ADVANCES IN MATHEMATICS 12, 178–268 (1974)

Ergodic Theory and Its Significance for Statistical Mechanics and Probability Theory*

GEORGE W. MACKEY

Department of Mathematics, Harvard University

DEDICATED TO GARRETT BIRKHOFF

1. INTRODUCTION

Ergodic theory is a relatively new branch of mathematics which from a mathematical point of view may be regarded as generated by the interaction of measure theory and the theory of transformation groups. Its basic concept of "metric transitivity" or "ergodicity" was introduced in 1928 in a paper of Paul Smith and G. D. Birkhoff on dynamical systems. However, the significance of this concept was not appreciated until late 1931 when J. von Neumann and G. D. Birkhoff proved the celebrated mean and pointwise ergodic theorems, and one may regard the nearly simultaneous appearance of these papers as marking the birth of the subject. Birkhoff's proof of the much more difficult pointwise ergodic theorem was stimulated by von Neumann's theorem and von Neumann, in turn, was stimulated by a key observation of B. O. Koopman.

Let Ω_E be a surface of constant energy E in the phase space Ω of some Hamiltonian dynamical system. Let $V_t(\omega)$ denote the point of phase space representing the "state" of the system t time units after it was represented by ω. Then, for each t, $\omega \to V_t(\omega)$ is a one-to-one transformation of Ω_E onto itself which conserves the natural volume element ζ_E in Ω_E induced in Ω_E by the Liouville measure $dq_1 \cdots dq_n\, dp_1 \cdots dp_n$. Moreover, $V_{t_1+t_2} = V_{t_1}V_{t_2}$ for all real numbers t_1 and t_2. Koopman's observation (not so obvious 40 years ago as now) was that we may obtain a unitary representation $t \to U_t$ of the additive group of the real line in the Hilbert space $\mathscr{L}^2(\Omega_E, \zeta_E)$ by defining $U_t(f)(\omega) = f(V_t(\omega))$.

* This article is a somewhat expanded version of a series of ten lectures given by the author at Texas Christian University June 19–23, 1972, in connection with a conference held there on the subject of the title.

Stone had just proved his celebrated theorem reducing the analysis of such a representation to that of the self-adjoint H which "generates it" via $U_t = e^{itH}$, and Koopman remarked that the recently developed complete classification of self-adjoint operators could now be used to provide significant invariants for dynamical systems.

Koopman's observation and remark suggested to von Neumann the possibility of an operator theoretic approach to the 60-year-old problem of justifying the interchange of space and time averages in the foundations of statistical mechanics. He recognized that metric transitivity was the proper substitute for the mathematically untenable "ergodic hypothesis" of Boltzman and his mean ergodic theorem provided the desired justification wherever metric transitivity could be established.

The impact of this work on statistical mechanics has been rather less than expected in large part because of the great difficulty of establishing metric transitivity for systems of physical interest. On the other hand, as we shall see in detail below, the ergodic theorem and the notion of ergodicity have had and continue to have considerable influence on the theory of "stationary stochastic processes" in probability theory.

Probability theory is much older than ergodic theory and is usually considered to have begun in 1654 with a correspondence between Fermat and Pascal about the proper division of stakes in certain gambling games. A chief concern during much of its history has been with the properties of sequences of "independent, identically distributed random variables" —in particular with the "law of large numbers" and the "central limit theorem". In the 1920's due to the work of Wiener, Steinhaus, and others, it came gradually to be realized that the new subject of measure theory, which had developed from the generalization and abstraction of ideas in Lebesgue's thesis of 1902, could be used to great advantage in providing a convenient and rigorous model for formulating and proving the results of probability theory. The decisive step was made by Kolmogoroff who systematically reformulated probability theory in measure theoretic terms in his fundamental and influential booklet "Grundbegriffen den Wahrsheinlichheitsrechnung" published in 1933. The next year several mathematicians recognized that the pointwise ergodic theorem, interpreted in probabilistic terms, was nothing more or less that a considerable generalization of the strong law of large numbers. Since these events, the measure theoretic point of view has more or less dominated probability theory—at least in its more theoretical aspects.

A key tool in ergodic theory—both in providing examples and in

analyzing those that exist— is provided by harmonic analysis and its generalization from the integers and the real line to locally compact groups. The main ideas in effecting this generalization were provided by Hermann Weyl in the 1920's for compact groups and by L. Pontrjagin, E. R. van Kampen, and A. Weil for locally compact commutative groups. It is perhaps noteworthy that the important Pontrjagin–van Kampen duality theorem was published the year after Kolmogoroff's treatise and the same year that the connection between the ergodic theorem and the law of large numbers was pointed out.

A second important tool—the concept of the entropy of an ergodic action—was introduced much later. In 1948 Claude Shannon, in studying certain problems in communication engineering, discovered that one could make precise mathematical sense out of the notion that some stochastic processes yield more "information" per observation than others and was thus able to assign a number to certain processes which he called *entropy* and which could be regarded as the rate at which they gave off information. In 1959 Kolmogoroff and Sinai showed that Shannon's notions could be used to assign an "entropy" to every ergodic action of the integers and that the resulting concept could be used to settle an old and important problem in pure ergodic theory. This breakthrough led to a rapid development of the properties of entropy and stimulated a renaissance in abstract ergodic theory which is still very much underway. The word "entropy" is used in information theory because of a certain analogy with the interpretation via statistical mechanics of the classical entropy of thermodynamics. Actually, recent work in statistical mechanics on the so-called "thermodynamical limit" and the hypothetical "infinite system" which realizes this limit makes it possible to recast statistical mechanics in a form which brings it into very close contact with the theory of stationary stochastic processes. Indeed, one may think of statistical mechanics in the infinite system limit as the theory of stochastic processes with "three-dimensional time". Conversely, one may think of an ordinary stochastic process with a discrete state space as providing the statistical mechanics of a one-dimensional gas. From this point of view, the entropy of information theory is identical with that of thermodynamics and statistical mechanics. Moreover, ergodicity and the ergodic theorem may be made to play a role in statistical mechanics analogous to the one it plays in the theory of stochastic processes and rather different from the classical role which gave rise to ergodic theory in the first place.

In these lectures I propose to present a connected exposition of some

of the main ideas of ergodic theory both as a branch of pure mathematics and as the appropriate tool for analyzing the random phenomena encountered in the statistics of time series, in communication engineering, and in statistical physics. Emphasis will be placed upon the central importance of the concept of ergodicity or metric transitivity and its role in making possible a rigorous mathematical model for random behavior. Another theme that will be stressed is interplay between the problems of pure ergodic theory and the meaning for applications of the conceptual tools used to deal with them. In the hope of making the lectures conceptually self-contained and accessible to a wide audience, I have included introductory accounts of the necessary harmonic analysis as well as of the basic notions of probability, thermodynamics, and statistical mechanics.

2. Measure Theory and Probability

We shall do measure theory within the context of "standard Borel spaces" and begin by reminding the reader of the basic facts. A *Borel space* is a set S together with a distinguished σ field of subsets; that is, a family of subsets closed under the countable Boolean operations. The members of the distinguished family are called the *Borel subsets* of S. A function f from one Borel space to another is said to be a Borel function if $f^{-1}(E)$ is a Borel set in the first space whenever E is a Borel set in the second. A bijection f from one Borel space to another is said to be a *Borel isomorphism* if f and f^{-1} are both Borel functions. When there exists a Borel isomorphism from the Borel space S_1 to the Borel space S_2, one says that S_1 and S_2 are isomorphic Borel spaces. Every topological space may be made into a Borel space by defining the Borel sets to be the members of the smallest σ field containing the open and closed sets. A Borel space is defined to be *standard* if it is isomorphic to the Borel space defined by a Borel subset of a separable complete metric space. It is a remarkable fact that two standard Borel spaces having the same cardinal number are isomorphic as Borel spaces and that the continuum hypothesis holds for standard Borel spaces. Thus there is to within isomorphism just one noncountable standard Borel space. Let Γ be a countable set, and for each γ in Γ, let A_γ be a Borel space. Then $\prod_\gamma A_\gamma$, the Cartesian product of the A_γ, may be identified with the set of all functions g defined on Γ and such that $g(\gamma) \in A_\gamma$ for all $\gamma \in \Gamma$. For each $\gamma_0 \in \Gamma$ and each Borel subset E of A_{γ_0}, let $\tilde{E}$ denote the set of all $g \in \prod_\gamma A_\gamma$

such that $g(\gamma_0) \in E$. $\prod_\gamma A_\gamma$ is given a Borel structure, called the *product Borel structure*, by defining a set to be a Borel set if its is contained in the smallest σ field containing all the sets $\tilde{E}$. It follows that $g \to g(\gamma_0)$ is a Borel function for all $\gamma_0 \in \Gamma$. It is easy to show that $\prod_\gamma A_\gamma$ is standard whenever all of the A_γ are standard. In particular, choosing all of the A_γ to be the same, one sees that the set A^Γ of all functions from a countable set Γ to a standard Borel space A is itself a standard Borel space in a natural way.

By a *measure* in the Borel space S, we shall mean a function μ from the Borel subsets of S to the nonnegative real numbers and ∞ such that $\mu(E_1 \cup E_2 \cdots) = \mu(E_1) + \mu(E_2) + \cdots$ whenever $E_i \cap E_j = 0$ for $i \neq j$, and such that $S = A_1 \cup A_2 \cdots$ where $\mu(A_j) < \infty$. In other words, measures for us will always be σ *finite*. The measure μ is said to be *finite* if $\mu(S) < \infty$, and is said to be a probability measure if $\mu(S) = 1$. If $\mu(\{s\}) > 0$ for some $s \in S$, we call s an atom and say that μ is *free of atoms* if there are no points s such that $\mu(\{s\}) > 0$. Combining the theorem stated above about standard Borel spaces with a well-known result of Halmos and von Neumann, one shows that given measures μ_1 and μ_2 in standard Borel spaces S_1 and S_2 which are free of atoms and such that $\mu_1(S_1) = \mu_2(S_2) \neq 0$, then there exists an isomorphism f of S_1 on S_2 which is *measure preserving* in the sense that $\mu_2(f(E)) = \mu_1(E)$ for all Borel subsets E of S_1.

Let μ be a probability measure in the standard Borel space Ω. By a real- or complex-valued *random variable*, one means a real- or complex-valued Borel function defined on Ω. Of course, one can define random variables with values in an arbitrary Borel space, but we shall deal mainly with those that are real- and complex-valued. Let f be a real-valued random variable. Setting $\alpha_f(E) = \mu(f^{-1}(E))$ for all Borel subsets of the real line, we obtain a probability measure α_f in the real line which is called the *distribution* of the random variable f. One thinks of $\alpha_f(E)$ as the "probability" that an observation of f will lead to a value in E. The integral $\int_{-\infty}^{\infty} x \, d\alpha_f(x) = \int f(a) \, d\mu(a)$ (if it exists) is called the expected value e of the random variable f, and $\int (x - e)^2 \, d\alpha_f(x) = \int (f(\omega) - e)^2 \, d\mu(\omega)$ (if it exists) is called the *variance*. Analogous definitions and remarks can be made for complex-valued random variables.

