\alpha) = (x,x)$. From (12.5) we have

$$\mathcal{G}x = \textstyle\sum_\alpha P_\alpha x_\alpha,$$

and therefore

$$|(x,\mathcal{G}x)| = |\textstyle\sum P_\alpha (x,x_\alpha)| \leq \sum P_\alpha |(x,x_\alpha)| \leq \sum P_\alpha [(x,x)(x_\alpha,x_\alpha)]^{\frac{1}{2}} = (x,x),$$

where the Schwarz inequality has been used. We conclude that for all x,

$$|g(x)| \leq 1, \tag{14.7}$$

with equality if and only if $x_\alpha = x$ for all α. This is evidently the case if x is a multiple of the unit matrix; thus (14.4) is always an eigenvector of $\mathcal{G}$ with the eigenvalue unity. Only in exceptional circumstances could $\mathcal{G}$ have any other eigenvalue of magnitude unity; this would require that some x other than (14.4) must exist which is invariant under all the unitary transformations $U^{(\alpha)}$.

By a similar argument, one can derive a slightly weaker inequality than (14.7):

$$(\mathcal{G}x, \mathcal{G}x) \leq (x,x), \tag{14.8}$$

which shows that $\mathrm{Tr}[\rho^2(t_n)]$ is a non-increasing function of n, which must converge to some definite final value.

From these relations several features of the long-time behavior may be inferred. First consider $\mathcal{G}$ to be brought, by similarity transformations, to the canonical form

$$T\mathcal{G}T^{-1} = \begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \cdot & \\ & & & A_r \end{bmatrix}, \tag{14.9}$$

[18] J. L. Doob, Ann. Math. **43**, 351 (1942).

[19] W. Feller, *An Introduction to Probability Theory and its Applications* (John Wiley and Sons, Inc., New York, 1950).

where each A_i contains all those, and only those, terms which arise from the eigenvalue λ_i. If λ_i is nondegenerate, A_i is simply the number λ_i. If λ_i is an m-fold multiple root of $|\mathcal{G}-\lambda 1|=0$, then A_i may be the $(m\times m)$ diagonal matrix $\lambda_i 1$, or it may have one or more "superdiagonal" terms[20]

$$A_i=\begin{pmatrix}\lambda_i & 1 & 0 & 0 & \cdot\\ 0 & \lambda_i & 1 & 0 & \cdot\\ 0 & 0 & \lambda_i & 0 & \cdot\\ \cdot & \cdot & \cdot & \cdot & \cdot\end{pmatrix}. \tag{14.10}$$

The simplest type of step-relaxation process to describe is the one in which all of the matrices $\mathcal{G}(t_n,t_{n-1})$ are equal; i.e., $t_n=n\tau$, and each of the possible perturbations $V^{(\alpha)}(t)$ is periodic with period τ. The general conclusions will be the same regardless of whether this is the case. We now have

$$\rho(t_n)=\mathcal{G}^n\rho(0), \tag{14.11}$$

and those parts of the canonical form $T\mathcal{G}^nT^{-1}$ arising from the eigenvalue $\lambda=0$ are annihilated in a finite number of steps, while the sections $A_i{}^n$ for which $0<|\lambda_i|<1$ are exponentially attenuated. Thus, the situation as $n\to\infty$ depends only on those $A_i{}^n$ for which $|\lambda_i|=1$. There are two possibilities:

(a) The ergodic case. If $\mathcal{G}$ has only one eigenvalue with $|\lambda_i|=1$ [which must therefore correspond to the eigenvector (14.4)], the sequence $\{\mathcal{G}^n\}$ converges to the projection onto (14.4); i.e.,

$$\lim_{n\to\infty}\rho(t_n)=N^{-1}1, \tag{14.12}$$

independently of $\rho(0)$. The information contained in the initial distribution becomes completely lost, and the limiting entropy is the maximum possible value (14.3). In practice, this would be the usual situation.

(b) If $\mathcal{G}$ has more than one eigenvalue with $|\lambda_i|=1$, the density matrix does not necessarily approach any fixed limit. Nevertheless, the entropy $S(t_n)$ must do so. Therefore, by an argument like that of Appendix A, the ultimate behavior must be one in which a certain similarity transformation is repeated indefinitely. For example, this ultimate transformation could consist of a permutation of the rows and columns of ρ. In this case, traces of the initial information are never lost, and the limiting entropy is less than $\ln N$.

These results correspond closely to those of the theory of long-range order in crystals,[21,22] in which one introduces a stochastic matrix which relates the probability distribution of one crystal layer to that of an adjacent one. The existence or nonexistence of probability influences over arbitrarily long distances depends on the degeneracy (in magnitude) of the greatest eigenvalue of this matrix.

[20] S. Lefschetz, *Lectures on Differential Equations* (Princeton University Press, Princeton, 1946), Chap. I.
[21] J. Ashkin and W. E. Lamb, Jr., Phys. Rev. **64**, 159 (1943).
[22] G. F. Newell and E. W. Montroll, Revs. Modern Phys. **25**, 353 (1953).

15. PERTURBATION BY A STATIONARY STOCHASTIC PROCESS

We now investigate the change in our knowledge of the state of a system for which the perturbing Hamiltonian $V(t)$ is a stationary random function of time. Certain aspects of irreversible processes may be described in terms of such a model, although we will find that other essential features, such as the mechanism by which thermal equilibrium is established, require better approximations in which the quantum nature of the perturbing forces is taken into account.

