Theoret. Chim. Acta (Berl.) 41, 95–117 (1976)

Original Investigations

The Principle of Increasing Mixing Character and Some of Its Consequences

Ernst Ruch

Institut für Quantenchemie der Freien Universität Berlin

Alden Mead

Chemistry Department, University of Minnesota at Minneapolis

Received November 12, 1975

The "Principle of Increasing Mixing Character", previously postulated by one of us (and derived for the case of an ensemble of isolated systems obeying a "master equation") as a stronger version of the second law of thermodynamics, is re-derived using von Neumann's density matrix formulation of statistical mechanics. To make the principle more convenient for applications, it is reformulated in terms of "Mixing Homomorphic Functions", a set of state functions all of which must increase in an allowed irreversible process in an isolated system. The entropy is one such function, but no one function, and no finite set of functions, suffices to determine the increase of mixing character. The principle is extended to the case of a system which is not isolated, but in contact with a heat bath, for which it takes a form which we name the "Principle of Decreasing Mixing Distance" from the equilibrium distribution. As examples, applications are made to two simple cases: diffusion in an ideal solution, and chemical reactions in ideal gas mixtures.

Key word: Irreversible process

1. Introduction

The time development of a Gibbs (microcanonical) ensemble of isolated systems proceeds in such a way that the mixing character increases monotonically.

The foregoing is a statement of the "Principle of Increasing Mixing Character", first obtained in an earlier paper by one of us [1]. Its meaning may be understood as follows:

If we denote the probability (or population, depending on normalization) of the α'th state by p_α, then a statistical distribution is characterized by $\{p_\alpha\}$, the set of all these quantities. We associate with $\{p_\alpha\}$ a "diagram distribution" $\gamma = \{p_j\}$, which consists of the same numbers arranged in nonincreasing order:[1]

$$p_1 \geqslant p_2 \geqslant p_3 \geqslant \cdots$$

---

[1] Throughout this paper, probabilities with Latin indices are assumed to be arranged in nonincreasing order, with p_j thus denoting the probability of the j'th most probable state. Greek indices indicate that the order is arbitrary, with p_α denoting the probability of state α. At times, both types of indices will be used with the same distribution in different equations.

Different probability distributions which correspond to the same diagram distribution are called “mixing equivalent”. Such distributions differ from one another by permutations of the probabilities p_α among the states α.

The mixing character marks a property which is the same for mixing equivalent distributions and is therefore sufficiently determined, if defined for diagram distributions. Let $m[\gamma]$ denote the mixing character of γ, the relations “greater than or equal to” be expressed through “$\succ$”: then the increase of mixing character is defined as follows:

$$\underset{r}{\forall}\left(\sum_{j=1}^{r} p'_j \leqslant \sum_{j=1}^{r} p_j\right) \Leftrightarrow m[\gamma'] \succ m[\gamma]. \tag{1}$$

In part B of Ref. [1], which we refer to from now on as [1B], the concept of mixing character was arrived at through the analysis of the role of row and column partitions in Young diagrams if they are used to represent classifications of finite sets. For Young diagrams, however, a partial ordering and lattice structure was defined and studied five years ago [2]. This structure has relevance in mathematics, as was pointed out in part A of Ref. [1]. It turned out that this partial ordering corresponds precisely to the relation “more mixed than”, according to

$$m[\gamma'] \succ m[\gamma] \Leftrightarrow \gamma' \subset \gamma. \tag{2}$$

If γ and γ' are two different Young diagrams with the same total number of boxes, then the relation $\gamma' \subset \gamma$, according to Ref. [2], can be defined in three equivalent ways, viz:

γ' is obtainable from γ by moving boxes exclusively downward and to the left. (3a)

$$\sum_{j=1}^{r} \nu'_j \leqslant \sum_{j=1}^{r} \nu_j \tag{3b}$$

for all r, where the ν_j and ν'_j are the row lengths of γ and γ' respectively, arranged in nonincreasing order.

$$\sum_{j=1}^{r} \mu'_j \geqslant \sum_{j=1}^{r} \mu_j \tag{3c}$$

for all r, where the μ_j and μ'_j are the column lengths of γ and γ' respectively, arranged in nonincreasing order.

The close analogy between (3b) and definition (1) above is evident at a glance, and becomes precise if, for an ensemble with a finite number of members, one simply identifies the populations of the various states with the row lengths of a Young diagram.

It is important to realize that the “greater than” relation $\gamma \supset \gamma'$ is a *partial* ordering of distributions or diagrams, and that there can be pairs of distributions which are not mixing equivalent and also not comparable with one another as regards mixing character, i.e., such that neither is more mixed than the other. For example, if γ has $p_1 = p_2 = \frac{1}{2}$, all others zero, and γ' has $p'_1 = \frac{2}{3}$, $p'_2 = p'_3 = \frac{1}{6}$, others zero, one easily verifies from definition 1 that neither is more mixed than the other. The diagram analog of this is the pair of diagrams with row lengths (3, 3) and (4, 1, 1), for which the definitions (3a, b, c) all lead to the conclusion that neither is greater than the other.

In [1B], it was shown that increase of the Shannon entropy $S = -\sum_\alpha p_\alpha \ln p_\alpha$ is a necessary, but not sufficient, condition for increasing mixing character. It was also shown that the mixing character always increases in ensembles of isolated systems obeying a "master equation". It thus seems probable that the principle of increasing mixing character may hold with considerable generality, i.e., that in allowed irreversible processes for an isolated system the mixing character always increases. If true, this principle would afford a strengthening of the second law of thermodynamics, since it would permit one to rule out processes, allowed by conventional thermodynamics, for which the entropy increases but the mixing character does not. For example, in the case of the distributions $\gamma = (\frac{1}{2}, \frac{1}{2}, 0, 0, \ldots)$ and $\gamma' = (\frac{2}{3}, \frac{1}{6}, \frac{1}{6}, 0 \ldots)$ cited above, it is easily verified that the latter has the greater Shannon entropy. Thus, a transition from a distribution characterized by γ to one characterized by γ' would be allowed by conventional thermodynamics, but ruled out by the principle of increasing mixing character.

In Section 2 of the present paper, we present a second argument for the principle of increasing mixing character, which does not depend on the "master equation", but is based on von Neumann's density matrix formulation of the second law. It is perhaps a matter of taste which of these two arguments is to be preferred; together, however, we find them quite persuasive. Thus fortified in our belief that the principle does in fact hold with great generality, we proceed in the following sections to reformulate and develop it in a way convenient for applications. Finally, we give two very simple applications. In order to understand the present paper properly, it is desirable (though not absolutely essential in a first reading) that the reader be acquainted with [1B].

2. Derivation with Density Matrix Formulation

We consider a Gibbsian ensemble described by a quantum mechanical density (statistical) matrix ρ. According to von Neumann [3], the act of measurement corresponds to replacing ρ by a new statistical matrix $\rho^{(d)}$ consisting only of the diagonal elements of ρ in the representation determined by the measurement

$$\rho \to \rho^{(d)}, \quad \rho^{(d)}_{\alpha\beta} = \rho_{\alpha\alpha}\delta_{\alpha\beta} = p'_\alpha \delta_{\alpha\beta} \tag{4}$$

where p'_α is the probability that a randomly chosen member of the ensemble is in the state α after the measurement. The measurement, in other words, simply determines how many ensemble members are in each eigenstate of the variable measured, and leaves the phases between these states random. It was shown by von Neumann that the process (4) always leads to an increase in the entropy, defined here by

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

We now proceed to show that the process (4) also leads to an increase in the mixing character.

Since the matrix ρ is related to its diagonal form $\tilde{\rho}$ by a unitary transformation:

$$\rho = U\tilde{\rho}U^{\dagger}, \quad \tilde{\rho}_{\mu\nu} = p_{\mu}\delta_{\mu\nu}$$

it follows that the new probabilities after the process (4) are given by

$$p'_{\alpha} = \rho_{\alpha\alpha} = \sum_{\mu,\nu} U_{\alpha\mu}\tilde{\rho}_{\mu\nu}U^{\dagger}_{\nu\alpha} = \sum_{\mu} B_{\alpha\mu}p_{\mu} \tag{5}$$

where

$$B_{\alpha\mu} = |U_{\alpha\mu}|^2.$$

From the unitarity of U, one sees immediately that the matrix B is "bistochastic", i.e., that all its elements are real and non-negative and that

$$\sum_{\mu} B_{\mu\beta} = \sum_{\beta} B_{\mu\beta} = 1.$$

We now invoke a theorem due to Hardy, Littlewood and Pólya [4], which was also used in [1B]. It states that, for two probability distributions $\{p_{\alpha}\}$ and $\{p'_{\alpha}\}$,

$$p'_{\alpha} = \sum_{\beta} B_{\alpha\beta}p_{\beta} \quad \Leftrightarrow \quad \forall_{r}\left(\sum_{j=1}^{r} p'_{j} \leqslant \sum_{j=1}^{r} p_{j}\right) \tag{6}$$

where B is bistochastic.