As long as one is concerned with a single random variable, the space Ω is superfluous; one need only know the probability measure α_f. In other words, a random variable in isolation is completely described by its distribution. One introduces the space Ω in order to be able to discuss

relationships between random variables. If f and g are two real-valued random variables, then in addition to the probability measures α_f and α_g in the real line which are their distributions, one may construct a certain probability measure $\alpha_{f,g}$ in the plane which is called their *joint distribution.* This is the probability measure $E \to \mu(\varphi^{-1}(E))$, where φ is the Borel function $\omega \to f(\omega)$, $g(\omega)$ from Ω to the space of pairs of real numbers. α_f and α_g are immediately deduceable from $\alpha_{f,g}$ by projecting on the coordinate axes, but many different measures $\alpha_{f,g}$ are consistent with a given pair α_f, α_g. One such is the product measure $\alpha_f \times \alpha_g$, and when $\alpha_{f,g} = \alpha_f \times \alpha_g$, one says that the random variables f and g are *independent.* Intuitively, this means that an observation of f does not change in any way the distribution of possible values of g. At the other extreme, one might have $g = f^2$ so that, having measured f, there is no uncertainty left for g. In this case, $\alpha_{f,g}$ is supported by the curve $y = x^2$ in the space of all pairs x, y; that is, the $\alpha_{f,g}$ measure of the complement of the curve is zero. Of course, there are many intermediate possibilities.

Consider the special case in which α_f is supported by a countable set $\{a_1, a_2, \ldots\}$. It follows at once from the definitions that $\alpha_{f,g}$ is supported by the union of the lines L_j, where L_j is the set of all x, y with $x = a_j$. For each Borel subset E of the real line, let E_j denote the set of all x, y with $x = a_j$ and $y \in E$. Then $E \to \alpha_{f,g}(E_j)/\alpha_{f,g}(L_j)$ is a probability measure β_j in the real line. One interprets $\beta_j(E)$ as the probability that y is in E *given that* $x = a_j$ and calls β_j the *conditional probability distribution* for g given that $f = a_j$. Saying that f and g are independent as defined above is equivalent to saying that β_j is the same for all j. On the other hand, when g is a function of f, then β_j is concentrated in a single point for all j but that point will vary with j. The notion of conditional probability distribution can be defined even when α_f is not concentrated in a countable set. However, it requires more sophisticated mathematics to do so and it will be convenient to postpone the details to the next section.

If $f_1, f_2, \ldots, f_n$ is a set of n random variables, then $\omega \to f_1(\omega), f_2(\omega), \ldots, f_n(\omega) = \varphi(\omega)$ is a Borel map of Ω into the space R^n of all n-tuples of real numbers. Setting $\alpha_{f_1,f_2,\ldots,f_n}(E) = \mu(\varphi^{-1}(E))$, one obtains a probability measure in R^n called the joint distribution of the n-tuple $f_1, \ldots, f_n$. If $\alpha_{f_1,\ldots,f_n} = \alpha_{f_1} \times \alpha_{f_2} \cdots \times \alpha_{f_n}$, one says that the f_j are mutually independent. To say that each pair f_i, f_j are independent does *not* in general imply that $f_1, \ldots, f_n$ are mutually independent.

By a *discrete stochastic processes* we shall mean a doubly infinite sequence $\ldots f_{-2}, f_{-1}, f_0, f_1, f_2 \ldots$ of real or complex random variables.

One usually thinks of the index n in f_n as a time parameter. The nth random variable f_n might be the result of the nth spin of a roulette wheel or the temperature at a certain point in space at a certain hour on the nth day. The real or complex number $f_n(\omega)$ is a function of the two variables n and ω and one may think of this function of two variables as either a family of functions of ω parametrized by the integer n or as a family of functions of n (doubly infinite sequences) parametrized by ω. In the first instance, $\omega \to f_n(\omega)$ are random variables; in the second, the functions $n \to f_n(\omega)$ are called *sample functions* or *sample sequences*. Each sample function represents a possible sequence of observations of the infinitely many random variables f_n. For example, in tossing a coin, f_n might be the random variable which is 1 when the nth toss is "heads" and -1 when it is tails. A sample function would then be a doubly infinite sequence of 1's and -1's representing the results of a doubly infinite sequence of coin tosses. One thinks of a so-called (discrete) time series in statistics as being the restriction to the nonnegative integers of some sample function of a discrete stochastic process. One seeks to recover the process from the time series and to make predictions about the "future" values of the sample functions.

Given the discrete stochastic process $\ldots f_{-2}, f_{-1}, f_0, f_1, \ldots$, let $s(\omega)$ denote the sample function with parameter ω; that is, let $s(\omega)$ denote the doubly infinite sequence of real or complex numbers $\ldots f_{-2}(\omega), f_{-1}(\omega), f_0(\omega), f_1(\omega), \ldots$. Then $\omega \to s(\omega)$ is a mapping of Ω into the space R^∞ or C^∞ of all possible doubly infinite sequences of real or complex numbers. Moreover with respect to the product Borel structure in R^∞ (resp. C^∞), the mapping s is a Borel function. Hence, $E \to \mu(s^{-1}(E)) = \alpha(E)$ is a probability measure α on R^∞ (resp. C^∞) which is the joint probability distribution of the whole infinite collection $\{f_n\}$ of random variables. If the mapping s is injective, one says that the random variables *separate* the points of Ω. When they do not, one can introduce an equivalence relation in Ω by saying that $\omega_1 \sim \omega_2$ if $f_j(\omega_1) = f_j(\omega_2)$ for all j. The space $\tilde{\Omega}$ of all equivalence classes may be made into a Borel space by defining the Borel sets to be those sets F such that $r^{-1}(F)$ is a Borel set in Ω. Here, $r(a)$ denotes the equivalence class to which a belongs. Defining $\tilde{\mu}(F) = \mu(r^{-1}(F))$ and $\tilde{f}_j(r(\omega)) = f_j(\omega)$, one obtains a new stochestic process $\tilde{f}_{-2}, \tilde{f}_{-1}, \tilde{f}_0, \ldots$ defined on $\tilde{\Omega}$, $\tilde{\mu}$ whose random variables separate the points of $\tilde{\Omega}$. For all probability purposes, this new process is equivalent to the original one so that there is no essential loss in generality in assuming that the random variables do separate. While it is not always true that $\tilde{\Omega}$ is standard, one can always find a Borel set N

with $\tilde{\mu}(N) = 0$ so that $\tilde{\Omega} - N$ is standard and $\tilde{\Omega}$ can be replaced by $\tilde{\Omega} - N$.

When the random variables do separate (as we may always assume), the mapping s is a Borel isomorphism of Ω onto a Borel subset of R^∞ (resp. C^∞) which carries μ into the restriction of α to the image of Ω. Since the complement of this image has α measure zero, there is no loss of generality in replacing Ω by R^∞ (resp. C^∞) and μ by α. In other words, a discrete real-valued (respectively complex-valued) stochastic process may be equivalently defined as a probability measure in R^∞ (resp. C^∞). The space Ω is then the set of all real-valued (respectively complex-valued) functions on the group Z of all integers and the random variable $\omega \to f_n(\omega)$ is the function $\omega \to \omega(n)$. One advantage of this definition is that it permits us to introduce a natural homomorphism of the integers into a group of automorphisms of Ω as a Borel space. For each $\omega \in \Omega$ and each integer n, let $\omega n = \omega'$, where $\omega'(m) = \omega(n + m)$. Then $\omega \to \omega n$ is an automorphism of Ω and $((\omega)n)\, n' = [\omega](n + n')$. Of course, in general, these automorphisms will not preserve the measure μ, but the case in which they do will be of central importance for us. Given this "action" of the group of integers in Ω, one need only know one of the random variables $\ldots f_{-2}, f_{-1}, f_0, f_1, f_2$ in order to know them all. Indeed, one has $f_n(\omega) = f_0([\omega]n)$. Thus, a third equivalent definition of a discrete stochastic process and the one that will be most convenient for us is the following. A discrete stochastic process is the system consisting of (i) a standard Borel space Ω, (ii) an assignment of an automorphism $\omega \to [\omega]n$ of Ω to each integer n such that $[[\omega]n]m = [\omega](n + m)$, (iii) a probability measure μ on Ω, and (iv) a real- or complex-valued function f on Ω. To get the random variables of the process as originally defined, one simply sets $f_n(\omega) = f([\omega]n)$.

A discrete stochastic process is said to be stationary if, for each r, the joint distribution of $f_{m+1}, f_{m+2}, \ldots, f_{m+r}$ in the space of r-tuples is the same for all m. It is easy to see that a process is stationary if and only if the probability measure α defined in R^∞ (resp. C^∞) is invariant under the natural action of the integers. Thus the measure μ in our third definition of a stochastic process may be taken as invariant whenever the process is stationary.

The notion of discrete stochastic process may be generalized in two obvious directions. First of all, one may replace the discrete time parameter by a continuous one and consider families of random variables $\{f_t\}$ parametrized by a real number t. The measure theoretic technicalities are less straightforward than in the discrete case, but one shows in analogy

with the above that such a continuous parametric process may be equivalently defined as the system consisting of (i) a standard Borel space Ω, (ii) an assignment of an automorphism $\omega \to [\omega]t$ of Ω to each real number t in such a manner that $[[\omega]t]\ t' = [\omega](t + t')$ for all t and t' and t, $\omega \to [\omega]t$ is a Borel function from $\Omega \times R$ to Ω, (iii) a probability measure μ in Ω, and (iv) a Borel function f on Ω. As in the discrete case, one calls the functions $t \to f([\omega]t)$ the sample functions of the process, and for *stationary processes* the measure μ is *invariant* in the sense that $\mu([E])t) = \mu(E)$ for all t and E. Let f_ω denote the sample function $t \to f([\omega]t)$, and let ψ denote the mapping $\omega \to f_\omega$ of Ω into the space $\mathscr{F}$ of all real- or complex-valued functions of a real variable. Setting $\tilde{\mu}(E) = \mu(\psi^{-1}(E))$, one obtains a measure $\tilde{\mu}$ defined on a certain σ field of subsets of $\mathscr{F}$, or equivalently, on a certain σ field of subsets of $\mathscr{F}_0$, where $\mathscr{F}_0$ is the image of Ω under ψ. Unlike the discrete case, $\mathscr{F}$ itself cannot be made into a standard Borel space and we cannot define the process by simply giving a measure $\tilde{\mu}$ defined on certain Borel sets specified in advance. One must also give $\mathscr{F}_0$ or allow the Borel sets to vary with the process. The difficulties produced by this circumstance have been extensively investigated by Doob but will not concern us further here.

The other obvious direction of generalization consists in replacing the one-dimensional time parameter of the process whether discrete or continuous by a parameter varying over some more general multi-dimensional set such as physical space. Whenever the parameter space has or can be given the structure of a group in a natural way, one can introduce a notion of invariance for the family of random variables which reduces to stationarity in the special case in which the group is the group of translations in time. Thus one is led to study the system consisting of a more or less general group G acting as a group of automorphisms of a standard Borel space Ω, a real- or complex-valued Borel function f defined on Ω, and a probability measure μ in Ω—special interest attaching itself to the case in which μ is invariant.

Still further generalizations are possible and in fact important. However, it will be convenient to postpone their description until Section 8.