In classical statistical mechanics an ergodic hypothesis facilitated the mathematics by allowing one to replace time averages by ensemble averages. We now find the reverse situation; that calculation of $\mathcal{G}(t,t')$ is facilitated by an ergodic principle that enables us to replace the "ensemble average" (13.11) by a time average, and then to make use of correlation functions and the Wiener-Khintchine theorem. In Eq. (13.10), $G^{(\alpha)}(t,t')$ may be regarded as a certain functional $F[V^{(\alpha)}(t)]$ of $V^{(\alpha)}(t)$, which depends on the values assumed by this operator in the time interval $(t'\to t)$. The statement that $V(t)$ is a stationary stochastic process implies that the average of this functional

$$\bar{F}_0{}^\alpha=\sum_\alpha P_\alpha F[V^{(\alpha)}(t)] \tag{15.1}$$

is not affected by which particular sample of the function $V^{(\alpha)}(t)$ is involved in (15.1); i.e., if we were to insert instead the values assumed by $V^{(\alpha)}(t)$ in some other equal time interval $(t'+\tau\to t+\tau)$, the average

$$\bar{F}_\tau{}^\alpha=\sum_\alpha P_\alpha F[V^{(\alpha)}(t+\tau)] \tag{15.2}$$

would be independent of τ. Conversely, if

$$\bar{F}_\tau{}^\alpha=\bar{F}_0{}^\alpha$$

for all functionals and all values of τ, this implies that $V(t)$ has exactly the same statistical properties after any time translation, so that $V(t)$ must be a stationary stochastic process. Under these conditions the expression (15.1) will not be affected by averaging it over all time translations;

$$\bar{F}^\alpha=\bar{F}^{\alpha\tau}=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}\sum P_\alpha F[V^{(\alpha)}(t+\tau)]d\tau. \tag{15.3}$$

Our ergodic assumption is that in this formula the averaging over P_α is redundant; i.e.,

$$\bar{F}^\alpha=\bar{F}^\tau=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}F[V(t+\tau)], \tag{15.4}$$

in which the parameter α may be dropped.

The preceding paragraph was written in a conventional kind of language which made it appear that a substantial assumption has been introduced; one whose correctness should be demonstrated if the resulting

theory is to be valid. Such conventional modes of expression, however, do not do full justice to the situation as it is presented to us in practice. To see this, we need only ask, "What do we really mean by the functions $V^{(\alpha)}(t)$ and the probabilities P_α?" In most cases there is only one function $V(t)$. Knowledge of the statistical properties of V cannot then be obtained by observing the frequency with which the particular function $V^{(\alpha)}(t)$ appears in an ensemble of similar situations, because no such ensemble exists. By the probability P_α we could mean only the average frequency, over long periods of time, with which a configuration locally like $V^{(\alpha)}$ occurs in the *single* function $V(t)$. *The means by which the probabilities P_α are defined already involve a time-averaging procedure.* Therefore (15.4) is not an assumption at all; it is merely the natural way of stating a fact which is expressed only awkwardly by (15.1). Equation (15.4) carries out in a single step both the averaging procedure in (15.1) and the process by which the $V^{(\alpha)}$ and P_α are determined.

The problem is thus reduced to a calculation of $\mathcal{G}(t,t')=\mathcal{G}(t-t')$, where

$$\mathcal{G}(t)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[U(t+\tau,\tau)\times U^*(t+\tau,\tau)]d\tau. \quad (15.5)$$

The exact evaluation of $\mathcal{G}(t)$ would require a rigorous solution of the Schrödinger equation (12.3) for arbitrary $V(t)$. In practice one must resort to approximate solutions at this point, and it is fortunate that in many practical situations $\mathcal{G}(t)$ is determined to a good approximation by the use of second-order perturbation theory. The characteristic feature of such problems is found by noting that although $\mathcal{G}(t,t')$ does not in general possess the group property (13.12), an equality of this form may be approximately correct for certain choices of times, provided the perturbation is weak and has a short correlation time. Thus, suppose that $t''<t'<t$, and we try to represent $\mathcal{G}(t,t'')$ by a product

$$\mathcal{G}(t,t'')\simeq\mathcal{G}(t,t')\mathcal{G}(t',t''). \quad (15.6)$$

The approximation involved in (15.6) consists of the discarding, at time t', of mutual correlations which were built up in the time interval $(t''\to t')$ between possible functions $V(t)$ and the corresponding simple arrays. If $V(t)$ is a weak perturbation, it can change the state of the system only slowly, and a long time is required for any strong correlations to develop. However, if the time τ_c over which appreciable autocorrelations persist in $V(t)$ is very short compared to $(t'-t'')$, the mutual correlations discarded were actually accumulated only during an interval τ_c just prior to t', and will be relatively unimportant; thus (15.6) may be a very good approximation. On the other hand, it will never be an exact equality, because the values of $V(t)$ just prior to t' will necessarily have some influence on its behavior just after t', whose effect is lost in the approximation.

These considerations lead to a means for approximate calculation of $\mathcal{G}(t-t')$. Divide the time interval $(t'\to t)$ into n equal intervals: $(t-t')=n\tau$, and set

$$\mathcal{G}(t-t')\simeq[\mathcal{G}(\tau)]^n. \quad (15.7)$$

If $\tau\gg\tau_c$, this is a good approximation, and if in addition it is possible to choose τ short enough so that the change of state during time τ is given adequately by second-order perturbation theory, this leads to a feasible method of calculation. With this approximation, the theory is reduced in its essentials to that of the step-relaxation process of the preceding section.