Because of (6), Eq. (5) expresses the fact that the distribution $\{p'_{\alpha}\}$ is more mixed than $\{p_{\alpha}\}$. Therefore the principle of increasing mixing character holds whenever the time development of a system is described by process (4).

A. Uhlmann has defined a "more mixed"-relation between density matrices [5] that corresponds precisely to the one which we derived in terms of Young diagrams [1, 2]. Uhlmann's definition can be presented in the following form:

$$m[\rho'] \succ m[\rho] \quad \Leftrightarrow \quad \rho' = \sum_{\lambda} c_{\lambda}U^{(\lambda)}\rho U^{(\lambda)\dagger} \tag{7}$$

where the $U^{(\lambda)}$ are unitary and $c_{\lambda} \geqslant 0$, $\sum_{\lambda} c_{\lambda} = 1$.

Eq. (4) is easily seen to be a special case of the equation on the right-hand side of (7). This equation, however, may also be interpreted as describing a process that connects ρ' at a later time with ρ at an earlier time. A process of this type includes cases in which the loss of informations is caused by unknown perturbations unrelated to measurement, or by a measurement of less than a complete set of commuting observables, which only partially wipes out the phases. With reference to this interpretation, definition (7) expresses the fact that processes of this more general type also effect an increase of the mixing character, since one can easily show by exactly the same method of proof used above, that the right-hand side of (7) also implies that ρ' is more mixed than ρ in our sense.

The arguments presented here, together with those given in [1B], encourage us in the belief that the range of validity of the principle of increasing mixing character is as wide as that of thermodynamics itself, or nearly so. Neither derivation, for example, depends on any special properties of the Hamiltonian of the system, such as its time-independence, or symmetry properties. Of course, we have not removed the difficulties associated with the transition from mechanics (which is time-reversal invariant) to thermodynamics (which is not). However, the difficulties do not take any more severe a form in the present case than in attempted derivations of the entropy principle: the assumptions which one needs to make to derive the entropy principle also lead to increasing mixing character.

We therefore feel justified in postulating that the principle of increasing mixing character holds with a generality close to that of thermodynamics itself.

3. Representation of the Mixing Character through Functions

3.1 *Mixing Homomorphic Functions*

For purposes of applications, it is useful to have functions of the probabilities making up a distribution which never decrease when the mixing character increases[2]. Such functions were called "mixing homomorphic" in [1B], because each one can be understood as a homomorphism of the partially ordered mixing character onto the ordered real numbers. One example of such a function is, of course, the entropy. Entropy increase, however, is only a necessary, not a sufficient, condition for increasing mixing character. Thus, by defining more mixing homomorphic functions and requiring them to increase, we can derive new inequalities that must be obeyed by allowed irreversible processes.

Consider therefore a function of the form $m(g;\gamma)=\sum_\alpha g(p_\alpha)$ or in density matrix notion $m(g;\rho)=\operatorname{Tr} g(\rho)$, where $g(x)$ is a continuous function of x. $m(g;\gamma)$ will be mixing homomorphic if and only if

$$m[\gamma'] \succ m[\gamma] \quad \Longrightarrow \quad m(g;\gamma') \geqslant m(g;\gamma). \tag{8}$$

According to Ref. [1B] the idea of increasing mixing character is based on the effect of mixing, and mixing operators have been shown to be represented by bistochastic matrices. The definition of increasing mixing character, related to this idea (Def. 1e in Ref. [1B]), is given through

$$m[\gamma'] \succ m[\gamma] \quad \Leftrightarrow \quad p'_\alpha = \sum_\beta B_{\alpha\beta} p_\beta \text{ where } B \text{ is bistochastic;}^3 \tag{9}$$

therefore, (8) means that $m(g;\gamma)$ will be mixing homomorphic if and only if, for all p_α, and bistochastic B, the function g satisfies

$$\sum_\alpha g\left(\sum_\beta B_{\alpha\beta} p_\beta\right) \geqslant \sum_\alpha g(p_\alpha). \tag{10}$$

To see what (10) implies, we first consider a special case. Suppose that $p_1 \neq p_2$, $p'_1 = p'_2 = \frac{1}{2}(p_1+p_2)$, $p'_\alpha = p_\alpha$, $\alpha > 2$. This corresponds to $B_{11}=B_{12}=B_{21}=B_{22}=\frac{1}{2}$, $B_{\alpha\alpha}=1$, $\alpha>2$, all other $B_{\alpha\beta}=0$. For this case, (10) becomes, after subtracting off terms which are equal on both sides, and dividing by 2,

$$g\left(\tfrac{1}{2}(p_1+p_2)\right) \geqslant \tfrac{1}{2}\left(g(p_1)+g(p_2)\right) \text{ for all } p_1, p_2. \tag{11a}$$

---

[2] To save writing, we will sometimes simply say that such a function "must increase". In such cases, it should be understood that it may also remain unchanged.

[3] That this is an equivalent definition of increasing mixing character also follows from the theorem of Hardy, Littlewood, and Pólya [4] already cited in Sect. 2.

Eq. (11a), however, is just one definition of a convex function, a function such that the midpoint of every chord lies on or below the curve of the function [6]. Hence, (10) can only be satisfied if g is convex. If g is even piecewise continuous (11a) is equivalent to

$$g(\sum_{\beta} c_{\beta} p_{\beta}) \geqslant \sum_{\beta} c_{\beta} g(p_{\beta}) \tag{11b}$$

for arbitrary $c_{\beta} \geqslant 0$ with $\sum_{\beta} c_{\beta} = 1$. Because of the identity

$$\sum_{\alpha} \sum_{\beta} B_{\alpha\beta} g(p_{\beta}) = \sum_{\alpha} g(p_{\alpha})$$

(provable by summing first over α on the left-hand side), Eq. (10) can be transformed into

$$\sum_{\alpha} g\,(\sum_{\beta} B_{\alpha\beta} p_{\beta}) \geqslant \sum_{\alpha} (\sum_{\beta} B_{\alpha\beta} g(p_{\beta})).$$

Because of (11b), this is satisfied term by term if g is convex, since the $B_{\alpha\beta}$ with fixed α satisfy the criteria for c_{β} in (11b). We have thus proved

Theorem 1. A function $m(g;\gamma)$ is mixing homomorphic if and only if g is convex.

Either (11a) or (11b) may be used as a definition of convexity. If $g(p)$ is twice differentiable, an equivalent definition is $g''(p) \leqslant 0$.

To get an intuitive feel for the meaning of the requirement that all $m(g;\gamma)$ must increase, consider a process in which an infinitesimal probability δp is transferred from state α to state β. For this process,

$$\delta m(g;\gamma) = \{g'(p_{\beta}) - g'(p_{\alpha})\}\delta p$$

which, with convex g, will be positive only if $p_{\alpha} > p_{\beta}$. Thus, our principle in this case is equivalent to the common-sense requirement that less populated (probable) states gain population at the expense of more populated ones.