3. Actions of Separable Locally Compact Groups and the Notion of Ergodicity

Having seen how groups of measure-preserving transformations present themselves in a natural way in problems in probability theory,

we turn our attention in this section and the next to the study of such groups from the point of view of pure mathematics. Let G be a locally compact topological group which is separable in the sense that there is a countable basis for the open sets. By a standard Borel G space, we shall mean a standard Borel space S together with a Borel mapping $s, x \to sx$ of $S \times G$ into S such that $[sx]\, y = s(xy)$ and $se = s$ for all x, y in G and all $s \in S$, where e denotes the identity element of G. It follows at once that for each fixed x, the mapping $s \to sx$ is an automorphism of S and that the mapping so defined from G to automorphisms of S is a homomorphism. A measure μ in S is said to be *invariant* if $\mu([E]x) = \mu(E)$ for all x in G and all Borel subsets E of S. For any measure μ in S and any x in G, let $\mu_x(E) = \mu([E]x)$. μ is said to be *quasi-invariant* if μ and μ_x have the same null sets for all x. While applications to probability theory are chiefly concerned with invariant probability measures, much of abstract ergodic theory may be developed for measures which are only quasi-invariant. In dealing with quasi-invariant measures, one can replace the given measure by any measure having the same null sets without affecting the central concepts of the theory, and it is often convenient to deal at once with all of these measures at once. We define a *measure class* to be the set of all measures having the same null sets (sets of measure zero) as any one of them and define a measure class to be invariant if for each X in G, and each μ in the class, μ_x is also in the class. Of course, it amounts to the same thing to say that each μ in the class is quasi-invariant and this is implied by the quasi-invariance of any member of the class. An invariant measure class may or may not contain an invariant measure.

By an action of the separable locally compact group G, we shall mean the system consisting of a standard Borel G space S and an invariant measure class C in S. For each group G, we shall be interested in classifying the actions of G up to "isomorphism", two actions on S_1, C_1 and S_2, C_2 being said to be isomorphic if there exist G invariant Borel null sets N_1 and N_2 in S_1 and S_2, respectively, and a Borel isomorphism φ of $S_1 - N_1$ or $S_2 - N_2$ such that

(1) If E is a Borel subset of $S_1 - N_1$, then $\varphi(E)$ is a C_2 null set if and only if E is a C_1 null set.

(2) $\varphi(sx) = \varphi(s)x$ for all $s \in S$ and $x \in G$.

Given an action of G with measure class C in the standard Borel G space S, suppose that S admits a Borel subset E which is invariant and such that neither $S - E$ nor E is a C null set. Then $S - E$ and E are

both standard Borel G spaces, and the restrictions of the measures in C to $S - E$ and E define invariant measure classes and, hence, actions of G. The original action is in an obvious sense a "direct sum" of the two "subactions" defined by $S - E$ and E. Quite generally, if $S_1, S_2, \ldots$ are standard Borel G spaces with invariant measure classes $C_1, C_2, \ldots$, we may make $S = S_1 \cup S_2 \ldots$ the disjoint union of the S_j into a standard G space by declaring the Borel subsets of S to be the sets $E_1 \cup E_2 \ldots$, where E_j is a Borel subset of S_j. If $\mu_j \in C_j$, we obtain a measure μ in S by defining $\mu(E) = \sum_{j=1}^{\infty} \mu_j(E \cap S_j)$ whose class C depends only on the C_j. C is invariant, and we obtain a new action which we call the *direct sum* of the actions on the S_j. An action which is *not* isomorphic to a direct sum of two or more nontrivial actions is said to be *ergodic* or *metrically transitive*. It follows at once from the definition that an action is ergodic if and only if every invariant Borel subset of S is either of measure zero (a null set) or the complement of a set of measure zero. As we shall see below, the ergodic actions are the fundamental building blocks from which all other actions can be constructed. While it is not true that every action is a direct sum of ergodic actions, it is true that every action is in a certain sense a "direct integral" of ergodic actions and that this decomposition is essentially unique. It follows that to a large extent one can reduce the problem of finding all isomorphism classes of actions of a given group to that of finding the isomorphism classes of *ergodic actions*.

Given an action of G with space S and invariant measure class C, one defines the *orbit* of $s \in S$ to be the set of all points into which s may be transformed by the elements of G; that is, the orbit of s is the set of all sx with $x \in G$. Any two orbits are either disjoint or identical, and when there is only one orbit, one says that the action is *transitive*. Obviously, transitive actions are always ergodic. Indeed, they can be defined as actions in which there are no invariant subsets at all except for the empty set and the whole space. The ergodic actions then are those which are transitive in a measure theoretic sense, that is, in which there are no invariant Borel subsets except sets of measure zero and their complements. Hence the term "metric transitivity", The shorter and more convenient synonym "ergodic" comes from two Greek words meaning "work" and "path" and refers to the original (untenable) hypotheses of statistical mechanics to the effect that the trajectory of a point in phase space goes through every point of a constant-energy hypersurface.

Given any ergodic action, one can consider its orbits and ask whether any are of positive measure. It can be proved that all orbits are Borel

sets and it follows at once from ergodicity that at most one can be of positive measure. Thus we have the following dichotomy: either

(a) There exists a unique orbit of positive measure and its complement is a null set so that the action is in fact isomorphic to a transitive action.

or

(b) Every orbit is of measure zero.

In the first case, we shall say that the action is *essentially transitive* and in the second that it is *properly ergodic*.

The fact that properly ergodic actions can exist is at first rather surprising. Consider the following example. Let S be the set of all points on the unit circle in the complex plane, and let C be the measure class of the usual Lebesgue "arc length" measure. Let G be the additive group of all the integers, and let $[e^{i\theta}]n = e^{i\theta}e^{i\alpha\pi n}$, where α is some fixed irrational number; in other words, let the generator of the infinite cyclic group G act on S by rotating through an irrational multiple of a complete revolution. Every orbit of this action is countable and hence of measure zero, and intuition suggests that the action cannot be ergodic. Surely it must be possible to collect a number of these "very thin" orbits into a Borel set of positive measure without taking almost all of them. This intuition, however, turns out to be false. Assuming the existence of an invariant Borel set, one has only to consider the Fourier coefficients of its characteristic function to be led at once to the conclusion that it is either a null set or the complement of one. Our action is properly ergodic.

It is the existence of properly ergodic actions which gives ergodic theory its special flavor. If all ergodic actions were essentially transitive, then ergodic theory would consist of the theorem on decomposition into ergodic parts and little else. The ergodic theorem and other results of the theory would reduce to corollaries of the decomposition theorem and elementary facts about harmonic analysis on groups. On the other hand, properly ergodic actions have a "pathological flavor" in that they cannot exist unless the orbit structure of the action of G on the standard Borel space S is of a particularly nasty character. Specifically, one can prove the following theorem. Let S be a standard Borel G space (where G is separable and locally compact), and suppose that there exists a Borel set A in S which meets each orbit just once. Then for every ergodic invariant measure class C in S, the corresponding action is essentially transitive.

Roughly speaking, this theorem says that when one has a properly ergodic action, it is impossible to parametrize the orbits in a reasonable way. In the author's opinion, this apparently pathological character of properly ergodic actions is a reflection of the fact to be explained below that such actions occur in an essential way in mathematical models for random behavior.

The fundamental problem of finding (to within isomorphism) all possible ergodic actions of each separable locally compact group G now breaks naturally into two parts: (a) Find all possible transitive actions, and (b) find all possible properly ergodic actions. Part (a) is much the easier. As we shall see shortly, it is completely equivalent to finding all possible closed subgroups of G (where conjugate subgroups are identified), and this is a problem which can be explicitly solved for many groups G of interest in ergodic theory. For example, if G is the additive group of the real line, the most general closed subgroup other than G itself is the set of all integer multiples of a fixed nonnegative real number λ. To see how this reduction comes about, let H be a closed subgroup of the separable locally compact group G, and let G/H denote the set of all right H cosets Hx. G/H becomes a separable locally compact space if we define the open sets in G/H to be the sets whose inverse images in G are open and the associated Borel structure is standard. Setting $(Hx)\,y = Hxy$, one verifies at once that G/H becomes a transitive standard Borel G space. Conversely, let S be any transitive standard Borel G space and choose a point s_0 in S. It can be proved that the subgroup H_{s_0} of all x in G with $s_0x = s_0$ is closed, and it then follows easily that S is isomorphic as a G space to G/H_{s_0}. Changing s_0 to $s_1 = s_0x$ of course changes H_{s_0} to $H_{s_1} = x^{-1}H_{s_0}x$. Finally, it can be proved that G/H always admits a unique invariant measure class. Indeed, let ν be a Haar measure in G; that is, the unique (up to a multiplicative constant) measure which is finite on compact sets and invariant under right translation. Let ν_1 be any finite measure in the class with ν, and let $\tilde{\nu}_1(E) = \nu_1(\psi^{-1}(E))$, where ψ is the map $x \to Hx$ of G onto G/H. The class $\tilde{\nu}_1$ is independent of the choice of ν_1 and is the unique invariant measure class in G/H. This invariant measure class may or may not contain an invariant measure but if it does, this invariant measure is unique up to a multiplicative constant. More generally, if an ergodic invariant measure class contains an invariant measure, this measure is unique up to a multiplicative constant. If it is finite, then, of course, the measure class contains a unique invariant probability measure.

Part (b) is not only much harder but there are no groups for which it

has been solved except those for which it is trivial. It is not hard to show that compact groups have no properly ergodic actions so that the simplest group which has them is the additive group of all integers. The properly ergodic actions of this group have been much more assiduously studied than those of any other group and a considerable amount of information has been accumulated. However, we are still quite far from anything approaching a complete classification. In spite of these differences in difficulty, it is well to keep in mind that parts (a) and (b) of the classification problem are analogous problems. In particular, the various properly ergodic actions of a given G relate to one another in a manner suggestive of the way that the subgroups of a group behave, and, indeed, it turns out to be fruitful to think of each properly ergodic action as defined by a sort of "ideal" or "generalized" subgroup which we call a "virtual subgroup". We shall not study virtual subgroups systematically in these lectures but will invoke the point of view they suggest from time to time as a heuristic and motivational guide.

Let S_1, C_1 and S_2, C_2 be the spaces and measure classes for actions of the separable locally compact commutative groups G_1 and G_2. Then $S_1 \times S_2$ becomes a standard $G_1 \times G_2$ space if we define $(s_1, s_2)(x, y)$ to be $(s_1 x, s_2 y)$. Moreover, if μ_1 and μ_2 are members of C_1 and C_2, then $\mu_1 \times \mu_2$ is a member of an invariant measure class in $S_1 \times S_2$ whose class depends only on C_1 and C_2 and may be denoted by $C_1 \times C_2$. It is more or less immediate that the resulting action of $G_1 \times G_2$ is ergodic if and only if the actions of G_1 and G_2 are both ergodic. Moreover, the product action is essentially transitive if and only if *both* the G_1 and G_2 actions are transitive. We see in particular then that we can construct properly ergodic actions of a product of other groups whenever we can construct a properly ergodic action of at least one of the factors. It is of course not true that every ergodic action of $G_1 \times G_2$ is isomorphic to a product of ergodic actions of G_1 and G_2, respectively. To see this, we have only to look at the transitive case where the actions are defined by closed subgroups of $G_1 \times G_2$. Those that are products of transitive actions of G_1 and G_2 are those whose defining subgroups are of the rather special form $H_1 \times H_2$, where H_1 and H_2 are closed subgroups of G_1 and G_2, respectively.