The most important feature of the final solution can be seen directly from (15.7). The change of state with time has a simple "stroboscopic" property: if we observe the density matrix only at the instants $t_m=m\tau$, we see the approach to equilibrium take place in a stepwise exponential fashion, describable by relaxation times. This result is already guaranteed by the nature of the approximation in (15.7) quite independently of any further details, and in particular independently of any assumptions concerning the level spacings of the system. However, the level spacings are important in determining the appropriate form of the solution. For example, if the correlation time τ_c is extremely short compared to all characteristic times of the system, we may, while satisfying the condition $\tau\gg\tau_c$, still have $|\omega_{kl}|\tau\ll 1$ for all transitions frequencies ω_{kl}. In this case, the change in ρ during time τ is very small, and (15.7) may be replaced by a linear differential equation with constant coefficients. Thus, defining K_1 by

$$K_1=[\mathcal{G}(\tau)-1]/\tau, \quad (15.8)$$

we have approximately

$$d\rho/dt\simeq K_1\rho. \quad (15.9)$$

K_1 has N^2 eigenvalues λ_i, one of which must be zero since K_1 annihilates (14.4). By an argument like that leading to (14.7) one shows that $\mathrm{Re}(\lambda_i)\leq 0$. Thus each element of ρ will relax to a final state according to a superposition of exponentials $\exp(\lambda_i t)$, with several different relaxation times in general.

The right-hand side of (15.9) is generally a poor approximation to the instantaneous time derivative of ρ, but gives only the average rate of change over the period τ. Similarly, the matrix K_1 resembles a time derivative of $\mathcal{G}$; in the following section we present reasons for expecting that a slightly different definition of K_1 will render (15.9) more accurate as far as giving the long-term drift is concerned.

16. EXACTLY SOLUBLE CASE

In the case where the perturbation $V(t)$ commutes with H_0, it is possible to evaluate (15.5) exactly without use of perturbation theory. This case is a very special one, since the perturbation causes no transitions but only a loss of coherences; nevertheless it has found some

applications in the theory of pressure-broadening of spectral lines[23,24] and exchange narrowing[25] in paramagnetic resonance.

The perturbing forces represented by $V(t)$ often arise as a superposition of many small independent effects, and in this case the central limit theorem of probability theory shows that the distribution of $V(t)$ will be Gaussian. Furthermore, in most applications one will not have enough information about $V(t)$ to determine any unique objective probability distribution; we may know, for example, only the average energy density, therefore the mean-square value, of the perturbing fields, plus a few features of their spectral density. Maximum-entropy inference would then be needed in order to represent our knowledge of $V(t)$ in a way free of arbitrary assumptions. Since a Gaussian distribution has maximum entropy for a given variance, one should always use a Gaussian distribution if the available information consists only of the first and second moments. In the following we consider only the Gaussian case.

The Hamiltonian has matrix elements

$$H_{kl}(t)=[\omega_k+V_k(t)]\delta_{kl}. \tag{16.1}$$

The solution of (12.3) for the time-development matrix is substituted into (15.4) to give

$$(kk'|\mathcal{G}(t,t')|ll') = \delta_{kl}\delta_{k'l'}e^{i\omega_{k'k}(t-t')}\left\langle \exp\left[i\int_{t'}^{t} f_{k'k}(t'')dt''\right]\right\rangle, \tag{16.2}$$

where $\omega_{k'k}=\omega_{k'}-\omega_k$, and

$$f_{k'k}(t)=V_{k'}(t)-V_k(t) \tag{16.3}$$

is a real Gaussian random function with mean value zero (by definition, since any constant part of V may be included in H_0). So also, therefore, is the function

$$g(t)=\int_0^t f(t')dt', \tag{16.4}$$

where we have dropped the subscripts for brevity. The probability distribution of $g(t)$ is determined by its second moment

$$\sigma(t)=\langle g^2(t)\rangle=\int_0^t dt'\int_0^t dt''\langle f(t')f(t'')\rangle$$

$$=\int_0^t dt'\int_0^t dt''\,\varphi(t'-t''), \tag{16.5}$$

where

$$\varphi(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} f(t+\tau)f(t)dt \tag{16.6}$$

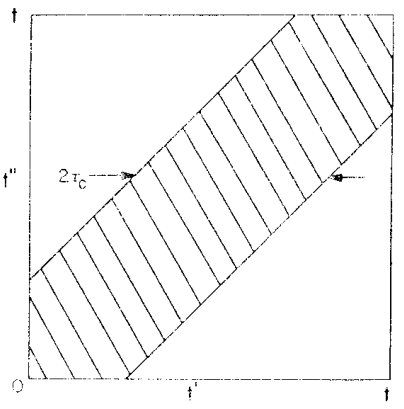

FIG. 3. Region of integration in Eq. (16.5). Appreciable contributions to the integral come only from shaded part.

is the autocorrelation function of $f(t)$. A short calculation shows that for a Gaussian function with variance $\sigma(t)$, the average required in (16.2) is

$$\langle e^{ig}\rangle=e^{-\frac{1}{2}\sigma(t)}, \tag{16.7}$$

and thus the exact solution (13.10) of the relaxation problem is

$$\rho_{kk'}(t)=e^{i\omega_{k'k}t}\rho_{kk'}(0)e^{-\frac{1}{2}\sigma_{k'k}(t)}. \tag{16.8}$$

Since $\sigma_{kk}=0$, the diagonal elements of ρ are unchanged, but the off-diagonal elements relax to zero in a manner described by (16.5).[26]

We assume that there exists a correlation time τ_c such that the correlation function (16.6) is essentially zero whenever $|\tau|>\tau_c$. The region of integration in (16.5) may be represented by a square as in Fig. 3, and it is seen that although $\sigma(t)$ necessarily starts out proportional to t^2 for small t, it approximates a linear function of time when $t>\tau_c$. The function $\sigma(t)$ therefore has the form of Fig. 4, and for $t>\tau_c$ it reduces to

$$\sigma(t)\cong 2\pi I(0)[t-\tau_1]. \tag{16.9}$$

The quantity

$$I(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(t)e^{-i\omega t}dt \tag{16.10}$$

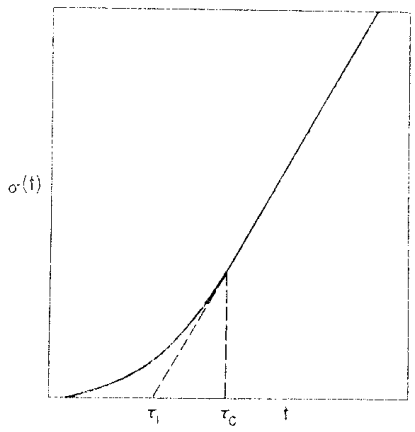

FIG. 4. The function $\sigma(t)$.