Any convex $g(p)$ defines a mixing homomorphic function $m(g;\gamma)$, and thus leads to requirements on allowed thermodynamic processes. In particular cases, one may wish to choose g's which lead to the strongest results, or those which require the least effort to evaluate the m's. The best-known example, of course, is the first of the following three examples (12a, b, c). Functions of the second type may be used for purposes of computational convenience. The third class of functions will be discussed in detail in the subsequent subsection:

$$\begin{aligned} &g_s(p) = -p \ln p \\ &m(g_s;\gamma) = S(\gamma), \text{ the Shannon entropy;} \end{aligned} \tag{12a}$$

$$\begin{aligned} &g_n(p) = -p^n, \quad n > 1 \\ &m(g_n;\gamma) = -\sum_{\alpha} p_{\alpha}^n; \end{aligned} \tag{12b}$$

$$\begin{aligned} &g_{\lambda}(p) = \frac{\lambda - p}{2} - \left|\frac{\lambda - p}{2}\right| = \begin{cases} 0 & \text{for } p \leqslant \lambda \\ \lambda - p & \text{for } p > \lambda \end{cases} \quad (\lambda \geqslant 0) \\ &m(g_{\lambda};\gamma) = -\sum_{\{\alpha \mid p_{\alpha} \geqslant \lambda\}} (p_{\alpha} - \lambda). \end{aligned} \tag{12c}$$

3.2 Mixing Isomorphic Sets of Functions

No single mixing homomorphic function is sufficient to determine the increase of the mixing character; however, it may happen that a set of such functions, all of which are required to increase, is sufficient. To investigate this possibility, we make the following definition:

Definition 1. A set of mixing homomorphic functions $\{m(g;\gamma)\}$ is called mixing isomorphic if, for every γ, γ' with $m[\gamma']\nsucc m[\gamma]$, there is at least one $m(g;\gamma)$ in $\{m(g;\gamma)\}$ for which $m(g;\gamma')<m(g;\gamma)$.

Thus, if a mixing isomorphic set exists, the requirement that all its members increase is a sufficient as well as a necessary condition for increasing mixing character. Any mixing isomorphic set of functions may serve as definition of the partial ordering in the set of diagram distribution.

We now state and prove

Theorem 2. The functions $m(g_\lambda;\gamma)$ defined by (12c) with $0\leqslant\lambda\leqslant 1$ form a mixing isomorphic set.

What needs to be proved is the conclusion from the left to the right side of (13):

$$\underset{\lambda}{\forall}\left(\sum_{\{\alpha\,|\,p'_\alpha\geqslant\lambda\}}(p'_\alpha-\lambda)\leqslant\sum_{\{\alpha\,|\,p_\alpha\geqslant\lambda\}}(p_\alpha-\lambda)\right)\ \Leftrightarrow\ \underset{r}{\forall}\left(\sum_{j=1}^{r}p'_j\leqslant\sum_{j=1}^{r}p_j\right). \tag{13}$$

To show this, we choose $\lambda=p_r$ and note that

$$\begin{aligned}\sum_{\{\alpha\,|\,p_\alpha\geqslant p_r\}}(p_\alpha-p_r)&=\sum_{j=1}^{r}(p_j-p_r)=-rp_r+\sum_{j=1}^{r}p_j\\ \sum_{\{\alpha\,|\,p'_\alpha\geqslant p_r\}}(p'_\alpha-p_r)&\geqslant\sum_{j=1}^{r}(p'_j-p_r)=-rp_r+\sum_{j=1}^{r}p'_j.\end{aligned} \tag{14}$$

The inequality between the two sums in (14) follows since, whenever both sums differ, the right-hand sum either does not contain certain members $(p'_\alpha-p'_r)$ contained in the left sum, all of which are positive; or it contains members $(p'_\alpha-p_r)$ not contained in the left-hand sum, all of which are negative. From (14) follows the conclusion from left to right in (13). Since the conclusion in the opposite direction is obvious, theorem 2 is proved.

From theorem 1 and 2 one immediately obtains the

Corollary. The set of all $m(g;\gamma)$ with arbitrary convex g is a mixing isomorphic set.

In the appendix some necessary conditions for mixing isomorphic sets are derived.

Eq. (13) represents two equivalent definitions of the partial ordering of the mixing character. With the abbreviations

$$\begin{aligned}o_r(\gamma)&=\sum_{i=1}^{r}p_i\\ u_\lambda(\gamma)&=1-\sum_{\{\alpha\,|\,p_\alpha\geqslant\lambda\}}(p_\alpha-\lambda)\end{aligned} \tag{15}$$

we may write:

$$m[\gamma'] \succ m[\gamma] \;\Leftrightarrow\; \gamma' \subset \gamma \;\Leftrightarrow\; \underset{r}{\forall}(o_r(\gamma') \leqslant o_r(\gamma)) \;\Leftrightarrow\; \underset{\lambda}{\forall}(u_\lambda(\gamma') \geqslant u_\lambda(\gamma)). \tag{16}$$

A diagram distribution γ is illustrated in Fig. 1a, where the shaded area is assumed to be 1 as the condition $\sum_j p_j = 1$ holds. The $o_r(\gamma)$ and $u_\lambda(\gamma)$ correspond to the areas shaded in Fig. 1b and 1c respectively. This illustration indicates the comparison with the case of diagrams with a finite number of boxes. If we replace p_i by v_i/n with v_i representing the numbers of boxes in each row of a diagram with n boxes, then $no_r(\gamma)$ and $nu_\lambda(\gamma)$ with $\lambda = 1/n, 2/n, \ldots n/n$ denote the partial sums of rows and columns respectively. Thus (16) represents besides (3b) the definition (3c) without the limitation to diagrams with a finite number of boxes.

Fig. 1a

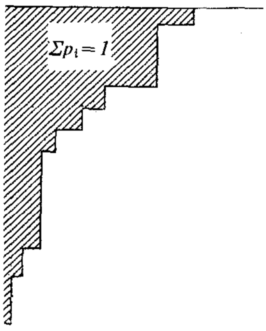


Fig. 1b

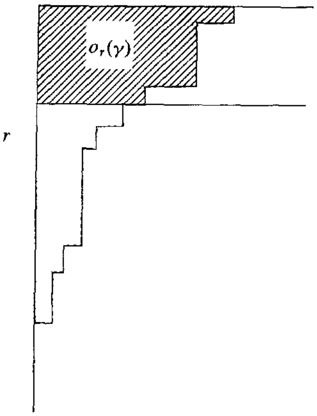
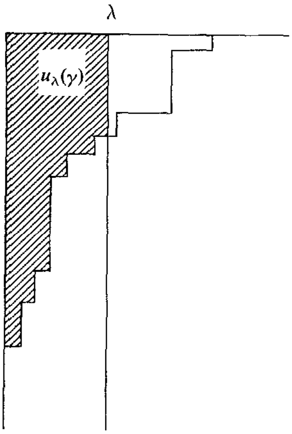


Fig. 1c



In Ref. [1B], the row and column partitions of a diagram have been associated with the concepts of statistical order and disorder. The mixing character criteria according to (16) refer to both these aspects also in the case where general probability distributions are concerned. The advantage of having at hand mixing isomorphic sets of functions will become evident, if we try to transfer the principle of increasing mixing character to systems which are not isolated. We shall now discuss this problem.

4. System in Interaction with Bath

Up to now, we have been concerned exclusively with isolated systems, while most practical applications are concerned with systems in weak interaction with some sort of bath. In conventional thermodynamics, one can start with the entropy principle for an isolated system, apply it to a combined system consisting of bath and system of interest, and derive results such as the decrease of Gibbs or Helmholtz free energy for systems with constant temperature and pressure or volume. In this section, we carry out the analogous derivation for the principle of increasing mixing character.

Consider, therefore, a system Σ which can be mentally divided into two parts: a system of interest A, and a bath B, in weak interaction with one another. Denote the states of A by α, β, etc., those of B by capital letters Φ, etc. The direct product states $|\alpha\Phi\rangle = |\alpha\rangle|\Phi\rangle$ form a complete set for Σ. In most cases, only certain states will be accessible, e.g., those with given total energy for Σ. This means that, in general, only certain Φ will be "compatible" with each α, i.e., only certain Φ will be accessible to B when it is given that A is in the state α. In the cases normally considered, two states α and β either possess no compatible Φ in common, or all Φ compatible with α are also compatible with β. For example, if Σ has fixed total energy W, the Φ compatible with α are those with $E_\Phi = W - E_\alpha$. If $E_\alpha = E_\beta$, all these will also be compatible with β; otherwise, none of them will. In the beginning, we shall not assume this property of "total mutual compatibility", but we shall see in subsection 4.2 that when it is not satisfied no unique equilibrium state for the bath can be defined when A is not in equilibrium.