Let S be the unit disk $|z| \leqslant 1$ in the complex plane, and let G be the additive group of all integers. Make S into a standard Borel G space by giving it the Borel structure defined by the topology of the complex plane and defining $(z)^n$ for each integer n to be $ze^{2\pi i\alpha n}$, where α is some fixed irrational number. Then μ is invariant, but the action of G defined

by μ and by z, $n \to ze^{2\pi i\alpha n}$ is clearly very far from being ergodic. Given $0 \leqslant r_1 < r_2 \leqslant 1$, the set of all z with $r_1 \leqslant |z| \leqslant r_2$ is invariant of positive measure and has a complement of positive measure. Moreover, it is easy to see that there are no invariant subsets of positive measure on which the action is ergodic. On the other hand, let S_r denote the circle $|z| = r$, and let μ_r denote the "arc length measure" on S_r. Then the measure μ is a "direct integral" of the measures μ_r in the precise sense that for any Borel set E in S, $\mu(E) = \int_0^1 \mu_r(E \cap S_r)\, dr$. Moreover, each S_r is an invariant Borel set, and the pair S_r, μ_r defines an ergodic action of G for all r. We have in an obvious sense decomposed our given action as a "direct integral" or "continuous sum" of ergodic actions.

This example is a typical special case of the general decomposition theorem alluded to above and to whose general formulation we now turn our attention. Let μ be a finite measure in the standard Borel space S, and let φ be a mapping of S onto some space R. Then R becomes a Borel space if we define a subset E of R to be a Borel set whenever $\varphi^{-1}(E)$ is a Borel subset in S and setting $\tilde{\mu}(E) = \mu(\varphi^{-1}(E))$ gives us a measure $\tilde{\mu}$ in R. We think of the sets $\varphi^{-1}(r)$ for $r \in R$ as providing a "fibering" of S analogous to the fibering of the disk into circles in the above example. The fundamental theorem on the fibering or decomposition of measures may be stated as follows.

THEOREM. *Let the fibering R, φ be such that R as a Borel space is countably separated; that is, that there exist countably many Borel subsets $E_1, E_2, \ldots$ such that for any p and q in R with $p \neq q$, there exists j such that $p \in E_j$ and $q \notin E_j$. Then there exists an assignment $r \to \mu_r$ of a measure μ_r in $\varphi^{-1}(r)$ to each $r \in R$ such that for all Borel subsets E of S $\mu(E) = \int \mu_r(E \cap \varphi^{-1}(r))\, d\tilde{\mu}(r)$. This assignment is essentially unique in the sense that if $(r \to \mu_r')$ is another with the same properties, then $\mu_r = \mu_r'$ almost everywhere with respect to $\tilde{\mu}$.*

It is easy to see that if μ^1 and μ^2 are in the same class, then $\tilde{\mu}^1$ and $\tilde{\mu}^2$ are in the same class and $\tilde{\mu}_r^1$ and $\tilde{\mu}_r^2$ are in the same class for almost all r. Since every measure class contains finite measures, the theorem just stated implies an analogous theorem about the fibering of measure classes. Based on this result, we may formulate the fundamental theorem on the decomposition of actions into ergodic parts as follows.

THEOREM. *Let G be a separable locally compact group, and let C be an invariant measure class in the standard Borel G space S. Then there exists a fibering R, φ of S such that the following conditions hold:*

(1) *R is countably separated.*
(2) *Each $\varphi^{-1}(r)$ is an invariant Borel set.*
(3) *In the fibering of C defined by R, φ, almost all of the measure class C_r are invariant and ergodic under the action of G.*

If R', φ' is a second fibering with the same properties, then there exists a one-to-one Borel map θ of almost all of R onto almost all of R' such that the action defined by C_r in $\varphi^{-1}(r)$ is isomorphic so that defined by $C_{\theta(r)}$ in $(\varphi')^{-1}(\theta(r))$.

4. Group Duality and Examples of Properly Ergodic Actions

Let S and C denote the standard Borel space and invariant measure class, respectively, for some action of the separable locally compact G. If C contains an invariant measure μ, we may imitate the construction of Koopman mentioned in the introduction and define a "unitary representation" $x \to V_x^{\mu}$ of G in the Hilbert space $\mathscr{L}^2(S, \mu)$ by defining $V_x^{\mu}(f)(s)$ to be $f(sx)$. The representation V can in fact be constructed even when C does not contain an invariant measure. One simply chooses μ to be an arbitrary member of C and changes the definition to read

$$V_x^{\mu}(f)(s) = \sqrt{\rho(s, x)}\, f(sx),$$

where $\rho(s, x)$ is a Borel function which for each x in G is a Radon Nikodym derivative of μ_x with respect to μ. Inserting the factor $\sqrt{\rho(s, x)}$ compensates for the noninvariance of μ in such a fashion as to render the operators V_x^{μ} unitary. It is easily seen that $(V_x^{\mu}(\varphi) \cdot \psi)$ is a Borel function of x for each φ and ψ in the Hilbert space $\mathscr{L}^2(S, \mu)$ and that $V_{xy}^{\mu} = V_x^{\mu} V_y^{\mu}$ for all x and y in G. Quite generally, one defines a unitary representation of separable locally compact group representation W of G to be a homomorphism $x \to W_x$ of G into the group of all unitary operators in some separable Hilbert space $H(W)$ such that $x \to W_x$ has one and hence all of the equivalent properties.

(1) For each φ and ψ in $H(W)$, $x \to (W_x(\varphi) \cdot (\psi))$ is a Borel function.
(2) For each φ and ψ in $H(W)$, $x \to (W_x(\varphi) \cdot (\psi))$ is continuous.
(3) For each φ in $H(W)$, $x \to W_x(\varphi)$ is continuous.

Moreover, one defines two unitary representations W^1 and W^2 to be *equivalent* if there exists a unitary operator U from $H(W^1)$ to $H(W^2)$ such that $UW_xU^{-1} = W_x^2$ for all x in G. For many purposes, one does

not distinguish between equivalent representations. Since it is quite easy to show that V^{μ_1} and V^{μ_2} are equivalent unitary representations of G whenever μ_1 and μ_2 are in C, it follows that our construction is essentially independent of the choice of μ and gives us a natural mapping of isomorphism classes of actions of G into equivalence classes of unitary representations of G. While this mapping is far from being one-to-one, it provides a very useful partial classification of actions, and, in fact, for a certain class of properly ergodic actions is one-to-one.

Given two unitary representations W^1 and W^2 of the same G, their direct sum $W^1 \oplus W^2$ is defined as the representation whose space is $H(W^1) \oplus H(W^2)$ such that $(W^1 \oplus W^2)_x = W_x{}^1 \oplus W_x{}^2$. Similarly, one defines $W^1 \oplus W^2 \cdots \oplus W^n$ and, more generally, $W^1 \oplus W^2 + \cdots$. It is easy to see that the unitary representation defined by the direct sum of finitely or countably many actions is equivalent to the direct sum of those defined by the actions separately. Moreover, one can define "direct integral" for representations and prove a corresponding result for the unitary representation defined by a direct integral of other actions. Thus, one is interested primarily in finding the unitary representations defined by ergodic actions. A closed subspace M of the Hilbert space $H(W)$ of a unitary representation W is said to be *invariant* if $W_x(\varphi) \in M$ for all $\varphi \in M$. Defining $W_x{}^M$ to be the operator W_x restricted to elements of M, one obtains a new representation $x \to W_x{}^M$, called the subrepresentation, defined by the invariant subspace M. An obvious argument shows that the orthogonal complement $M^\perp$ of M is invariant whenever M is and that W is equivalent to the direct sum of W^M and $W^{M^\perp}$. Thus W is equivalent to a proper direct sum if and only if $H(W)$ has nontrivial closed invariant subspaces. A representation which has no nontrivial closed invariant subspaces is said to be *irreducible*. One attempts to analyze representations by decomposing them as direct integrals of irreducible representations. In the special case in which G is compact, every unitary representation is a *direct sum* of representations which are finite dimensional and irreducible. Moreover, whether or not G is compact, whenever a unitary representation can be written as direct sum of irreducibles, the irreducibles which occur are uniquely determined to within equivalence by specifying which irreducibles occur and with what multiplicity.

Ergodic theory has up to now been mainly developed for groups G which are commutative as well as separable and locally compact, and when G is commutative its unitary representation theory takes an especially simple form because of a theorem asserting that an irreducible

unitary representation of a commutative group is necessarily one dimensional. Note that if V is a one-dimensional unitary representation, then $V_x = \chi(x)I$, where I is the identity operator and χ is a continuous complex-valued function on G, such that $\chi(xy) = \chi(x)\,\chi(y)$ and $|\chi(x)| = 1$. Conversely, given any such function; that is, any continuous homomorphism $x \to \chi(x)$ of G into the group of all complex numbers of modulus one, then $V_x = \chi(x)I$ is a one-dimensional irreducible unitary representation of G. The functions χ are called the one-dimensional characters of G, and in the commutative case, the "one dimensional" is usually omitted. Determining the irreducible unitary representations of a commutative G is equivalent to determining the characters of G.

A fact which will be of great importance for us is that the set $\hat{G}$ of all one-dimensional characters of a given separable locally compact G is itself a separable locally compact group. Indeed, it is obvious that the product $\chi_1\chi_2$ of two one-dimensional characters is again such and so are the complex conjugates $\bar{\chi}_1$ and $\bar{\chi}_2$ of χ_1 and χ_2. Thus $\hat{G}$ is a commutative group with respect to ordinary pointwise multiplication. For each $\epsilon > 0$ and each compact subset C of G, one defines $N_{C,\epsilon}$ to be the set of all $\chi \in \hat{G}$ with $|\chi(x) - 1| < \epsilon$ for all $x \in C$. The sets $N_{C,\epsilon}$ can be shown to form a complete set of neighborhoods of the identity for a topology in $\hat{G}$ which makes it into a separable locally compact group. In the case in which G is commutative, $\hat{G}$ is called the *dual group* or the *character group of* G.

Perhaps the simplest example is that in which G is the additive group of all integers. Then $\chi(n) = (\chi(1))^n$ by definition, and χ is completely known as soon as we know $\chi(1)$. On the other hand, if we choose a to be an arbitrary complex number of modulus one, then setting $\chi(n) = a^n$ defines a character. $\hat{G}$ in this case is isomorphic to the group of all complex numbers of modulus one or, equivalently, to the quotient group obtained from the additive group of the real line by factoring out the subgroup of all integer multiples of 2π. Moreover, the topology of $\hat{G}$ turns out to be the usual topology of the complex numbers so that $\hat{G}$ is a *compact* topological group. This is a special case of a general result. It is quite easy to prove that $\hat{G}$ is compact whenever G is discrete. Another simple and important example is that in which G is the additive group of all real numbers. If y is any real number, then setting $\chi_y(x) = e^{iyx}$ clearly defines a character, and it may be shown that every character is of the form χ_y for some uniquely determined y. Since $\chi_{y_1}\chi_{y_2} = \chi_{y_1+y_2}$, it follows in this case that $\hat{G}$ is also isomorphic to the additive group of the real line. Moreover, one verifies that the topology of $\hat{G}$ as the dual

of G agrees with the usual real-line topology. Other duals of the important examples such as the additive group of Euclidean n space can be computed from the above using the following easy theorem about direct products: The characters of $G_1 \times G_2$ are precisely the functions $x, y \to \chi_1(x)\chi_2(y) = \chi_1\chi_2(x, y)$, where χ_1 and χ_2 are arbitrary members of $\hat{G}_1$ and $\hat{G}_2$, respectively. Moreover, $\chi_1, \chi_2 \to \chi_1\chi_2$ is an isomorphism of $\hat{G}_1 \times \hat{G}_2$ and $\widehat{G_1 \times G_2}$ as topological groups. It follows in particular that the most general character of the additive group of Euclidean n space is

$$x_1, \ldots, x_n \to e^{i(x_1y_1+\cdots+x_ny_n)}$$

where $y_1, \ldots, y_n$ are fixed real numbers.