[23] P. W. Anderson, Phys. Rev. **76**, 647 (1949). Earlier references are given in this paper.

[24] S. Bloom and H. Margenau, Phys. Rev. **90**, 791 (1953).

[25] P. W. Anderson and P. R. Weiss, Revs. Modern Phys. **25**, 269 (1953).

[26] In some cases it may be possible to evaluate (16.7) directly even though $\langle g^2\rangle$ does not exist. For example, we may have $f(t)=$constant, with probability distribution $p(f)df$. Then (16.7) is a Fourier transform, and with Lorentzian $p(f)$ we obtain a decay law exactly exponential for all times.

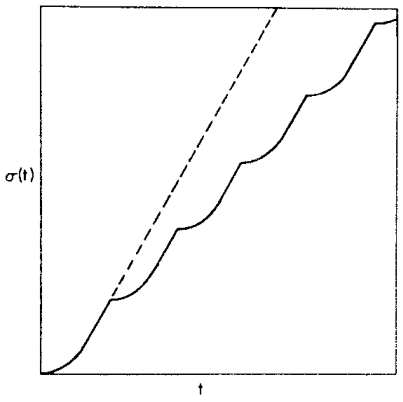

FIG. 5. Slip effect caused by discarding correlations. The approximate solution is represented by the solid line, while the dashed line is the exact solution.

is the spectral density of $f(t)$ for frequency ω, and τ_1 is a short time somewhat less than τ_c, indicated on Fig. 4. Thus when $t>\tau_c$, the relaxation process goes into an exponential damping, the element $\rho_{kk'}$ having a relaxation time $T_{kk'}$, where

$$1/T_{kk'}=\pi I_{k'k}(0). \tag{16.11}$$

Note that although the final formulas involve only the spectral density at zero frequency, the condition that $\varphi(t)$ should be very small for $|t|>\tau_c$ implies certain conditions on $I(\omega)$ at other frequencies. It is required not only that $I(\omega)$ be large over a band width $\sim\tau_c^{-1}$ of frequencies, but also that it be a sufficiently smooth function of frequency. Discontinuities in $I(\omega)$ produce oscillations in $\varphi(t)$ and $\sigma(t)$ which may persist for long periods, rendering (16.9) inaccurate.

It is of interest to compare the exact solution (16.8) with the one which would be obtained using the approximation of (15.7). Here we stop the integration process of (16.5) after each interval τ, throw away mutual correlations between ρ and $V(t)$, and use the density matrix thus obtained as the initial condition for the next period. The resulting $\sigma(t)$ is illustrated in Fig. 5. It is seen that the approximation "slips behind" the exact solution by a time delay τ_1 each time the mutual correlations are discarded.

There is an apparent paradox in this result. It seems natural to suppose that any mathematical approximation must "lose information," and therefore increase the entropy. However, we find the relaxation process taking place more rapidly in the exact treatment than in the approximate one: $S_{\text{exact}}(t)\geq S_{\text{approx}}(t)$. Thus, the approximation has not "lost information," but has "injected false information." The reason for this can be visualized as follows. Suppose that at time $t=0$ the array consisted of a single point, i.e., a pure state. At later times it will consist of a continuous distribution of points filling a certain volume, which continually expands as t increases. It is very much like an expanding sphere of gas, where strong correlations will develop between position and velocity; a molecule near the edge of the sphere is very likely to be moving away from the center. This corresponds roughly to the correlations between different states of the array and different possible perturbing signals $V(t)$. Now suppose that in an expanding gas sphere these correlations are suddenly lost; the set of velocities existing at time τ is suddenly redistributed among the molecules at random. Then a molecule near the edge is equally likely to be moving toward or away from the center. The general expansion is momentarily interrupted, but soon resumes its former rate.

This paradox shows that "information" is an unfortunate choice of word to describe entropy expressions. Furthermore, one can easily invent situations where acquisition of a new piece of information (that an event previously considered improbable had in fact occurred) can cause an increase in the entropy. The terms "uncertainty" or "apparent uncertainty" come closer to carrying the right connotations.

Note that, if we were to use the slope of the approximate curve in Fig. 5 just before time τ, instead of the average drift over period τ, to calculate the relaxation time, we would obtain a more accurate value whenever $\tau>\tau_c$.

17. PERTURBATION THEORY APPROXIMATION

Returning to the general case, we conjecture that a similar situation to that just found will occur: i.e., that the differential equation

$$d\rho/dt=K_2\rho, \tag{17.1}$$

where

$$K_2=\left(\frac{d\mathcal{G}}{dt}\right)_{t\to\tau_-}, \tag{17.2}$$

will give a slightly more accurate long-term solution than will (15.9). The evaluation of $\mathcal{G}(\tau)$ using perturbation theory is in essence identical with the treatments of nuclear spin relaxation given by Wangsness and Bloch,[27] Fano,[28], Ayant,[29] and Bloch.[30,31] Only a brief sketch of the calculations is given here, although we wish to point out certain limitations on the applicability of previous treatments.