We present three derivations, all leading to the same result, in subsections 4.1, 4.2, and 4.3. The first of these is in terms of classical probabilities, the second in terms of quantum mechanical density matrices, and the third, which makes no specific reference to the bath, is in terms of a master equation. The results are summarized in subsection 4.4. by means of the "principle of decreasing mixing distance from the equilibrium distribution". The connection of all this with conventional thermodynamics is discussed in subsection 4.5. The related problem of partial equilibrium is treated in subsection 4.6.

4.1 Classical Probability Formulation

In this subsection, we consider the system Σ to be entirely describable in terms of classical probabilities. The system of interest A is not necessarily in equilibrium: the probabilities

$$p_\alpha = \sum_\Phi p_{\alpha\Phi}$$

are fixed but arbitrary. The bath, however, must be assumed to be in equilibrium in order to obtain unambiguous results. In other words, the mixing character of the total system Σ must be maximized subject to the constraint of fixed p_α. With this assumption, our task is to express an arbitrary mixing homomorphic function for $m(g; \gamma)$ in terms of the variables of A alone.

We first maximize the mixing character of Σ subject to fixed p_α. We require

$$\delta\{m(g;\gamma) - \sum_\alpha l_\alpha p_\alpha\} = 0$$

where the l_α are Lagrange multipliers. (The convexity of g guarantees that any stationary point will be maximum.) Writing out the variation explicitly, we obtain

$$\sum_{\alpha\Phi} g'(p_{\alpha\Phi})\delta p_{\alpha\Phi} - \sum_{\alpha\Phi} l_\alpha \delta p_{\alpha\Phi} = 0.$$

Since the $\delta p_{\alpha\Phi}$ are arbitrary, this equation is satisfied for arbitrary convex g if and only if all the $p_{\alpha\Phi}$ for given α are equal. We therefore have

$$p_{\alpha\Phi}=\frac{p_\alpha}{\Omega_\alpha}$$

where Ω_α is the number of bath states compatible with α. Mixing homomorphic functions of the system Σ according to (8) can now be expressed as sums over states of the system A alone,

$$m(g;\gamma)=\sum_{\alpha\Phi} g(p_{\alpha\Phi})=\sum_{\alpha\Phi} g\left(\frac{p_\alpha}{\Omega_\alpha}\right)=\sum_{\alpha}\Omega_\alpha g\left(\frac{p_\alpha}{\Omega_\alpha}\right).$$

By introducing the equilibrium probabilities w_α of the system A, that are given by

$$w_\alpha=\frac{\Omega_\alpha}{\Omega}\quad\text{where } \Omega \text{ is the total number of accessible states in } \Sigma,$$

we arrive at:

$$m(g;\gamma)=\Omega\sum_\alpha w_\alpha g\left(\frac{p_\alpha}{w_\alpha\Omega}\right)=\sum_\alpha w_\alpha G\left(\frac{p_\alpha}{w_\alpha}\right)$$

where $G(x)=\Omega g(x/\Omega)$ is again an arbitrary convex function of its argument, which however, now ranges from zero to infinity. Thus mixing homomorphic functions of the system Σ expressed in terms of variables, that refer to the system A alone, may be written as follows:

$$a(G;p,w)=\sum_\alpha w_\alpha G\left(\frac{p_\alpha}{w_\alpha}\right). \tag{17}$$

Our principle now states that, for a system in interaction with a bath, all the $a(G;p,w)$ for arbitrary convex G must increase in an allowed irreversible process.

4.2 *Density Matrix Formulation*

Here we treat basically the same problem as in the previous subsection, but with Σ now described by a density matrix ρ_Σ with elements $\langle\alpha\Phi|\rho_\Sigma|\beta\Phi'\rangle$, it being assumed that there exists a complete set of accessible states which are all of the form $|\alpha\Phi\rangle$. The state of A alone is described by the reduced density matrix ρ, whose elements are given by

$$\langle\alpha|\rho|\beta\rangle=\sum_\Phi\langle\alpha\Phi|\rho_\Sigma|\beta\Phi\rangle.$$

We note that, if α and β possess no compatible Φ in common, then $\langle\alpha|\rho|\beta\rangle$ is necessarily zero. As before, our task is to express a mixing homomorphic function $m(g;\rho)=\mathrm{Tr}\,g(\rho_\Sigma)$ in terms of ρ alone, assuming that the mixing character has been maximized subject to fixed ρ. We carry out this constrained maximization first, using this time a particular mixing homomorphic function $-\mathrm{Tr}\,\rho^2$. Introducing Lagrange multipliers $l_{\alpha\beta}$, we find:

$$\delta\{-\mathrm{Tr}\,\rho_\Sigma^2-\sum_{\alpha\beta}l_{\alpha\beta}\langle\alpha|\rho|\beta\rangle\}$$

$$=-2\sum_{\alpha,\Phi,\beta,\Phi'}\langle\alpha\Phi|\rho_\Sigma|\beta\Phi'\rangle\delta\langle\beta\Phi'|\rho_\Sigma|\alpha\Phi\rangle-\sum_{\alpha,\beta,\Phi}l_{\alpha\beta}\delta\langle\alpha\Phi|\rho_\Sigma|\beta\Phi\rangle=0.$$

We first equate to zero the coefficient of $\delta\langle\beta\Phi'|\rho_\Sigma|\alpha\Phi\rangle$ with $\Phi\neq\Phi'$ and obtain

$$\langle\alpha\Phi|\rho_\Sigma|\beta\Phi'\rangle=0,\quad \Phi\neq\Phi'.$$

For $\Phi=\Phi'$, we find

$$\langle\alpha\Phi|\rho_\Sigma|\beta\Phi\rangle=-\tfrac{1}{2}l_{\alpha\beta},\tag{18}$$

independently of Φ for given α, β. Since we must have

$$\sum_{\Phi}\langle\alpha\Phi|\rho_\Sigma|\beta\Phi\rangle=\langle\alpha|\rho|\beta\rangle,$$

(18) is satisfied only if

$$\langle\alpha\Phi|\rho_\Sigma|\beta\Phi\rangle=\frac{1}{\Omega_{\alpha\beta}}\langle\alpha|\rho|\beta\rangle\tag{19}$$

where $\Omega_{\alpha\beta}$ is the number of bath states compatible with both α and β, and of course, Φ itself must be compatible with both.

If instead of $g(p)=-p^2$, we had used (say) $g(p)=-p^n$, we would have obtained instead of (18),

$$\langle\alpha\Phi|\rho_\Sigma^{n-1}|\beta\Phi\rangle=-\frac{1}{n}l_{\alpha\beta}\tag{20}$$

independently of Φ for given α, β. Now consider the special case in which $\langle\alpha|\rho|\kappa\rangle$ is the only matrix element of ρ with $\langle\alpha|$ on the left, and let Φ be a state of the bath which is compatible with both α and κ, Φ' a state compatible with α but not with κ. Then we obtain from (19)

$$\langle\alpha\Phi|\rho_\Sigma^2|\alpha\Phi\rangle=\frac{1}{\Omega_{\alpha\kappa}^2}\langle\alpha|\rho|\kappa\rangle\langle\kappa|\rho|\alpha\rangle$$

and

$$\langle\alpha\Phi'|\rho_\Sigma^2|\alpha\Phi'\rangle=0,$$

which contradicts (20), according to which these two matrix elements should be equal. We conclude, then, that the bath has a definite equilibrium for each ρ (independently of the function used to determine it) only if our system has the property of "total mutual compatibility" defined in the introduction to this section. Since all the cases normally considered have this property in any case, we restrict ourselves from now on to systems of this kind. With this restriction, either

α and β have no compatible states Φ in common (in which case $\langle\alpha|\rho|\beta\rangle=0$), or $\Omega_{\alpha\beta}=\Omega_\alpha=\Omega_\beta$. Thus, Eq. (19) becomes

$$\langle\alpha\Phi|\rho_\Sigma|\beta\Phi\rangle=\frac{1}{\Omega_\alpha}\langle\alpha|\rho|\beta\rangle.$$

We can now introduce an operator σ in the space of the $|\alpha\rangle$ alone, with matrix elements

$$\langle\alpha|\sigma|\beta\rangle=w_\alpha\delta_{\alpha\beta}\,;\quad w_\alpha=\frac{\Omega_\alpha}{\Omega}.$$

σ commutes with ρ, since the latter has matrix elements only between states with compatible bath states in common, hence with the same eigenvalue of σ. σ and ρ are both in block diagonal form, and within each block, σ is just the unit matrix for the block multiplied by Ω_α/Ω. In this representation, the full density matrix ρ_Σ is also block-diagonal, the blocks being those of ρ, each multiplied by $1/\Omega_\alpha$ and repeated Ω_α times. Thus, ρ_Σ and ρ can be diagonalized simultaneously by diagonalizing the blocks of ρ_Σ separately, and we have for $m(g;\rho_\Sigma)$

$$m(g;\rho_\Sigma)=\operatorname{Tr} g(\rho_\Sigma)=\Omega\operatorname{Tr}\left(\sigma g\left(\frac{1}{\Omega}\rho\sigma^{-1}\right)\right).$$

Making the same change of scale as in subsection 4.1, we obtain functions in terms of the reduced density matrix

$$a(G;\rho,\sigma)=\operatorname{Tr}\left(\sigma G(\rho\sigma^{-1})\right)\tag{21}$$

where σ is the equilibrium reduced density matrix with eigenvalues w_α, the equilibrium probabilities, and G is again an arbitrary convex function of its argument between zero and infinity. We require that all the functions (21) increase in allowed processes, which is clearly the quantum-mechanical analog of the result of subsection 4.1.