Consider $\chi(x)$, where x is a *fixed* element of G and χ is a *variable* element of $\hat{G}$. The resulting complex-valued function on $\hat{G}$ is a member f_x of $\hat{\hat{G}}$, the character group of $\hat{G}$. Indeed, $f_x(\chi_1\chi_2) = \chi_1\chi_2(x) = \chi_1(x)\,\chi_2x = f_x(\chi_1)f_x(\chi_2)$, and continuity is equally easy to establish. Thus we have a natural mapping $x \to f_x$ of G into $\hat{\hat{G}}$ which is seen at once to be a continuous homomorphism. The celebrated Pontryagin van Kampen duality theorem says much more; the mapping $x \to f_x$ is one to-one and onto, and its inverse is continuous so that it defines an isomorphism between G and $\hat{\hat{G}}$ as topological groups. If we use this isomorphism to identify G and $\hat{\hat{G}}$, we see that the relation between G and $\hat{G}$ is a reciprocal one. G is just as much the dual of $\hat{G}$ as $\hat{G}$ is the dual of G. Separable locally compact commutative groups (actually, the duality theorem is true without separability) occur in dual pairs. Moreover, in this duality, the duals of compact group are always discrete so that every compact commutative group is the dual of some discrete one and conversely.

Let θ be a continuous homomorphism of G_1 into G_2, where G_1 and G_2 are locally compact commutative groups. Then for each $\chi \in \hat{G}_2$, $x \to \chi(\theta(x))$ is a member $[\chi]\,\theta^*$ of $\hat{G}_1$, and $\chi \to [\chi]\,\theta^*$ is a continuous homomorphism of $\hat{G}_2$ into $\hat{G}_1$, called the dual of θ. Of course, $\theta^{**} = \theta$ if we identify $\hat{\hat{G}}$ with G. When G_1 is a closed subgroup of G_2 and θ is the identity mapping, then θ^* maps $\hat{G}_2$ onto $\hat{G}_1$, and its kernel is the closed subgroup $G_1^{\perp}$ of $G_2^{\perp}$ consisting of all χ_1 with $\chi_1(x) = 1$ for all x in G_1. The subgroup $G_1^{\perp}$ is called the annihilator of G_1. Of course, $G_1^{\perp\perp} = G_1$ when $\hat{\hat{G}}_1$ is identified with G_1.

Let $K_1, K_2, \ldots$ be an infinite sequence of compact commutative groups (some or all of which may be finite), and let K denote their complete direct product $\prod_{j=1}^{\infty} K_j$. With the product topology, K is a compact

commutative group which is separable if all of the K_j are. Now, let $\chi_1, \chi_2, \chi_3, \ldots$ be a sequence such that $\chi_j \in \hat{K}_j$ and *such that* $\chi_y \equiv 1$ *for all but a finite number of* j. Then $\prod_{j=1}^{\infty} \chi_j(x_j)$ is well-defined for all $x_1, x_2, \ldots \in K = \prod_{j=1}^{\infty} K_j$ and defines a character of K. It is not difficult to show that every character of K has this form for some uniquely determined sequence $\chi_1, \chi_2 \ldots$. The dual of $\prod_{j=1}^{\infty} K_j$ is thus naturally isomorphic to a certain *discrete subgroup* of the product $\prod_{j=1}^{\infty} \hat{K}_j$. Members of this class of examples will occur several times in the sequel.

Let S be a standard Borel G space, where G is a separable locally compact *commutative* group; let C be an ergodic invariant measure class in S; and let V be the associated unitary representation of G. One says that the action of G defined by S and C has a *pure point spectrum* if V is a direct sum of irreducible representation (i.e., of one-dimensional characters), and the set of characters which appear is called the *spectrum* of the action. As we shall see, it is possible to give a complete analysis of all possible ergodic actions with pure point spectrum. First of all, it is not difficult to show that if $H(V)$ has any one-dimensional invariant subspaces at all, then C contains a finite invariant measure. Thus, if V has a pure point spectrum, then C contains a unique invariant probability measure μ. If g is contained in the invariant subspace corresponding to the character χ, then for all $x \in G$, $g(sx) = \chi(x)\, g(s)$ for almost all s in S, whence it follows that $|g(sx)| \equiv |g(s)|$ in that by ergodicity, $|g(s)|$ is almost everywhere constant and in particular may be taken to be bounded. Hence, if $g_1(sx) \equiv \chi_1(x)\, g_1(s)$ and $g_2(sx) = \chi_2(x)\, g(s)$, then $g_1\bar{g}_2$ is in $\mathscr{L}^2(S, \mu)$ and $g_1\bar{g}_2(sx) \equiv \chi_1(x)\, \bar{\chi}_2(x)\, \bar{g}_1\bar{g}_2(s)$. If $\chi_1 = \chi_2$, it follows that $g_1\bar{g}_2(sx) \equiv g_1\bar{g}_2(s)$ and, hence, by ergodicity, that $g_1\bar{g}_2$ is a constant. Hence, g_2 is a constant multiple of g_1. On the other hand, if $\chi_1 \neq \chi_2$, then the subspace spanned by $g_1\bar{g}_2$ is invariant and is associated with the character $\chi_1\bar{\chi}_2 = \chi_1\chi_2^{-1}$. In other words, each character which occurs, occurs with multiplicity one, and the set of characters making up the spectrum is a finite or countable subgroup of $\hat{G}$. It is a remarkable fact that an action with a pure point spectrum is determined to within isomorphism by its spectrum. Moreover, every countable subgroup of $\hat{G}$ occurs as the spectrum of some ergodic action of G with a pure point spectrum. More specifically, one can prove the following.

THEOREM. *Let Γ be a finite or countable subgroup of $\hat{G}$, where G is a separable locally compact acommutative group and Γ is not necessarily closed. Let θ be the natural homomorphism* of Γ *as a discrete countable group into $\hat{G}$. Let θ^* denote the dual homomorphism of G into the compact dual $\hat{\Gamma}$*

of Γ and make $\hat{\Gamma}$ into a standard Borel G space by setting $[\chi]x = \chi\theta^*(x)$ for all $\chi \in \hat{\Gamma}$ and all $x \in G$. Then Haar measure in $\hat{\Gamma}$ is invariant and ergodic with respect to the G action. The resulting ergodic action of G has pure point spectrum with Γ as spectrum, and every such action of G is isomorphic to this one. The actions is properly ergodic if and only if Γ is not closed. When Γ is closed, the associated transitive action of G is that on the coset space of $G/\Gamma^{\perp}$.

If follows at once from this theorem that the isomorphism classes of properly ergodic actions with a pure point spectrum for a given G correspond one-to-one to the nonclosed countable subgroups of G and, hence, exist in great variety whenever G is not compact. In the special case in which G is the additive group of the integers so that $\hat{G}$ is the multiplicative group of all complex numbers of modulus one, the simplest possible example for Γ is a nonfinite cyclic group, that is, the group of all $e^{in\theta}$ where $\theta/2\pi$ is irrational. This class of examples is precisely the one discussed in Section 3 and the only class we have presented up to the beginning of the current discussion. We see, in particular, that every noncompact G admits at least continuum many mutually nonisomorphic properly ergodic actions.

The examples of actions with pure point spectra just constructed are such that the measure-preserving transformations $s \to sx$ are all translations on a compact commutative group $\hat{\Gamma}$. Now Haar measure on a compact group is preserved not only by all left and right translations but also by all automorphisms of the group. We shall now investigate groups of automorphism of separable compact commutative groups and so obtain examples of ergodic actions which not only fail to have pure point spectra but are in various ways at the opposite extreme from actions with pure point spectra. Let K be a separable compact commutative group, and let α be an automorphism of K. Then setting $[k]n = \alpha^n(k)$ and taking C to be the measure class of Haar measure μ, we obtain an action of Z, the additive group of all the integers. Let V denote the associated unitary representation of Z. Now if μ is chosen so that $\mu(K) = 1$, the characters of K form a complete orthonormal set. Indeed, by the theorem on pure point spectra cited above, the transitive action of K defined by the identity subgroup has $\hat{K}$ as its spectrum, and this implies that the members of $\hat{K}$ have the property in question. As first noticed by Halmos, this fact may be used to determine the structure of V since $V_n(\chi) = [\chi](\alpha^*)^n$, where α^* is the automorphism of $\hat{K}$ dual to α. Let O_χ denote the "orbit of χ", that is, the set of all $[\chi](\alpha^*)^n$ for

fixed χ and variable n. Then the characters in each orbit span a subrepresentation of V and V is the direct sum of these subrepresentations. This subrepresentation associated with the orbit of χ is at once seen to be equivalent to the representation associated with the corresponding *transitive* action of Z on O_χ and hence determined by the subgroup Z_0 of all n with $\chi(\alpha^*)^n = \chi$. This subgroup Z_0 is either the group of all multiples of some fixed positive integer m_0 or consists of the identity alone. In the first case, the orbit has m_0 elements and the corresponding component of V is the direct sum of the m_0 characters $n \to \exp(2\pi il/m_0)$, $l = 0, 1, 2, \ldots m_0 - 1$. In the second case, the orbit is infinite and the corresponding component of V is the representation W whose space is the space of all complex-valued functions $n \to f(n)$ on Z such that $\sum_{n=\infty}^{\infty} |f(n)|^2 < \infty$ with $W_n(f)(m) = f(m + n)$. This is called the *regular representation* of Z and an easy application of Fourier analysis shows that it has no irreducible subrepresentations at all. Notice that the one-dimensional identity occurs once in the component associated with each *finite* orbit. On the other hand, where there is a finite invariant measure in the measure class, it follows at once from the definition of ergodicity that an action is ergodic if and only if the only invariant square summable functions are the constants. Since the identity of $\hat{K}$ is always an orbit, the action will be ergodic if and only if every other orbit is infinite. Thus we have outlined the proof of Halmos' theorem: Let α be an automorphism of the separable compact commutative group K, and let α^* denote the dual automorphism of the discrete character group $\hat{K}$. Then the action of the integers on K and Haar measure defined by α is ergodic if and only if every α^* orbit in $\hat{K}$ other than the identity has infinitely many elements. When it is ergodic, the associated unitary representation of the integers is the direct sum of the one-dimensional identity and has as many replicas of the regular representation as there are infinite orbits. In particular, the point spectrum contains just one element.

To produce examples of pairs K, α in which the conditions of Halmos' theorem hold so that an ergodic action of Z is defined, one can proceed as follows. Let A be a separable compact commutative group which may be finite, and let K be the group A^Z of all functions from Z to A made into a compact separable topological group by regarding it as a direct product of replicas of A. Then the mapping $f \to \alpha(f)$, where $\alpha(f)(n) = f(n + 1)$, is clearly an automorphism of K. We have already discussed the character group of K, and a straightforward computation shows that α^* has countably many infinite orbits and no finite orbits other than

the identity. The corresponding action of Z is thus ergodic, and the associated group representation V is the direct sum of the one-dimensional identity and of countably many replicas of the regular representation of Z. Notice that V is the same no matter what A is. Thus it is not immediately clear whether changing A changes the isomorphism class of the associated action as well. It is easy to see that the isomorphism class of the action depends only on the structure of the measure space A becomes when Haar measure is introduced, and, hence, only on the cardinal number of A. When A is finite with n elements, the corresponding α is called the *n shift*, and for many years it was an unsolved problem to decide whether the n shift and the n' shift could define isomorphic actions when n and n' were different. In fact, for no two distinct values of n could it be decided whether or not the actions were isomorphic. The problem was finally settled in 1959 by the work of Kolmogorov and Sinai who introduced a new invariant based on the concept of entropy in information theory and showed that the n shift and the n' shift define isomorphic actions if and only if $n = n'$. We shall have more to say about this in later sections.