One solves the equation of motion (12.3) by use of time-dependent perturbation theory, retaining terms through the second order. The result of substituting this solution into (15.5) is expressed compactly as follows. Define a matrix $\varphi(t)$ whose elements consist of all correlation functions of V_{kl}, $V_{k'l'}$:

$$(kk'|\,\varphi(t-t')\,|ll')=\langle V_{kl}(t)V_{k'l'}^*(t')\rangle, \tag{17.3}$$

in which the average is taken over all time translations. $\varphi(t)$ has the symmetry properties

$$(kk'|\,\varphi(t)\,|ll')=(ll'|\,\varphi(t)\,|kk')^*=(l'l|\,\varphi(-t)\,|k'k). \tag{17.4}$$

We assume again that there exists a correlation time

[27] R. K. Wangsness and F. Bloch, Phys. Rev. **89**, 728 (1953).
[28] U. Fano, Phys. Rev. **96**, 869 (1954).
[29] Y. Ayant, J. phys. radium **16**, 411 (1955).
[30] F. Bloch, Phys. Rev. **102**, 104 (1956).
[31] F. Bloch, Phys. Rev. **105**, 1206 (1957).

τ_c such that all components of $\varphi(t)$ are essentially zero whenever $t>\tau_c$. In this case the "partial Fourier transforms" of φ, defined by

$$\Phi(\omega)=\int_0^\tau e^{-i\omega t}\varphi(t)dt \tag{17.5}$$

are independent of τ. Finally, we introduce the symbols

$$(kk'|nn')=(kk'|\Phi(\omega_{n'k'})|nn')=(nn'|kk')^*. \tag{17.6}$$

In terms of these quantities, we obtain

$$\begin{aligned}(kk'|\mathcal{G}(\tau)|nn')=e^{-i\omega_{kk'}\tau}\{&\delta_{kn}\delta_{k'n'}\\&-\delta_{kn}q(\omega_{n'k'})\textstyle\sum_p(pp|k'n')-\delta_{k'n'}q(\omega_{kn})\sum_p(kn|pp)\\&+q(\omega_{kn}-\omega_{k'n'})[(kk'|nn')+(n'n|k'k)]\},\end{aligned} \tag{17.7}$$

where

$$q(\omega)=(e^{i\omega\tau}-1)/i\omega.$$

In the case of extremely short correlation time, so that $|\omega_{kn}\tau|\ll 1$, as assumed in (15.9) and (17.1), $q(\omega_{kn})=\tau$ for all transition frequencies ω_{kn}, and (17.7) leads to the differential equation

$$\begin{aligned}\dot\rho_{kk'}+i\omega_{kk'}\rho_{kk'}=\sum_{n,n'}\{&[(kk'|nn')+(n'n|k'k)]\rho_{nn'}\\&-(nn|k'n')\rho_{kn'}-(kn|n'n')\rho_{nk'}\}.\end{aligned} \tag{17.8}$$

This case of perturbation by extremely wide-band "white noise" applies to many cases of nuclear spin relaxation in liquids,[32] its condition of validity being that the correlation time (roughly, period of molecular rotation) is short compared to the Larmor precession periods.

In the approximation of (17.8) the quantities $(kk'|nn')$ are real if $\varphi(t)$ is real, as will usually be the case:

$$(kk'|nn')\simeq\int_0^\infty\cos(\omega_{n'k'}t)(kk'|\varphi(t)|nn')dt. \tag{17.9}$$

The neglected term is small, since by hypothesis $\varphi(t)$ is very small before $\sin(\omega_{n'k'}t)$ attains an appreciable magnitude. Equation (17.9) is π times the "mixed spectral density," at frequency $\omega_{n'k'}$, of $V_{kn}(t)$ and $V_{k'n'}(t)$. To interpret (17.8) we transfer all terms containing $\rho_{kk'}$ to the left-hand side

$$\dot\rho_{kk'}+\left(\frac{1}{T_{kk'}}+i\omega_{kk'}\right)\rho_{kk'}=\text{"driving forces."} \tag{17.10}$$

The relaxation times $T_{kk'}$ are given by

$$1/T_{kk'}=\gamma_k+\gamma_{k'}-\gamma_{kk'}, \tag{17.11}$$

where

$$\begin{aligned}\gamma_k&=\textstyle\sum_p(kk|pp),\\ \gamma_{kk'}&=(kk'|kk')+(k'k|k'k).\end{aligned} \tag{17.12}$$

If the correlation time τ_c is not short compared to the periods $(\omega_{kn})^{-1}$, then the time of integration τ must be chosen so long that the formulation (17.8) in terms of a differential equation breaks down. In this case a different approach, used by Wangsness and Bloch,[27] may be attempted. Here one removes the rapid time variations of ρ due to H_0 by transforming to the interaction representation, in which the density matrix is

$$\bar\rho(t)=e^{iH_0t}\rho(t)e^{-iH_0t}, \tag{17.13}$$

and attempts to describe the relaxation process by a linear differential equation with constant coefficients, satisfied by the slowly varying $\bar\rho(t)$. This is not always possible, however, for Eqs. (15.5) and (15.7) hold only in the original Schrödinger representation. If H_0 is diagonal, the matrix $\mathcal{G}_I$ which gives the change of state in the interaction representation,

$$\bar\rho(t)=\mathcal{G}_I(t,t')\bar\rho(t'), \tag{17.14}$$

is related to the previous $\mathcal{G}$ by

$$\begin{aligned}&(kk'|\mathcal{G}_I(t+\tau,t)|nn')\\&\quad=e^{i(\omega_{kn}-\omega_{k'n'})t}e^{i\omega_{kk'}\tau}(kk'|\mathcal{G}(\tau)|nn'),\end{aligned} \tag{17.15}$$