4.3 Master Equation Formulation

In this subsection, we make no specific reference to the bath, but simply assume that our system (described by classical probabilities) is governed by a linear time-displacement operator expressing the probabilities at a later time, p'_α, in terms of those at time zero, p_α:

$$p'_\alpha=\sum_\beta M_{\alpha\beta}p_\beta.\tag{22}$$

Since the total probability must be conserved, we have

$$\sum_\alpha M_{\alpha\beta}=1.\tag{23a}$$

Also, it is assumed that there is an equilibrium distribution $\{w_\alpha\}$ which is not changed by the time-displacement operator:

$$\sum_\beta M_{\alpha\beta}w_\beta=w_\alpha.\tag{23b}$$

If the system obeys a master type equation, but with equilibrium distribution $\{w_\alpha\}$ instead of the uniform distribution, then all these conditions are satisfied.

We define the matrix μ by

$$\mu_{\alpha\beta} = w_\alpha^{-1} M_{\alpha\beta} w_\beta . \tag{24}$$

It follows from (23b) that

$$\sum_\beta \mu_{\alpha\beta} = 1. \tag{25}$$

Now let us compare the value of a function $a(G; p, w)$ before and after the time-displacement. We find

$$\begin{aligned} a(G; p', w) &= \sum_\alpha w_\alpha G\left(\frac{p'_\alpha}{w_\alpha}\right) = \sum_\alpha w_\alpha G\left(\frac{1}{w_\alpha}\sum_\beta M_{\alpha\beta} p_\beta\right) \\ &= \sum_\alpha w_\alpha G\left(\sum_\beta \mu_{\alpha\beta}\frac{p_\beta}{w_\beta}\right) \geqslant \sum_\alpha w_\alpha \sum_\beta \mu_{\alpha\beta} G\left(\frac{p_\beta}{w_\beta}\right) \\ &= \sum_{\alpha\beta} M_{\alpha\beta} w_\beta G\left(\frac{p_\beta}{w_\beta}\right) = \sum_\beta w_\beta G\left(\frac{p_\beta}{w_\beta}\right) = a(G; p, w) \end{aligned} \tag{26}$$

where we have used (23a), (24), and (25), and the inequality follows from the convexity of G and Eq. (10).

Eq. (26) is just what we wished to prove, that the functions $a(G; p, w)$ always increase for an allowed irreversible process in such a system.

4.4 Summary: Principle of Decreasing Mixing Distance

The conclusion of the three preceding subsections is that the approach to equilibrium of a system in interaction with a bath is also determined by a partial ordering, in which all the functions $a(G; p, w)$ are required to increase. Moreover, it is easy to see that all these functions are actually maximized at the equilibrium distribution $\{w_\alpha\}$. For, maximizing an arbitrary $a(G; p, w)$ subject to $\sum_\alpha p_\alpha = 1$, using the Lagrange multiplier l, we find

$$\delta\left\{a(G; p, w) - l\sum_\alpha p_\alpha\right\} = \sum_\alpha \left\{G'\left(\frac{p_\alpha}{w_\alpha}\right) - l\right\}\delta p_\alpha = 0.$$

This is satisfied for arbitrary G if and only if all the p_α/w_α have the same value, say k. That k must be unity is obvious.

If $a(G; p', w) \geqslant a(G; p, w)$ for all convex G, one may speak of the distribution $\{p'_\alpha\}$ as being "closer" to the equilibrium distribution $\{w_\alpha\}$ than is $\{p_\alpha\}$, since a path from $\{p_\alpha\}$ to $\{w_\alpha\}$ is allowed to pass through $\{p'_\alpha\}$, but not vice versa. Accordingly, our partial ordering may be described in terms of the "mixing distance", defined as follows:

Definition 2. The mixing distance of a distribution $\{p_\alpha\}$ from a reference distribution $\{w_\alpha\}$ is said to be greater than that of $\{p'_\alpha\}$ from the same reference distribution if for all convex G

$$a(G; p', w) \geqslant a(G; p, w).$$

The mixing distance does not possess all the properties of a distance as normally used in mathematics. Nevertheless, we feel justified in our choice of terminology.

By means of the mixing distance, the results of subsections 4.1, 4.2, 4.3 may be summarized as a principle, which of course, is a consequence of the principle of increasing mixing character.

Principle of decreasing mixing distance: The approach to equilibrium of a system in interaction with a bath is always in the direction of monotonically decreasing mixing distance from the equilibrium distribution.

Referring as it does to a reference distribution $\{w_\alpha\}$, the partial ordering in terms of mixing distance is different from that of mixing character in that, for example, distributions which are mixing equivalent are not necessarily equivalent as regards mixing distance from a given reference distribution. For the case of an isolated system, or for any system whose equilibrium distribution is the uniform one, the functions $a(G; p, w)$ reduce, apart from scale factors, to $m(g; \gamma)$ so that the principle of decreasing mixing distance reduces to that of increasing mixing character, a fact whose implications will be discussed briefly in the next subsection.

To get a simple feel for the meaning of this principle, consider again the special case in which an infinitesimal probability increment δp is transferred from the state α to β: For the change in a function $a(G; p, w)$ we have

$$\delta a \cdot \left\{ G'\left(\frac{p_\beta}{w_\beta}\right) - G'\left(\frac{p_\alpha}{w_\alpha}\right) \right\}.$$

Since for convex G, G' is a monotonically decreasing function of its argument, this will be positive (corresponding to an allowed process) if and only if $p_\alpha/w_\alpha > p_\beta/w_\beta$.

We have thus proved

Theorem 3. The mixing distance relative to a reference distribution decreases only when states with low values of p_α/w_α gain probability (or population) at the expense of states with higher values of this quantity.

4.5 Connection with Conventional Thermodynamics

In conventional thermodynamics, it will be recalled, one can start with the entropy principle $\delta S \geqslant 0$ for an isolated system, then derive from it the relation $\delta S \geqslant q/T$ for a system in interaction with a heat bath, by considering the system and bath together to comprise an isolated system. Having done this, one may reverse the process: Starting with $\delta S \geqslant q/T$ for an arbitrary system, and considering the special case of an isolated system for which $q = 0$, one immediately recovers the entropy principle [7].

We have done much the same thing here in subsections 4.1 and 4.2, except that we have used an arbitrary mixing homomorphic function for the isolated system (instead of just the entropy, which is a special case), and for each such function have derived a corresponding function for the system of interest alone.

Putting it another way, the principle of increasing mixing character for an isolated system leads to that of decreasing mixing distance for a system in interaction with a bath. Again, the derivation can be reversed.

Applying the principle of decreasing mixing distance to the special case of an isolated system, for which the equilibrium distribution is the uniform one, one recovers the principle of increasing mixing character. The only difference is that the principle of decreasing mixing distance, as we have formulated it, requires prior knowledge of the equilibrium distribution for the system of interest, while in thermodynamics this can be calculated in principle (e.g., in the case of a system of given temperature and volume, by minimizing the Helmholtz free energy). Since conventional thermodynamics fully suffices to determine the equilibrium state, this circumstance is not a drawback.