Thus far, we have produced examples of properly ergodic actions other than those having pure point spectra only for the special case of the infinite cyclic group Z. However, the method is easily extended to arbitrary countable discrete groups—even noncommutative ones. First of all, Halmos' result about infinite cyclic groups of automorphisms of K extends at once to arbitrary countable groups. One looks at the orbits of the dual action on $\hat{K}$ and finds as before that the action on K is ergodic if and only if there are no nontrivial finite orbits. If it is ergodic, then the associated representation V is a direct sum of the one-dimensional identity and of a representation associated with each orbit. The representation associated with the orbit of χ is, however, not necessarily the regular representation of the group G but the representation associated with the transitive action of G on the coset space G/H, where H is the subgroup leaving a point fixed. This representation can be shown to have no finite-dimensional irreducible subrepresentations. To produce examples, let G be an arbitrary countably infinite discrete group, and let A be as above. Let $K = A^G$ denote the group of all functions from G to A topologized as a direct product of replicas of A. Then for each $x \in G$, $f \to f'$ is an automorphism of K if $f'(y) = f(yx^{-1})$. The dual action of G on $\hat{K}$ is easily seem to have no nontrivial finite orbits and to be such that all "isotopy subgroups" H for the infinite orbits reduce to the identity. Thus, Haar measure on K is ergodic under G,

and the associated group representation is a direct sum of the one-dimensional identity and of countably many replicas of the regular representation of G. A further class of examples may be obtained by replacing $K = A^G$ by the group of all functions from some infinite coset space G/H to A.

There is an analogue of the above construction which applies to arbitrary separable locally compact groups but it cannot be described immediately. Indeed, in the special case of the additive group of the real line, it reduces to the ergodic action of the real line associated with the celebrated Wiener measure in the space of continuous functions. We shall give details in a later section.

5. The Ergodic Theorem and Its Significance for the Theory of Stochastic Processes

Recall that a discrete stationary stochastic process is defined by giving three things: (1) A standard Borel Z space Ω, where Z is the additive group of the integers, (2) an invariant probability measure α in Ω, and (3) a real- or complex-valued Borel function f defined on Ω. (1) and (2) together define an action of Z, and we may accordingly classify processes by classifying the underlying actions of Z. Of course, there will be many distinct processes associated with a given action corresponding to different choices for f, but it will turn out that the nature of the underlying action sharply limits the possibilities for the process and that, conversely, we may classify actions by the nature of the processes which can be built upon them. This interplay between the classification and properties of actions and processes will be one of our principal themes.

To begin with, let us examine the meaning of the decomposition theorem of Section 3 for the nature of a process Ω, α, f. By that theorem, Ω may be fibered into Z invariant Borel subsets Ω_r each of which is equipped with a finite invariant measure α_r which is ergodic under the action of Z. The measures α_r have the property that

$$\alpha(E) = \int \alpha_r(E \cap \Omega_r)\, d\tilde{\alpha}(r)$$

for all Borel subsets E of Ω and are essentially uniquely determined by this fact. Moreover, the α_r are all probability measures. Let f_r denote the restriction of f to Ω_r. Then for each r, the triple Ω_r, α_r, f_r defines a

discrete stationary stochastic process for which the underlying action is ergodic. Moreover, in an obvious sense, one may think of the original process as being a direct integral of these ergodic processes. Now consider an arbitrary sample function $n \to f([\omega]n)$ of our process. The point ω will belong to one of the fibers Ω_r, and since Ω_r is Z invariant, $f([\omega]n) \equiv f_r([\omega]n)$ for all n. In other words, *every sample function coincides with a sample functions of one of the ergodic components of the process.* This simple fact is of considerable significance since, in many contexts, it permits us to confine our attention to the ergodic case. Note in particular that in the theory of time-series analysis, one is presented with a part of the past of *one* sample function and one wants to make deductions about the process and about the probability of various events in the future. A little reflection makes it clear that the past of one sample function can at most yield information about the ergodic part Ω_r in which that sample function lies and that in making predictions one can assume that the process is Ω_r, α_r, f_r instead of Ω, α, f. In other words, in the prediction theory of stationary stochastic processes, one may always assume that one is dealing with a process whose underlying action is ergodic.

Since every sample function is the sample function of an ergodic component of the process, it is interesting to investigate the effect on the sample functions of the division of ergodic actions into these which are properly ergodic and those which are essentially transitive. Now the only ergodic actions which arise from stationary stochastic processes have finite invariant measures, and the only transitive actions of Z which have finite invariant measures are those whose defining subgroup contains more than 0. Hence the transitive components in the decomposition of an arbitrary stationary stochastic process are *finite* sets on which 1 acts by a cyclic permutation of order k, where k is the number of elements in the set. Clearly, the sample functions associated with the elements of such a finite invariant set are periodic with period k. Thus for a stationary stochastic process whose underlying action is essentially transitive, almost all sample functions will be periodic functions with a fixed period k. More generally, given an arbitrary stationary process, we have the following dichotomy. Either (a) almost every sample functions is periodic or (b) for a set of r values of positive measure, the ergodic action of Z on Ω_r, α_r is *properly ergodic.*

This simple observation tells us that *if properly ergodic actions of the integers did not exist then for every stationary stochastic process, almost every sample function would be periodic. It is in this sense that one can regard*

the existence of properly ergodic actions as a necessary condition for the existence of a realistic mathematical model for random phenomenon.

Given a sample function $n \to f([\omega]n)$ of a discrete stationary process, one can consider the mean

$$\frac{f([\omega]1) + f([\omega]2) + \cdots + f([\omega]N)}{N}$$

of N successive observations. Intuitive ideas about the nature of probability suggest that this mean should have a limit as N tends to ∞ which is equal to the common expected value of the random variables of the process (that is to $\int f(\omega)\, d\alpha(\omega)$) whenever the latter exists. In mathematical terms, one expects that whenever f is in $\mathscr{L}'(\Omega, N)$, then

$$\frac{f([\omega]1) + f([\omega]2) + \cdots + f([\omega]N)}{N}$$

has a limit as $N \to \infty$ and that the limit is equal to $\int f(a)\, d\alpha(a)$. Of course, since f can be arbitrary on a set of α measure zero without affecting its integrability, one can at most hope that the limit in question exists almost everywhere. The celebrated pointwise ergodic theorem proved by G. D. Birkhoff in late 1931 states precisely this *provided only that the action is ergodic.* More exactly, Birkhoff's theorem (in its discrete form) says that whenever f is in $\mathscr{L}^1(\Omega, \alpha)$ and α is invariant, then

$$\lim_{N\to\infty} \frac{f([\omega]1) + f([\omega]2) + \cdots + f([\omega]N)}{N}$$

exists for almost all ω in Ω and that this limit is almost everywhere equal to $\int f(\omega)\, d\alpha$ whenever the action is ergodic. The existence of the limit is the difficult point to establish and this is quite independent of the ergodicity hypothesis. It is obvious that the limit is measurable and invariant under the action and hence must be almost everywhere constant if ergodicity holds. To prove that this constant must be equal to $\int f(\omega)\, d\alpha$ is then an elementary exercise in the theory of the Lebesgue integral. When the action is not ergodic, the limit in question will depend upon the ergodic component to which ω belongs and only "by accident" will be equal to $\int f(\omega)\, d\alpha$. Thus the concept of ergodicity is essential in order that the so-called "strong law of large numbers" should hold in probability theory and the content of the G. D. Birkhoff ergodic theorem is that it does, in fact, hold whenever the underlying action is ergodic.

Actually, in its original form, the ergodic theorem was proved for actions of the real line rather than actions of the integers and correspondingly took the form of asserting the almost everywhere existence of

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f([\omega])t \, dt.$$

Moreover, it was proved under somewhat more restrictive hypotheses on Ω and the action. However, it is easy to derive one form of the theorem from the other and the superfluity of some of the original hypothesis was soon recognized. In proving his theorem, Birkhoff was motivated by the needs of statistical mechanics (cf. Section 9) rather than those of the theory of stochastic processes, and the connection with stochastic processes was not pointed out until two years later. As mentioned in the introduction, Birkhoff's work was inspired by von Neumann's proof of the (much easier) mean ergodic theorem and it was von Neumann who called attention to the significance of the concept of ergodicity in implying equality for the two means. Before the proof of the ergodic theorem, the strong law of large numbers was known to hold only in rather special cases such as that in which the random variables are independent.

Of course, the ergodic applies equally well to the action $\omega, n \to [\omega](-n)$ and so tells us that not only does

$$\int f(\omega)\, d\alpha = \lim_{N\to\infty} \frac{f([\omega]1) + f([\omega]2) + \cdots + f([\omega]N)}{N}$$

but that also

$$\int f(\omega)\, d\alpha = \lim_{N\to\infty} \frac{f([\omega](-1)) + f([\omega](-2)) + \cdots + f([\omega](-N))}{N}.$$

In other words, one can compute the invariant $\int f(\omega)\, d\alpha$ of the process from the "complete past" $\cdots f([\omega](-n)), f([\omega](-n+1)) \cdots f([\omega]-1)$ of almost any sample function. We show next that one can do much better than this. When ergodicity holds, one can reconstruct the entire process from almost any sample function. The procedure we shall outline for doing this involves both the ergodic theorem and a theorem about locally compact commutative groups called the Bochner–Herglotz theorem. We begin by describing the latter which will be used again in the sequel. Let G be a locally compact commutative group, and let χ

be a member of $\hat{G}$, the character group of G. Let $x_1, \ldots, x_n$ be a finite set of elements of G, and let $c_1, \ldots, c_n$ be complex numbers. Then, of course, $(c_1\chi_1(x_1) + \cdots c_n\chi(x_n))\overline{(c_1\chi_1(x_1) + \cdots c_n\chi_n(x_n)} \geqslant 0$. But the left-hand side is equal to $\sum_{i,j} c_i\bar{c}_j\chi(x_ix_j^{-1})$. In other words, χ is "positive definite" where a complex-valued function f on a topological group G is defined to be *positive definite* wherever it is continuous and satisfies the inequality, $\sum_{i,j=1}^{n} c_i\bar{c}_j f(x_ix_j^{-1}) \geqslant 0$ for arbitrary choices of the group elements $x_1, x_2, \ldots, x_n$ and the complex numbers $c_1, \ldots, c_n$. More generally, let μ be any finite measure on $\hat{G}$, and let $\hat{\mu}(x) = \int \chi(x)\, d\mu(\chi)$. One verifies at once that $\hat{\mu}$, the Fourier transform of μ, is a positive definite function on G. The Bochner–Herglotz theorem asserts that the mapping $\mu \to \hat{\mu}$ is injective and has every positive definite function in its range. In other words, for every positive definite function g on G, there exists a unique finite measure μ on $\hat{G}$ such that $g = \hat{\mu}$.