so that although $\mathcal{G}$ is a function only of $(t-t')$, this is not in general true of $\mathcal{G}_I$. Consequently an approximation of the form (15.7) cannot be valid in general for $\mathcal{G}_I$. However, it is seen that those elements of $\mathcal{G}_I$ for which

$$\omega_{kn}=\omega_{k'n'} \tag{17.16}$$

depend only on $(t-t')$. Therefore, if by any means one can justify discarding elements of $\mathcal{G}_I$ not satisfying (17.16), this method will work. Referring to (17.7), it is seen that the elements which satisfy (17.16) are just the "secular terms" which increase proportional to τ, while the unwanted terms are the oscillating ones. Therefore if the time τ is sufficiently long, and the level spacings are such, that the quantities

$$|\omega_{kn}-\omega_{k'n'}|\tau$$

are either large compared to unity, or zero, for all combinations of levels, the secular terms will be much larger than the oscillating ones and we obtain the approximate differential equation

$$\begin{aligned}\frac{\partial\bar\rho_{kk'}}{\partial t}=\sum_{n,n'}\{&\delta(\omega_{kn}-\omega_{k'n'})[(kk'|nn')\\&+(n'n|k'k)]\bar\rho_{nn'}-\delta(\omega_{k'n'})(nn|k'n')\bar\rho_{kn'}\\&-\delta(\omega_{kn})(kn|n'n')\bar\rho_{nk'}\}.\end{aligned} \tag{17.17}$$

If there is no degeneracy and the density matrix is initially diagonal, (17.17) reduces to

$$\partial\bar\rho_{kk'}/\partial t=2\pi\delta_{kk'}\textstyle\sum_n I_{kn}(\omega_{nk})(\bar\rho_{nn}-\bar\rho_{kk}), \tag{17.18}$$

where

$$I_{kn}(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}(kk|\varphi(t)|nn)dt \tag{17.19}$$

[32] Bloembergen, Purcell, and Pound, Phys. Rev. 73, 679 (1948).

is the spectral density, at frequency ω, of $V_{kn}(t)$. Equation (17.18) is to be compared to (11.7); we have a time-proportional transition probability satisfying the condition of microscopic reversibility. Note, however, that this result depends entirely on the assumptions as to spectral properties of $V(t)$ and the various approximations made, which ensured that off-diagonal elements of ρ would not appear. From the definition (15.5) of $\mathcal{G}$ it follows that, in the case that $\rho(0)$ is diagonal, the rigorous expression for diagonal elements at time t is

$$\rho_{kk}(t)=\sum_n\langle|U_{kn}(t,0)|^2\rangle\rho_{nn}(0)$$
$$=\sum_n\lambda_{kn}(t)\rho_{nn}(0), \quad (17.20)$$

so that in general the transition probabilities $\lambda_{kn}(t)$ are neither time proportional nor symmetric.[33] On the other hand, the so-called λ-hypothesis,[14] if stated in the form

$$\sum_k\lambda_{kn}(t)=\sum_n\lambda_{kn}(t)=1,$$

is always satisfied in this semiclassical theory, in consequence of the unitary character of U.[34]

In (17.17) we may again transfer all terms containing $\bar{\rho}_{kk'}$ to the left-hand side[35]:

$$\frac{\partial\bar{\rho}_{kk'}}{\partial t}+\left[\frac{1}{T_{kk'}}+i(\delta\omega_k-\delta\omega_{k'})\right]\bar{\rho}_{kk'}$$
$$=\text{"driving forces,"} \quad (17.21)$$

where (17.11) holds, but in place of (17.12) we now have

$$\gamma_k+i\delta\omega_k=\sum_p(kk|pp). \quad (17.22)$$

The quantities γ_k and $\delta\omega_k$ are defined to be real. We interpret these relations as follows. In consequence of the random perturbations, the energy of state k is uncertain by an amount γ_k (in frequency units), and in addition its average position is shifted by an amount $\delta\omega_k$. Because of this uncertainty in energy, different possible states of the array drift out of phase with each other, and the off-diagonal element $\bar{\rho}_{kk'}$ tends to relax to zero with a relaxation time $T_{kk'}$. The term

$$\gamma_{kk'}=(kk'|kk')+(k'k|k'k)$$
$$=\int_{-\infty}^{\infty}\langle V_{kk}(t)V_{k'k'}(0)\rangle dt \quad (17.23)$$

corrects for the fact that there may be correlations between the "instantaneous level shifts" $V_{kk}(t)$, $V_{k'k'}(t)$ so that the contributions of the level widths γ_k, $\gamma_{k'}$ to the rate of relaxation are not independent. Due to the terms $\gamma_{kk'}$ the uncertainty in energy γ_k is different from the reciprocal of the mean lifetime of state k against transitions. The predicted line widths are, of course, the reciprocals of the relaxation times $T_{kk'}$.