Since the choice $g(p) = -p \ln p$ for an isolated systems leads to $\delta S \geqslant 0$ one might expect that the choice $G(x) = -x \ln x$ for a system in interaction with a bath would lead to $\delta S \geqslant q/T$, or the direct consequences thereof, for that system. That this is indeed the case is easily verified.

Consider, for example, the canonical, constant volume, ensemble for which

$$w_\alpha = Q^{-1} \exp\left(\frac{E_\alpha}{kT}\right)$$

where Q is the partition function. With $G(x) = -x \ln x$, we obtain

$$\begin{aligned} a(G; p, w) &= -\sum_\alpha w_\alpha \left(\frac{p_\alpha}{w_\alpha}\right)\left(\ln p_\alpha + \frac{E_\alpha}{kT} + \ln Q\right) \\ &= -\sum_\alpha p_\alpha \ln p_\alpha - \frac{E}{kT} - \ln Q \\ &= -\frac{1}{kT}(E - TS) - \ln Q \\ &= -\frac{A}{kT} - \ln Q, \end{aligned}$$

where E is the average energy, A the Helmholtz free energy, and S is now expressed in conventional units, differing by a factor of k from S as previously used in this paper.

The requirement that this function increases is clearly equivalent to the well-known thermodynamic result that A must decrease under these conditions, since $\ln Q$ is constant at fixed T and V.

The relation between our principle and conventional thermodynamics, and between isolated and nonisolated systems, is summarized in the following table, in which arrows indicate directions in which derivations may proceed. The one dotted arrow refers to the fact that one needs to know the equilibrium distribution in order to derive the upper result from the lower one.

$$\begin{tikzcd}
 & \text{Conventional} & \text{Present Treatment} \\
\text{Isolated System} & \delta S \geqslant 0 \arrow[d, shift right] & \delta m(g;\gamma) \geqslant 0 \arrow[l] \arrow[d] \\
\text{System Interacting with Bath} & \delta S \geqslant q/T \arrow[u, shift right] & \delta a(G;p,w) \geqslant 0 \arrow[l] \arrow[u, dashed]
\end{tikzcd}$$

4.6 Partial Equilibrium

We consider the case of a system which is in equilibrium with respect to some degrees of freedom, but not necessarily with respect to all. For example, the translational, vibrational and rotational motion might be in equilibrium, with the spins aligned in a non-equilibrium distribution; or the various motions in a gas may be in local equilibrium everywhere, but the concentration non-uniform, etc.

Let β denote the complete set of quantum numbers for those degrees of freedom not necessarily in equilibrium and Γ the same for the other degrees of freedom. Let $p_{(\beta)\Gamma}$ denote the conditional probability that the system is in a state β, Γ given that it is in a state with quantum numbers β and that it is in equilibrium with respect to the Γ. Since we are assuming this equilibrium, the unconditional probability $p_{\beta,\Gamma}$ (and $w_{\beta,\Gamma}$ in equilibrium) is then

$$\begin{array}{l} p_{\beta,\Gamma}=p_\beta p_{(\beta)\Gamma} \\ w_{\beta,\Gamma}=w_\beta p_{(\beta)\Gamma} \end{array}, \quad \sum_\Gamma p_{(\beta)\Gamma}=1,$$

where p_β is the total probability, independent of Γ, that the system is in a state β. Inserting these relations into (17) we have:

$$a(G;p,w)=\sum_{\beta,\Gamma} w_\beta p_{(\beta)\Gamma} G\left(\frac{p_{\beta,\Gamma}}{w_{\beta,\Gamma}}\right)=\sum_\beta w_\beta G\left(\frac{p_\beta}{w_\beta}\right). \tag{27}$$

This shows that we may use the functions $a(G;p,w)$ where p, w denote the probabilities p_α, w_α of those degrees of freedom which are not in equilibrium. Hence, the requirement

$$\delta a(G;p,w)\geqslant 0 \quad \text{for all convex } G(x) \text{ over } 0\geqslant x$$

governs the approach to equilibrium also in the case of a system in partial equilibrium.

5. Examples

To illustrate the kind of information derivable from this method, we treat in this section two very simple examples. The first has to do with diffusion in ideal solutions, the second with chemical reactions in ideal gas mixtures. In neither case are the results particularly startling, but in both cases they go beyond what one could derive with conventional thermodynamics, which we show in each case by direct comparison.

5.1 Diffusion in Ideal Solution

We consider a dilute, ideal solution, not necessarily of uniform concentration, described by a solute density $\rho(r)$. Since the solution is ideal, the solute particles are statistically independent of one another, so they can be considered to be members of an ensemble. Thus, a state of a particle (ensemble member) is characterized by its position in space, plus other variables. We assume equilibrium with respect to everything except position in space, and use the result of Sec. 4.6, Eq. (27), with the sum replaced by an integral over position. Since the equilibrium distribution in this case is the uniform one, (27) becomes

$$a(G; p, w) = \int G(\rho) d^3r$$

with

$$\delta a = \int G'(\rho)\delta\rho d^3r \geqslant 0. \tag{28}$$

Because the total number of particles is conserved, we must have a continuity equation obeyed:

$$\delta\rho = -\nabla \cdot \boldsymbol{J} \tag{29}$$

where $\boldsymbol{J}$ is a solute current density vector. Insertion of (29) into (28) gives

$$\delta a = -\int G'(\rho)(\nabla \cdot \boldsymbol{J}) d^3r \geqslant 0 \tag{30}$$

which, after integration by parts, becomes

$$\delta a = \int G''(\rho)(\boldsymbol{J} \cdot \nabla\rho) d^3r \geqslant 0 \tag{31}$$

$G''(\rho)$ is never positive, but is otherwise arbitrary. Accordingly, (31) is satisfied for arbitrary convex G only if

$$\int_{\rho=\rho_0} \frac{\boldsymbol{J} \cdot \nabla\rho}{|\nabla\rho|} d^2r \geqslant 0, \text{ all } \rho_0, \tag{32}$$

where the integral goes over the surface or surfaces where ρ has a fixed but arbitrary value ρ_0, and the inequality must hold for all ρ_0. Apart from unphysical cases in which the current at a point is assumed to be influenced by the current at distant points which happen to have the same value of ρ, (32) means to all intents and purposes that

$$\boldsymbol{J} \cdot \nabla\rho \leqslant 0 \quad \text{everywhere.}$$

In other words, the diffusion current must always be in the direction opposed to the concentration gradient, i.e., from regions of high concentration into low concentration. This result is certainly not surprising, but it is more than one could obtain from thermodynamics alone. To compare with the result of conventional thermodynamics, we evaluate (31) with the "conventional" choice,

$$G(\rho) = -\rho \ln \rho, \quad G''(\rho) = -\frac{1}{\rho}. \tag{33}$$

Insertion of (33) into (31) gives

$$\delta s = -\int \frac{1}{\rho}(\boldsymbol{J}\cdot\nabla\rho)d^3r \tag{34}$$

where by definition $s = a(G; p, w)$ with G defined by (33). We show by means of an example that it is possible to satisfy (34) without satisfying (32). Consider the phenomenological diffusion equation

$$\boldsymbol{J} = -D\nabla\rho + L\rho\nabla(\nabla^2\rho), \tag{35}$$

where D and L are phenomenological constants, to be determined perhaps by fitting empirical data. Substitution of (35) into (34) gives

$$\delta s = D\int(\nabla\rho)^2 d^3r - L\int(\nabla\rho)\cdot\nabla(\nabla^2\rho)d^3r.$$

We now integrate the second term by parts, obtaining

$$\delta s = D\int(\nabla\rho)^2 d^3r + L\int(\nabla^2\rho)^2 d^3r \geqslant 0. \tag{36}$$

Eq. (36) is satisfied if $D, L \geqslant 0$. However, (32) is only satisfied if $L = 0$, since for any nonzero value of L, it is possible to have the second term of (35) larger than the first and of opposite sign. Thus, our principle gives more information than one obtains from thermodynamics. In this case, of course, the result (32) could also have been obtained in other ways, either by requiring local (not just overall) entropy increase, or simply by common sense. Nevertheless, this affords a very simple example of the ability of our method to yield results beyond those strictly obtainable by thermodynamics.