Given a real-valued discrete stationary stochastic process, a simple change of variable such as $y \to e^{i \arctan y}$ converts it into a discrete stationary process whose values lie in the compact group T of all complex numbers of modulus one. Then replacing Ω by T^∞ the group of all functions from the integer to T, we see that Ω is a compact commutative group—the direct product of countably many replicas of T. If $\omega \in \Omega = T^\infty$, then $[\omega](m) = \omega(n + m)$, and our basic random variable of which the rest are translates will be $\omega \to \omega(0) = g(\omega)$. Our process will be fully described as soon as we are given the invariant probability measure μ in Ω, and we shall show how to compute it from the complete past of almost any sample function. The basic idea is that by the Bochner–Herglotz theorem it is enough to determine the Fourier transform $\hat{\mu}$ of μ, and $\hat{\mu}$, being a function on the countable dual group $\hat{\Omega}$, takes on only countably many values. We shall use the ergodic theorem to derive an explicit formula for computing each of these values. Given any pair of integers n, m, let us define $\chi_{n,m}(\omega)$ to be $[\omega(m)]^n$. It is clear that $\chi_{n,m}$ is in $\hat{\Omega}$ and easy to show that every member of $\hat{\Omega}$ is uniquely a finite product of the form $\chi_{n_1m_1}\chi_{n_2,m_2} \cdots \chi_{n_k,m_k}$, where $m_1 < m_2 \cdots < m_k$ and $n_1, n_2, \ldots, n_k$ are arbitrary. It suffices then to compute

$$\int \omega(m_1)^{n_1}\omega(m_2)^{n_2}\omega(m_k)^{n_k}\, d\mu(\omega)$$

for every $2k$-tuple of integers $m_1 < m_2 \cdots < m_k, \ldots n_1, n_2, \ldots, n_k$. But $\omega \to \omega(m_1)^{n_1}\omega(m_2)^{n_2}\omega(m_k)^{n_k}$ is bounded and measurable and therefore integrable. It follows then from the ergodic theorem that

$$\hat{\mu}(\chi_{n_1,m_1} \cdots \chi_{n_k,m_k})$$
$$= \int \omega(m_1)^{n_1} \cdots \omega(m_k)^{n_k}\, d\mu(\omega)$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \omega(m_1 - n)^{n_1}\, \omega(m_2 - n)^{n_2} \ldots\, \omega(m_k - n)^{n_k}$$
$$= \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \omega(m_1 - m_k - n)^{n_1}\, \omega(m_2 - m_k - n)^{n_2} \ldots\, \omega(-n)^{n_k}$$

for almost ω. But this final expression involves only $\omega(j)$ for $j \leqslant 0$ and, hence, is known whenever the complete past of the sample function is known. Note that each value of $\hat{\mu}$ is just an average of product of powers of "shifts" of the sample past.

Although the argument uses ergodicity in an essential way, the construction of $\hat{\mu}$ can be carried out in any case and leads to a function on $\hat{\Omega}$ which is readily seen to be positive definite and hence the Fourier transform of a unique measure ν in Ω. Moreover, ν is a probability measure since $\hat{\nu}(e) = 1$. What is the relationship of this probability measure ν to the nonergodic μ with which we started? Consider the decomposition of μ into ergodic components described above. Each component lies in an invariant subset of Ω, and every point Ω lies in one of these invariant subsets. Clearly, ν is just the (normalization) of the ergodic component of μ associated with the invariant subset to which our particular sample function belongs. In other words, whether an original process is ergodic or not, almost every sample function will belong to an ergodic component, and the past of almost all of the sample functions belonging to this component can be used just as above to compute the measure in this ergodic component. On the other hand, these sample functions yield no information about the rest of the process. Note however that from the point of view of predicting the future from the past, one is interested only in the ergodic component to which the given sample function belongs. *Hence, as already indicated above, in prediction problems there is no loss in generality in assuming that the underlying action is in fact ergodic.*

6. The Spectra of Actions and Processes

Let G be a separable locally compact group. As explained in Section 4, every measure class preserving action of G is canonically associated with a unitary representation of G whose equivalence class is therefore an

invariant of the action. When G is commutative and the action is ergodic, a special role is played by those actions for which the associated unitary representation is a discrete direct sum of irreducibles. In this case (as we saw in Section 4), the isomorphism class of the action is uniquely determined by the equivalence class of the representation. This equivalence class is, in turn, uniquely determined by a countable subgroup of $\hat{G}$ which is called the *spectrum* of the action. When the representation V is not a discrete direct sum of irreducibles, it need not determine the isomorphism class of the actions, but its equivalence class is still an important invariant. Moreover, this equivalence class is not describable by anything so elementary as a subset of $\hat{G}$. The analogue of the spectrum of an action must be defined in a more subtle fashion.

In the special case in which G is the additive group of the real line, Stone's theorem tells us that any unitary representation V of the real line may be expressed uniquely in the form $V_t = e^{itH}$, where H is a self-adjoint operator and two unitary representations V^1 and V^2 are equivalent if and only if the corresponding self-adjoint operators H_1 and H_2 are equivalent in the sense that $H_2 = UH_1U^{-1}$ for some unitary operator U. In this case then, the problem of determining the equivalence classes of unitary representations of G is equivalent to the problem of determining the equivalence classes of self-adjoint operators. This problem is trivial when H_1 and H_2 are diagonalizable, it being obvious that they are then equivalent if and only they have the same eigenvalues with the same multiplicities. When H_1 and H_2 have a so-called "continuous spectrum", the problem is much more difficult but has been completely solved. The case in which H_1 and H_2 are bounded was analyzed by Hahn and Hellinger in extension of Hilbert's spectral theorem. A few years later in the 1920's, the theorems of Hilbert, Hahn, and Hellinger were generalized to the unbounded case and thrown into abstract form by von Neumann and Stone. A detailed exposition of the unitary equivalence theory appears as Chapter VII in Stone's now classic treatise and a more modern version in a book by Halmos entitled "Introduction to Hilbert space and the theory of spectral multiplicity". When these theorems about self-adjoint operators are translated into corresponding theorems about unitary representations of the additive group of the real line, they take a form which is meaningful for arbitrary separable locally compact commutative groups. These more general theorems turn out to be true and provide a complete analysis of all equivalence classes of unitary representations for every separable locally compact commutative group G. In the next few paragraphs, we shall describe the analysis thus made possible.

Let G be a separable locally compact group, and let V be a unitary representation of G. In the special case in which V is a discrete direct sum of irreducibles, let χ_1, χ_2,... denote the members of $\hat{G}$ corresponding to the irreducible constituents which actually occur, and for each j, let P_{χ_j} denote the projection operator on the maximal subspace on which V reduces to the form $V_x = \chi_j(x)I$, where I is the identity. Then for each pair φ, ψ of vectors in $\mathscr{H}(V)$, we have $V_x(\varphi) = \sum_j \chi_j(x) P_{\chi_j}(\varphi)$, $\psi = \sum_j P_{\chi_j}(\psi)$, and $(V_x(\varphi) \cdot \psi) = \sum_j \chi_j(x)(P_{\chi_j}(\varphi) \cdot P_{\chi_j}(\psi) = \sum_j \chi_j(x)(P_{\chi_j}(\varphi) \cdot \psi)$. Hence V can be completely reconstructed given the χ_j and the projections P_{χ_j}. Now for each Borel subset E of $\hat{G}$, let $P_E = \sum_{\chi_j \in E} P_{\chi_j}$ where the sum is to be interpreted in the sense of strong convergence; i.e., $P_E(\varphi) = \sum_{\chi_j \in E} P_{\chi_j}(\varphi)$ for every φ in $\mathscr{H}(V)$. Then $E \to P_E$ is a so-called "projection valued measure" on $\hat{G}$. Quite generally, if S is an arbitrary Borel space, one defines a *projection-valued measure* P on S to be an assignment of a projection operator P_E on a separable Hilbert space $\mathscr{H}(P)$ to each Borel set $E \subseteq S$ in such a fashion that the following conditions are satisfied:

(i) $P_E P_F = P_F P_E = P_{E \cap F}$ for all E and F.
(ii) $P_0 = 0$ and $P_S = I$.
(iii) Whenever E_1, E_2,... are mutually disjoint, then $P_{E_1 \cup E_2 \cup \cdots} = P_{E_1} + P_{E_2} + \cdots$.

The projection-valued measure P on $\hat{G}$ defined above is special in that there exists a countable set $(\{\chi_1\} \cup \{\chi_2\} \cup \cdots) = A$ such that $P_{\hat{G}-A} = 0$. One says that P has *countable support*. Conversely, it is easy to see that every projection-valued measure with countable support arises as indicated from a unique representation V of G which is a discrete direct sum of irreducibles. One has a natural one-to-one correspondence between all projection-valued measures on $\hat{G}$ with countable support and all discretely decomposable unitary representations of G. The spectral theorem asserts that this correspondence may be extended to one between *all* unitary representations V of G and all projection-valued measures P on $\hat{G}$. Notice that for any projection-valued measure P on $\hat{G}$ and each vector φ in $\mathscr{H}(P)$, the mapping $E \to (P_E(\varphi) \cdot \varphi)$ is a measure on $\hat{G}$. Moreover, for each x in G, $\chi \to \chi(x)$ is a continuous function on $\hat{G}$ which may be integrated with respect to his measure. By an obvious abuse of notation, we write this integral as $\int \chi(x)\, d(P_x(\varphi) \cdot \varphi)$. More generally, if φ and ψ are two vectors, $E \to (P_E(\varphi) \cdot \psi)$ is a finite linear combination of measures and we may form the integral $\int \chi(x)\, d(P_x(\varphi) \cdot \psi)$. When P has countable support $\{\chi_1\} \cup \{\chi_2\} \cup \cdots$, it is evident that

$\int \chi(x)\, d(P_x(\varphi) \cdot \psi) = \sum \chi_j(x)(P_{x_j}(\varphi) \cdot \psi) = (V_x(\varphi) \cdot \psi)$, where V is the associated unitary representation. In the general case, it is easy to show that for each x in G, there exists a unique bounded linear operator V_x such that $(V_x(\varphi) \cdot \psi) = \int \chi(x)\, d(P_\chi(\varphi) \cdot \psi)$ for all φ and ψ. Moreover, it is only slightly more difficult to show that V_x is always unitary and that $x \to V_x$ is a unitary representation of V. The more difficult part of the spectral theorem (proved independently in the mid 1940's by Ambrose, Neumark, and Godement) states that every unitary representation V of G arises in this way from some projection-valued measure P on $\hat{G}$. It is easy to see that V determines P uniquely and that two V's are equivalent if and only if the corresponding P's are equivalent. Of course, one defines P^1 and P^2 to be equivalent if there exists a unitary map U from $\mathscr{H}(P^1)$ to $\mathscr{H}(P^2)$ such that $UP_E{}^1U^{-1} = P_E{}^2$ for all E.