The symbols $(kk|pp)$ may be expressed in terms of the spectral density of $V_{kp}(t)$. Inverting the Fourier transform (17.19) and substituting the result into (17.5), (17.6), we obtain

$$(kk|pp)=\pi I_{kp}(\omega_{pk})+iP\int_{-\infty}^{\infty}\frac{I_{kp}(\omega)d\omega}{\omega-\omega_{pk}}, \quad (17.24)$$

where P stands for the Cauchy principal value. Thus the level widths depend on the spectral density at the transition frequencies, while the level shifts depend mainly on the manner in which the spectral density varies near the transition frequencies. This can be stated in simpler form in the usual case where $V_{kp}(t)=Q_{kp}f(t)$, where Q_{kp} is constant, and $f(t)$ is a real random function. Let $\varphi(t)$ be the autocorrelation function of $f(t)$; then the level widths and level shifts are proportional to the cosine and sine transforms of $\varphi(t)$:

$$\gamma_k=\sum_p|Q_{kp}|^2\int_0^{\infty}\cos(\omega_{kp}t)\varphi(t)dt,$$
$$\delta\omega_k=\sum_p|Q_{kp}|^2\int_0^{\infty}\sin(\omega_{kp}t)\varphi(t)dt. \quad (17.25)$$

From this we see that the level shifts will be small compared to the level widths if $\varphi(t)$ becomes vanishingly small before $\sin(\omega_{kp}t)$ reaches its first maximum. This, however, is just the condition for validity of (17.8). Thus, whenever the correlation time τ_c is so long that (17.17) is required instead of (17.8) one may expect appreciable level shifts.

If the quantities $\omega_{kn}\tau$ are of order unity, neither of the differential Eqs. (17.17), (17.8) is applicable. In fact, it is clear already from the rigorous expression $\rho(t)=\mathcal{G}(t,t')\rho(t')$ that in general a relaxation process cannot be described by any differential equation, for the rate of change of ρ does not depend only on its momentary value, but is a functional of past conditions during the entire interval $(t'\rightarrow t)$. Thus, the formulation in terms of differential equations is fundamentally inappropriate. It is convenient in those special cases where it can be justified, because of the easy interpretation in terms of relaxation times and level shifts. However, the quantities necessary for comparison with experiment can always be inferred directly from (17.7), the validity of which does not depend on the magnitudes of the quantities $\omega_{kn}\tau$.[34]

The symmetry of the transition probabilities given by (17.18) arises only because the $V_{kn}(t)$ are here considered numbers. If in better approximation one

[33] A trivial exception occurs if the system has only two linearly independent states, for a (2×2) unitary matrix necessarily satisfies $|U_{12}|^2=|U_{21}|^2$. This is not true in any higher dimensionality.

[34] The possibility that λ_{kn} is not proportional to t may lead in some cases to a differential equation for $\bar{\rho}$ with time-dependent coefficients, analogous to Eq. (2.24) of reference 31.

[35] If there is no degeneracy and the level spacing is the most general type for which there is no relation of the form $\omega_{kn}=\omega_{k'n'}$ for $k\neq k'$, the right-hand side of (17.21) is zero for all off-diagonal elements $\rho_{kk'}$.

takes into account the quantum nature of the surroundings, they must be considered as operators which operate on the state vector of the perturbing system σ_2 (the "heat bath"). Then, as shown by Ayant,[29] the definition of correlation functions (17.5) remains valid, provided the brackets are now interpreted as standing for the expectation value taken over the system σ_2, and the differential Eq. (17.8) or (17.17) then represents an approximation in which mutual correlations between the two systems are discarded at intervals τ, in the manner of (11.5). One now finds that the probabilities of upward and downward transitions are no longer equal. In the treatment of Ayant, the question of equality of these transition probabilities is reduced to the question whether the spectral density of the perturbing forces is the same at frequencies $(+\omega)$ and $(-\omega)$. This is correct provided one always defines the perturbing terms to be real, as in (17.25); note, however, that the symmetry of transition probabilities in (17.18) does not require that the spectral density of $V_{kn}(t)$ be an even function of frequency. It is sufficient if the spectral density of V_{kn} at frequency $(+\omega)$ is equal to that of V_{nk} at $(-\omega)$, and this is always the case if V is Hermitian.

If one assumes a Boltzmann distribution for the heat bath and neglects the effect of the system of interest σ_1 in modifying this distribution, the solution of (17.17) tends to another Boltzmann distribution corresponding to the same temperature.[27,30] Treatment of this case and that of "secular equilibrium" from the subjective point of view will be considered in a later paper.

18. CONCLUSION

The foregoing represents the first stage of an attempt to provide a new foundation for the predictive aspect of statistical mechanics, in which a single basic principle and method applies to all cases, equilibrium or otherwise.

The phenomenon of nuclear spin relaxation is a particularly good one to serve as a guide to a general theory of irreversible processes. It is complicated enough to require most of the techniques of a general theory, but at the same time it is simple enough so that in many cases the calculations can be carried out explicitly. Nuclear induction experiments, in which the predictions of the Bloch-Wangsness theory[27,30,31] are verified down to fine details,[36] provide a good illustration of many of the above remarks. Here the experiments are performed on samples containing of the order of 10^{20} nuclei, and one measures the time dependence of their total magnetic moment when subject to various applied fields. In the theory, however, one usually calculates a density matrix $\rho_1(t)$ which operates only in the function space of a single spin, or of some small aggregate of spins such as those attached to a single molecule. The possibility of predicting mutual properties of different spin units is therefore lost.

[36] J. T. Arnold, Phys. Rev. **102**, 136 (1956); W. A. Anderson, Phys. Rev. **102**, 151 (1956).

It would, however, always be better in principle to adopt the "global" view in which the entire assemblage of spins in the sample is the system treated. To the extent that different molecular units behave independently, the complete density matrix ρ thus obtained would be a direct product of a very large number of matrices. However, this would hardly ever be true because some correlations between different spin units would be expected. Thus, the question is raised whether, and to what extent, predictions made only from ρ_1 can be trusted. At first glance it seems that they could not be, for in most cases the density matrix $\rho_1(t)$ differs only very slightly from a multiple of the unit matrix, and thus represents a very "broad" probability distribution. According to the discussions of maximum-entropy inference in I and the introduction to the present paper, it would appear that this is a case where the theory fails to make any definite predictions, so that unless the probabilities in ρ_1 could be established in the objective sense, the calculations of Sec. 17 would be devoid of physical content.