5.2 Chemical Reactions in Ideal Gas Mixtures

Consider a mixture of several ideal gas species, capable of undergoing chemical reactions, and not necessarily at equilibrium. Let c_τ be the equilibrium concentration of the τ'th species, and l_τ its instantaneous concentration. We assume equilibrium with respect to everything except concentrations, and again use the results of Sec. 4.5. The probability that there will be exactly ν_τ molecules of each species τ in a given unit element of volume is given by the Poisson distribution

$$p\{\nu_\tau\} = \prod_\tau \frac{l_\tau^{\nu_\tau}}{\nu_\tau!} e^{-l_\tau}, \tag{37}$$

while the equilibrium probability is

$$w\{\nu_\tau\} = \prod_\tau \frac{c_\tau^{\nu_\tau}}{\nu_\tau!} e^{-c_\tau}. \tag{38}$$

For this case, we have not been able to evaluate a general function $a(G; p, w)$, so we content ourselves with the power functions (12b) or rather the corresponding a-functions

$$a_n = -\sum_{\{\nu_\tau\}} w\{\nu_\tau\}\left[\frac{p\{\nu_\tau\}}{w\{\nu_\tau\}}\right]^n. \tag{39}$$

Substituting (37), (38) into (39), one obtains

$$\begin{aligned} a_n &= -\sum_{\{\nu_\tau\}} \prod_\tau \exp\left[-(nl_\tau-(n-1)c_\tau)\right] \frac{1}{\nu_\tau!}\left[\frac{l_\tau^n}{c_\tau^{n-1}}\right]^{\nu_\tau} \\ &= -\exp\left[\sum_\tau \left(\frac{l_\tau^n}{c_\tau^{n-1}} - nl_\tau + (n-1)c_\tau\right)\right] \end{aligned} \tag{40}$$

and

$$\delta a_n = a_n n \sum_\tau \left[\frac{l_\tau^{n-1}}{c_\tau^{n-1}} - 1\right] \delta l_\tau \geqslant 0. \tag{41}$$

Since a_n is negative, (41) means

$$\sum_\tau \left[1 - \left(\frac{l_\tau}{c_\tau}\right)^{n-1}\right] \delta l_\tau \geqslant 0, \quad \text{all} \quad n \geqslant 1. \tag{42}$$

If we had used $G(x) = -x \ln x$, corresponding to conventional thermodynamics, we would have obtained instead of (42)

$$\sum_\tau \left[\ln\left(\frac{l_\tau}{c_\tau}\right)\right] \delta l_\tau \leqslant 0.$$

We can obtain somewhat more information by making use of the generalized power functions

$$g_{nu}(p) = -(p-u)^{2n} \quad (n \text{ integer, } u \text{ real and variable}); \tag{43}$$

with

$$m(g_{nu}; \gamma) \equiv m_{nu} = -\sum_\alpha (p_\alpha - u)^{2n}, \quad m_n = -\sum_\alpha p_\alpha^n$$

we have

$$m_{nu} = \sum_{r=0}^{2n} \frac{(-1)^r (2n)!}{r!(2n-r)!} u^{2n-r} m_r. \tag{44}$$

Since m_0 and m_1 are both constants which do not change when the p's are varied, we have for the change in m_{nu} from (44):

$$\Delta m_{nu} = \sum_{r=2}^{2n} \frac{(-1)^r (2n)!}{r!(2n-r)!} u^{2n-r} \Delta m_r. \tag{45}$$

This must be non-negative for an allowed process, independently of u, and this leads to further constraints on the Δm_r, which, however, are simple only for $n=2$. For this case, we have from (45)

$$\Delta m_{2u} = \Delta m_4 - 4u\Delta m_3 + 6u^2 \Delta m_2. \tag{46}$$

Minimizing this with respect to u, and requiring that the minimum value be non-negative, we obtain

$$\Delta m_4 \Delta m_2 \geqslant \tfrac{2}{3}(\Delta m_3)^2. \tag{47}$$

We can also apply (47) to our example, since it obviously applies as well to the a_n as to the m_n. From (40), (41), we find

$$\begin{aligned}
\delta a_2 &= 2\exp\left[\sum_\mu\left(\frac{l_\mu^2}{c_\mu}-2l_\mu+c_\mu\right)\right]\sum_\tau\left(1-\frac{l_\tau}{c_\tau}\right)\delta l_\tau;\\
\delta a_3 &= 3\exp\left[\sum_\mu\left(\frac{l_\mu^3}{c_\mu^2}-3l_\mu+2c_\mu\right)\right]\sum_\tau\left(1-\frac{l_\tau^2}{c_\tau^2}\right)\delta l_\tau;\\
\delta a_4 &= 4\exp\left[\sum_\mu\left(\frac{l_\mu^4}{c_\mu^3}-4l_\mu+3c_\mu\right)\right]\sum_\tau\left(1-\frac{l_\tau^3}{c_\tau^3}\right)\delta l_\tau.
\end{aligned} \tag{48}$$

Substituting (48) into (47), one obtains, after some rearranging:

$$\exp\left[\sum_\mu\frac{l_\mu^2}{c_\mu}\left(1-\frac{l_\mu}{c_\mu}\right)^2\right]\left\{\sum_\tau\left(1-\frac{l_\tau}{c_\tau}\right)\delta l_\tau\right\}\left\{\sum_\sigma\left(1-\frac{l_\sigma^3}{c_\sigma^3}\right)\delta l_\sigma\right\} \geqslant \frac{3}{4}\left\{\sum_\tau\left(1-\frac{l_\tau^2}{c_\tau^2}\right)\delta l_\tau\right\}^2. \tag{49}$$

The volume element is still at our disposal, and can in particular be chosen so small that each l_τ is very small, and the exponential on the left-hand side of (44) is as close as we please to unity. This gives the result

$$\left\{\sum_\tau\left(1-\frac{l_\tau}{c_\tau}\right)\delta l_\tau\right\}\left\{\sum_\sigma\left(1-\frac{l_\sigma^3}{c_\sigma^3}\right)\delta l_\sigma\right\} \geqslant \frac{3}{4}\left\{\sum_\tau\left(1-\frac{l_\tau^2}{c_\tau^2}\right)\delta l_\tau\right\}^2. \tag{50}$$

Eqs. (42) and (50) are general requirements that must be obeyed by the increments in concentrations in ideal gas reaction mixtures. For the case of a single reaction proceeding toward equilibrium, they are of course automatically satisfied; in the case of a mechanism involving competing reactions or sequences of reactions, however, they are not trivial, and it may be possible to use them to exclude certain proposed mechanisms.

Since these results are obtained from the power functions only, they do not represent the most stringent conditions put on such systems by the principle of increasing mixing character.

6. Discussion: Possibilities for Future Work

The principle of increasing mixing character, which makes precise the common-sense notion that less populated states in an isolated system gain population at the expense of more populated ones, represents a stronger restriction on allowed thermodynamic processes than does the second law of thermodynamics. In this paper, we have given what we think is a rather convincing derivation of this principle, and then drawn some consequences from it. It would appear that many applications of this principle can be made, since it should be applicable to every system to which thermodynamics is applicable, and will in each case yield new information, though sometimes, as in the diffusion case treated in Sec. 5.1, this new information may be rather trivial.

Among the infinite number of mixing hormomorphic functions, the entropy continues to have a special place, since, in the first place, its exponential gives the actual probability of a given non-equilibrium state; and, in the second place, it is homogeneous of first order, hence additive, and expressible in purely phenomenological thermodynamic terms. It is perhaps worthwhile to examine the question of whether there are other mixing homomorphic functions with especially significant or convenient properties such as (but not identical with) these.

Another question not considered here is that of fluctuations, and to what extent they make possible deviations from the principle of increasing mixing character for systems of finite size.

We hope to consider these and other questions in future publications.

Appendix: Necessary Conditions for Mixing Isomorphic Sets

In this appendix, we derive some necessary conditions which must be satisfied if a set $\mathscr{M}$ of mixing homomorphic functions $m(g;\gamma)$ generated by a set $\mathscr{A}$ of convex functions $g(p)$ is to be mixing isomorphic. We concern ourselves primarily with requirements on the behavior in the neighborhood of a particular point $p=\lambda$ imposed on at least some of the g in $\mathscr{A}$. We restrict ourselves to convex functions $g(p)$ which are continuous and possess piecewise continuous first derivatives. Also, since a constant can always be added to $g(p)$ without affecting $\Delta m(g;\gamma)$, we shall assume $g(0)=0$ throughout.

The method we use will be to construct a process in which a very small amount of probability is transferred from a state with probability slightly less than λ to one with probability slightly more, with simultaneous probability transfer among the other states such as to lead to maximum increase in $m(g;\gamma)$. The overall process will be forbidden because some probability has been moved "in the wrong direction", hence if $\mathscr{M}$ is to be mixing isomorphic, there must be at least one g in $\mathscr{A}$ such that $m(g;\gamma)$ decreases. The strength of the results derived will depend on the magnitude of λ. Rather than give an exhaustive discussion, we confine ourselves to certain ranges of λ; the reader will easily be able to apply the ideas used here to other cases as needed. We first consider:

Case a: $\lambda<\frac{1}{5}$.

Let ε be a very small number, and define $v=\lambda+2\varepsilon$, $u=\lambda-3\varepsilon$. Consider an initial probability distribution as follows:

$$p_1=v+1-5\lambda;$$
$$p_2=p_3=v;$$
$$p_4=p_5=u.$$

All others zero.

For this distribution, we have

$$m(g;\gamma)=g(v+1-5\lambda)+2g(v)+2g(u).$$

We now deform the distribution in three steps, the first two of which will increase m, the last of which will decrease it, in such a way that the overall process is forbidden. The deformation is as follows:

1. We equalize the probabilities between the first two states, giving each a probability of $(2v+1-5\lambda)$. The change in $m(g;\gamma)$ for this process is

$$\Delta m_1=2g\left(\frac{2v+1-5\lambda}{2}\right)-g(v)-g(v+1-5\lambda).$$

This is just twice the distance of $g((2v+1-5\lambda)/2)$ above the chord drawn from $g(v)$ to $g(v+1-5\lambda)$, and is necessarily nonnegative for convex g.

2. The probability u of state 5 is distributed equally among N states, where N is allowed to approach infinity. The change in $m(g;\gamma)$ this time is

$$\Delta m_2=Ng\left(\frac{u}{N}\right)-g(u).$$

In the limit as N becomes large, we can replace $g(u/N)$ with $u/N(g'(0))$, giving the result

$$\Delta m_2 = u g'(0) - g(u).$$

This is just the distance of $g(u)$ below the extrapolated tangent to $g(p)$ at $p=0$. Again, for convex g it is nonnegative.

3. We transfer a very small probability δp from state 4 to state 3. The change in $m(g;\gamma)$ is now

$$\Delta m_3 = -(g'(u) - g'(v))\delta p,$$

and is negative or zero. The overall process is forbidden, as may be seen by examining the behavior of $m(g_\lambda, \gamma)$, Eq. (12c).

If $\mathscr{M}$ is to be mixing isomorphic, there must be at least one g in $\mathscr{A}$ for which $\Delta m_1 + \Delta m_2 + \Delta m_3 < 0$, i.e., for which the last, negative, term is large enough to compensate for the other two. There must be such a g for every ε and δp, no matter how small. One sees immediately that this can be accomplished only if there are g's which lie arbitrarily close to straight lines in the intervals $0 \leqslant p \leqslant u$, and $v \leqslant p \leqslant (v+1-5\lambda)$, even in the limit as both u and v approach λ; and which also have a finite change in first derivative in an arbitrarily small interval about λ. In other words, the function $g_\lambda(p)$, Eq. (12c), or a function differing from it by an additive linear function, must be arbitrarily well approximated by g in $\mathscr{A}$, at least in the interval $0 \leqslant p \leqslant (1-4\lambda)$. For example, if $\mathscr{A}$ only contains the power functions defined in Eqs. (12b) and (43), then $\mathscr{M}$ is definitely not mixing isomorphic.

This result holds only for $\lambda < 1/5$; in most applications, however, one needs only to consider infinitesimally small probabilities for single states. This means that, in most practical circumstances, a mixing isomorphic set must contain, or approximate arbitrarily well, all the $m(g_\lambda;\gamma)$ of Eq. (12c).

For other ranges of λ similar methods yield somewhat weaker results. For example, consider

Case b: $\frac{1}{5} < \lambda < \frac{1}{4}$.

Again let ε be small, and define $v = \lambda + \varepsilon$, $u = \lambda - 3\varepsilon$. Let the initial distribution be

$$\begin{aligned} &p_1 = v + q,\\ &p_2 = p_3 = v,\\ &p_4 = u,\\ &p_5 = 1 - 4\lambda - q,\\ &\text{all others zero.}\end{aligned}$$

where of course, $0 \leqslant q \leqslant (1-4\lambda)$.

Again, the deformation consists of equalizing between the first two states, distributing the probability of the fifth state among N states, and finally transferring an infinitesimal probability from state 4 to state 3. The change in $m(g;\gamma)$ is again the sum of three contributions, the first two positive, the third negative, as follows:

$$\begin{aligned} \Delta m_1 &= 2g\left(\frac{2v+q}{2}\right) - g(v+q) - g(v);\\ \Delta m_2 &= (1-4\lambda-q)g'(0) - g(1-4\lambda-q);\\ \Delta m_3 &= -(g'(u) - g'(v))\delta p.\end{aligned}$$

For the sum of these to be negative, we must have g's which lie arbitrarily close to straight lines for $0 \leqslant p \leqslant (1-4\lambda-q)$, and $v \leqslant p \leqslant (v+q)$, as well as a finite drop in first derivative in an arbitrarily small interval about λ. However, since $0 \leqslant q \leqslant (1-4\lambda)$, there is no restriction imposed on the behavior of $g(p)$ in the interval $(1-4\lambda) < p < \lambda$.

For larger λ, the results become still weaker, but the reader should have no difficulty deriving them as needed. The weakest case is

Case c: $\lambda > \frac{1}{2}$.

Here there can be no more than one state with probability λ or greater. The best that one can do is start with an initial distribution $p_1 = \lambda$, $p_2 = (1-\lambda)$, then first transfer δp from state 2 to state 1, afterwards distributing the remainder of p_2 among N states. The change in $m(g;\gamma)$ this time, to first order in infinitesimals, is

$$\Delta m = (1-\lambda)g'(0) - g(1-\lambda) - (g'(1-\lambda) - g'(\lambda))\delta p.$$

This can be made negative without any discontinuity in g. The only requirement is that g must undergo a finite change between $(1-\lambda)$ and λ, and that g lie arbitrarily close to a straight line in the interval $0 \leqslant p \leqslant (1-\lambda)$.

References

1. Ruch,E.: Theoret. Chim. Acta (Berl.) **38**, 167 (1975)
2. Ruch,E., Schönhofer,A.: Theoret. Chim. Acta (Berl.) **19**, 225 (1970)
3. von Neumann,J.: Mathematische Grundlagen der Quantenmechanik, pp. 202–12. Berlin: Springer-Verlag 1932; English translation: Mathematical foundations of quantum mechanics, trans. by R.T. Beyer, pp. 379–398. Princeton: Princeton University Press 1955
4. Hardy,G.H., Littlewood,J.E., Pólya,G.: Inequalities, 2nd Ed., p. 49. Cambridge: Cambridge University Press 1952.
5. Uhlmann,A.: Wiss.Z., Karl-Marx-Univ. Leipzig, Math-Naturwiss. R. **21**, 421 (1972); **22**, 139 (1973)
6. Hardy,G.H., Littlewood,J.E., Pólya,G.: Ref. [4], pp. 70–72. These authors actually define g to be convex if $(-g)$ satisfies Eq. (11). We feel, however, that our definition is more consistent with normal usage in the physical sciences.
7. Planck,M.: Vorlesungen über Thermodynamik, 10te Auflage, durchgesehen und erweitert von M. von Laue, §134, pp. 98–9, and §140, pp. 105–7. Berlin: Walter de Gruyter 1954; English translation: Treatise on thermodynamics, translated from 7th German Ed. (1922), by A. Ogg. New York: Dover Publications 1926, §133, p. 103, and §140, pp. 111–12. In Eq. (70) of the English edition, the symbol $\geqslant$ should be replaced by $\leqslant$.

Prof. Dr E. Ruch
Institut für Quantenchemie
Freie Universität Berlin
Holbeinstrasse 48
D-1000 Berlin 45

Prof. Dr. A. Mead
Chemistry Department
University of Minnesota
Minneapolis, Minnesota 55455
USA