The thing which rescues us from this situation is, of course, the fact that the experiments refer not to a single spin unit, but to a very large number of them. We must not, however, jump to the obvious conclusion that the "law of large numbers," or the central limit theorem,[19] automatically restores reliability to our predictions. To do so would be to make the logical error of the experimenter who thought that he could add three significant figures to his measurements merely by repeating them a million times. The correctness of the usual calculations can be demonstrated without explicit reference to the laws of large numbers, by application of the principles of Sec. 11. This is, in fact, the example *par excellence* of how much a prediction problem can be simplified by discarding irrelevant information.

Suppose that we had solved the problem from the global viewpoint, obtained the complete density matrix $\rho(t)$, and demonstrated that it gave a sharp distribution, and therefore reliable predictions, for the total magnetic moment $\mathbf{M}=\mathbf{M}_1+\mathbf{M}_2+\cdots+\mathbf{M}_N$. Then the only thing of further interest would be the value of $\langle\mathbf{M}\rangle$. According to Sec. 11, this can be calculated as well from the direct product matrix

$$\rho_1\times\rho_2\times\cdots\times\rho_N,$$

where ρ_k is the projection of ρ onto the space of the kth system. If the small systems are equivalent, the $\langle\mathbf{M}_k\rangle$ are all equal, and thus we obtain

$$\langle\mathbf{M}\rangle=\mathrm{Tr}(\rho\mathbf{M})=N\ \mathrm{Tr}(\rho_1\mathbf{M}_1).$$

This equation is exact regardless of whether correlations exist. Thus, *if ρ_1 embodies all of the available information about a single spin system, the predictions of total moment of N systems obtained from it are just as reliable as are those obtained from the global density matrix ρ.* We cannot estimate this reliability from ρ_1 alone; loss of that information is part of the price we had to pay for

simplification of the problem. If correlations between different spin units are strong, it will of course be very difficult to obtain ρ_1 without first solving a larger problem. Thus, in practice one will obtain only an approximate value of ρ_1; however, a one percent error in the calculated value of $\langle \mathbf{M}_1 \rangle$ leads only to a one percent error in $\langle \mathbf{M} \rangle$.

APPENDIX A. SUBJECTIVE H THEOREM

Consider the density matrix (12.5) with $t'=0$; at any particular time there exists a unitary matrix $V(t)$ which diagonalizes $\rho(t)$, so that (12.5) may be written in terms of the diagonal matrices,

$$d(t)=\sum_\alpha P_\alpha W_\alpha d(0) W_\alpha^{-1}, \tag{A.1}$$

where

$$W_\alpha = V(t)U^{(\alpha)}(t,0)V^{-1}(0) \tag{A.2}$$

is a unitary matrix. The eigenvalues $d_m(t)$ of $\rho(t)$ are thus related to the eigenvalues of $\rho(0)$ by

$$d_m(t)=\sum_n B_{mn} d_n(0), \tag{A.3}$$

where the quantities B_{mn} form a doubly stochastic matrix:

$$\sum_m B_{mn}=\sum_n B_{mn}=1. \tag{A.4}$$

The first of the inequalities (12.6) is then proved as follows:

$$\begin{aligned} S(t)-S(0) &= \sum_n d_n(0)\ln d_n(0) - \sum_m d_m(t)\ln d_m(t) \\ &= \sum_{mn} B_{mn} d_n(0)\ln[d_n(0)/d_m(t)] \\ &\geq \sum B_{mn}[d_n(0)-d_m(t)] = 0. \end{aligned} \tag{A.5}$$

Here use has been made of the fact that $\ln x \geq (1-x^{-1})$, with equality if and only if $x=1$. Thus, the equality sign in (A.5) holds if and only if $B_{mn}=0$ for each combination of m, n for which $d_n(0) \neq d_m(t)$. If $\rho(0)$ is nondegenerate, this means that the eigenvalues $d_m(t)$ must be a permutation of the $d_n(0)$.

The second of the inequalities (12.6) follows from the fact that for any given density matrix ρ, the "array entropy" S_A of Eq. (7.14) attains its minimum value, equal to $S=-\mathrm{Tr}(\rho \ln \rho)$ for the orthogonal array. To prove this, let the orthogonal array be the one with N states, where the state v_n has probability d_n, and let $\{\psi_m, w_m\}$ be any other array with M states, where $M \geq N$, which leads to the same density matrix. The two arrays are related by a transformation of the form (7.9)

$$\psi_m w_m^{\frac{1}{2}} = \sum_n v_n d_n^{\frac{1}{2}} U_{nm},$$

where U_{nm} is an $(M \times M)$ unitary matrix, and we define $d_n \equiv 0$, $N < n \leq M$. From this and the orthogonality of the v_n it follows that

$$w_m = \sum_n C_{mn} d_n, \tag{A.6}$$

where $C_{mn}=|U_{mn}|^2$ is a doubly stochastic matrix, and thus by the previous argument (A.5),

$$S \leq S_A. \tag{A.7}$$

Now in the case considered here, let $\rho(0)$ be represented by its orthogonal array $\{v_n(0), d_n(0)\}$. At time t, the density matrix (12.5) is represented by the array in which the state

$$\psi_{\alpha n}(t) = U^{(\alpha)}(t,0) v_n(0)$$

has probability $w_{\alpha n}=P_\alpha d_n(0)$. The array entropy is thus

$$S_A(t) = -\sum_{\alpha n} w_{\alpha n} \ln w_{\alpha n} = S(0)+S(P_\alpha) = \text{const}, \tag{A.8}$$

which, together with (A.7), proves the theorem.