It follows from the spectral theorem as just formulated that the problem of classifying unitary representations of G to within equivalence reduces completely to the corresponding problem for projection-valued measures on $\hat{G}$. On the other hand, the notion of projection-valued measure involves only the Borel structure on $\hat{G}$ and is completely independent of its group structure. Since $\hat{G}_1$ and $\hat{G}_2$ are isomorphic as Borel spaces except when they have different cardinal numbers, one can deduce the classification of projection-valued measures for any non-countable $\hat{G}_1$ from that for the real line and, hence, from the classical theory for self-adjoint operators. Moreover, one can forget about groups and describe the classification for an abstract Borel space M. For definiteness, we shall assume that M is standard, although a weaker restriction would suffice. For each σ finite measure μ on M, we define a projection-valued measure P^μ on M as follows. Its space is $\mathscr{L}^2(M, \mu)$ and for each f in $\mathscr{L}^2(M, \mu)$ and each Borel set E in M, $P_E{}^\mu(f) = \varphi_E f$, where φ_E is the function on M which is one for m in E and zero for m in $M - E$. Moreover, if g is any element of $\mathscr{L}^2(M, \mu)$ which never vanishes, then g is easily seen to be a *cyclic vector* for P^μ in the sense that the finite linear combinations of the vectors $P_E{}^\mu(g)$ are dense in $\mathscr{L}^2(M, \mu) = \mathscr{H}(P^\mu)$. In particular, if $\mu(M) < \infty$, we may choose $g \equiv 1$ and then $\mu(E) = (P_E{}^\mu(g) \circ g)$. Conversely, if $E \to P_E$ is any projection-valued measure on M which admits a cyclic vector g, then $E \to (P_E(g) \circ g) = \nu(E)$ is a measure on M such that $\nu(M) < \infty$ and one verifies easily that the mapping $c_1\varphi_{E_1} + \cdots c_n\varphi_{E_m} \to c_1P_{E_1}(g) + \cdots c_nP_{E_n}(g)$ is norm-preserving and linear from a dense subspace of $\mathscr{L}^2(M, \nu)$ to $\mathscr{H}(P)$ and that the unique continuous extension of this mapping to all of $\mathscr{L}^2(M, \nu)$ sets up an equivalence between P and P^μ. Since any finite

measure ν[1] on M with the same null sets as μ is of the form $E \to \int_E |g|^2 d\mu$ for some g in $\mathscr{L}^2(M, \nu)$ which is never zero and since $P_E{}^\mu = 0$ if and only if $\mu(E) = 0$, one deduces that: *A projection-valued measure P on M is equivalent to one of the form P^μ if and only if it admits a cyclic vector and P^μ and $P^{\mu'}$ are equivalent if and only if μ and μ' lie in the same measure class.* In other words, the mapping $\mu \to P^\mu$ sets up a one-to-one correspondence between measure classes on M and equivalence classes of those projection-valued measures P on M which possess a cyclic vector. In order to understand the significance of this result, we need to see what having a cyclic vector means from some other point of view. To this end, consider the special case in which P has countable support so that $P_{\{m\}_j} \neq 0$ for some countable subset $m_1, m_2, \ldots$ of M and $P_E = \sum_{m_j \in E} P_{\{m_j\}}$. It is straightforward to show that P has a cyclic vector if and only if $P_{\{m_j\}}$ has a *one-dimensional range* for all i. When $M = \hat{G}$, this is the same theory as saying that each irreducible constituent in the associated group representation occurs only once, that is, that there are *no multiplicities*. Accordingly, we define a general P to be *multiplicity free* if it has a cyclic vector ω so that the P^μ are just the multiplicity-free projection-valued measures. One can define "direct sums" of projection-valued measures by analogy with direct sums of group representations $(P^1 \oplus P^2 \cdots \oplus P^n \oplus \cdots)_E[f_1, f_2, \ldots]$ being $P_E{}^1(f_1), P_E{}^2(f_2), \ldots$, and one sees at once that the direct sum kP of k replicas of the same multiplicity-free countably supported projection-valued measure P is such that $(kP)_{\{m\}}$ has a k-dimensional range whenever $P_{\{m\}} \neq 0$. In this sense, kP is uniformly of multiplicity k. More generally, if P^1 and P^2 are arbitrary multiplicity-free projection-valued measures on M and k_1 and k_2 are positive integers or ∞, then one can show (but the proof is not obvious) that k_1P^1 and k_2P^2 are equivalent if and only if $k_1 = k_2$ and P^1 and P^2 are equivalent. Thus it makes sense to say that a projection-valued measure has uniform multiplicity k if it is equivalent to kP^0, where P^0 is multiplicity free and one sees that for each k, the projection-valued measures which have uniform multiplicity k correspond one-to-one to the measure classes in M. Thinking of the meaning of uniform multiplicity in the case in which P has countable support, it is natural to conjecture that every projection-valued measure is equivalent to direct sum of projection-valued measures each having uniform multiplicity k and that this decomposition is essentially unique when "overlapping" is avoided. This is in fact so. Let us define the projection-valued measures $P^1, P^2, \ldots$ to be *disjoint* if there exist disjoint Borel sets $E^1, E^2, \ldots$ such that $P^j_{M-E_j} = 0$. Then it can be proved that for every projection-

valued measure P on M, there exist mutually disjoint projection-valued measures P^∞, P^1, P^2,... such that P is equivalent to $P^\infty \oplus P^1 \oplus P^2 \oplus \cdots$ and P^j has uniform multiplicity j. Moreover, each P^j is uniquely determined up to equivalence. Of course, some terms may be missing. If C_j is the measure class whose null sets are those of P^j, then the C_j are mutually disjoint measure classes which are uniquely determined by the equivalence class of P and, in turn, uniquely determine this equivalence class. Obviously, every sequence C_∞, C_1, C_2,... of disjoint measure classes occurs.

Returning to the special case in which $M = \hat{G}$ for some separable locally compact commutative group, it follows from the spectral theorem and the foregoing analysis that the unitary representations of G are completely classified by the sequences C_∞, C_1, C_2,... of disjoint measure classes in $\hat{G}$. One verifies easily that the unitary representation of G, whose projection-valued measure is P^μ, is equivalent to the representation V^μ in $\mathscr{L}^2(\hat{G}, \mu)$ defined by the equation $V_x{}^\mu(f)(\chi) = \chi(x)f(\chi)$. Hence the representation associated with

$$C_\infty, C_1, C_2, \ldots \text{ is } \infty V^{\mu_\infty} \oplus V^{\mu_1} \oplus 2V^{\mu_2} \oplus \cdots,$$

where μ_j is any member of C_j. In an obvious sense, V^μ is the "direct integral with respect to μ" of the irreducible representations of G defined by the members of $\hat{G}$. Hence, our theorem implies in particular a canonical "direct integral decomposition" for every unitary representation V of G. If V is the unitary representation associated with an action of G, we shall call C_j the *multiplicity j component of the spectrum of the action.* A measure class which is supported by a countable set is completely determined by the minimal countable set which supports it so that such measure classes may be identified with subsets. With this convention, our definition of spectrum is consistent with that given earlier.

Recall the construction of the chief class of examples given so far of an ergodic action without pure point spectrum. Given a countable commutative group G, let A be a separable compact commutative group and form the compact group A^G of all functions from G to A. For each $f \in A^G$ and each x in G, let $f_x(y) = f(yx^{-1})$. Then $f \to f_x$ is an automorphism of A^G, and A^G thus becomes a G space. Haar measure μ in A^G is invariant and the associated action of G on A^G is ergodic whenever G is not finite. To compute the spectrum, we use the fact that the characters of A^G form a complete orthonormal set in $\mathscr{L}^2(A^G, \mu)$ and that the most

general character on A^G is defined by a function from G to $\hat{A}$ which is the identity for all but a finite number of members of G. One sees easily that, except for the identity character, the transforms of a character by the elements of G are all distinct. It follows that the representation V of G associated with the given action is the direct sum of the one-dimensional identity and of countably many replicas of the regular representation of G, the latter being the representation in $\mathscr{L}^2(G, \nu)$ defined by setting $W_x f(y) = f(yx)$, where ν in Haar measure in G. On the other hand, the duality theorem tells us that the members of G form a complete orthonormal set in $\mathscr{L}^2(\hat{G}, \hat{\nu})$, where $\hat{\nu}$ is Haar measure in $\hat{G}$, and it follows that the regular representation of G is equivalent to the representation $V^{\hat{\nu}}$ defined above. In other words, the regular representation is multiplicity free and has the measure class of Haar measure as its defining measure class. It now follows that the spectrum of the ergodic action of G on A^G has all components zero except the multiplicity-one component and the multiplicity-∞ component, the former being the measure class of a discrete measure supported by the identity and the latter being the measure class of Haar measure. In particular, the spectrum is independent of the choice of A. Since the one-dimensional identity is a component of V whenever there is a finite invariant measure, one often abbreviates and says simply that the spectrum is a countable Haar spectrum. In the special case in which $G = Z$, Haar measure is Lebesgue measure in the circle and one speaks of the spectrum as being countable Lebesgue.

If G_1 and G_2 are separable locally compact commutative groups and S_1, μ_1 and S_2, μ_2 define actions of G_1 and G_2, respectively, then one has a well-defined product action of $G_1 \times G_2$ on $S_1 \times S_2$, $\mu_1 \times \mu_2$, and it is natural to attempt to relate the spectrum of the product action to the spectra of the factors. One shows easily that the multiplicity j component of the product action is the measure class of $\sum_d \nu_d^1 \times \nu_{j/d}^2$, where d varies over the divisors of j and ν_l^k is a measure in the class defining the multiplicity l component of the action of G_k. When $j = \infty$, this sum must be replaced by $\sum_{n=1}^{\infty} \nu_n^1 \times \nu_\infty^2 + \sum_{n=1}^{\infty} \nu_\infty^1 \times \nu_n^2$. In the special case in which $G_1 = G_2 = G$, one can restrict the action of $G \times G$ on $S_1 \times S_2$ to the diagonal $\tilde{G}$ consisting of all x, y with $x = y$ and so obtain a new action of G. The spectrum of this action can be computed from that of the component actions in a straightforward way using the result on the action of $G \times G$ just described. In particular, a character χ of G is contained in the discrete part of the spectrum if and only if $\chi = \chi_1\chi_2$, where χ_1 and χ_2 are in the discrete part of the two component actions on S_1 and S_2, respectively. Moreover, it is contained

with a multiplicity equal to the number of different ways in which χ can be factored. It is a corollary that the product of two ergodic actions of G is also ergodic if and only if no χ except the identity is in the discrete part of the spectra of both actions. If one action is ergodic with pure point spectrum and the other is ergodic with countable Lebesgue spectrum, one sees easily that the product action is countable Lebesgue superposed upon a discrete spectrum equal to that of the component with a pure point spectrum.

Let G be the additive group of the integers or of the real line, and let μ be an invariant probability measure in the standard Borel G space S. Then, as we have seen, every real-valued member f of $\mathscr{L}^2(S, \mu)$ defines a stationary stochastic process. Let V be the unitary group representation associated with the action of G on S, μ, and let P^V be the projection-valued measure on $\hat{G}$ associated with V by the spectral theorem. Then $E \to (P_E{}^V(f) \cdot (f))$ is a finite measure ω on $\hat{G}$ absolutely continuous with respect to the spectrum of the action. This measure is called the *spectrum of the process*. It is not determined until one knows f as well as the action but is an actual measure instead of a measure class. As will be seen below, it is an important invariant of the process.

The Fourier transform of the spectrum is

$$\int \chi(x)\, d\omega(\chi) = \int \chi(x)\, d(P_x{}^V(f) \cdot (f),$$

which, by the spectral theorem, is equal to

$$(V_x(f) \cdot (f)) = \int f([s]x) f(s)\, d\mu(s).$$

In other words, the spectrum of the process is just the measure associated to the positive definite function $x \to (V_x(f) \cdot (f))$ by the Bochner Herglotz theorem. This positive definite function on G is called the *covariance function* and can be computed from the process without decomposing V. Actually, when the action is ergodic, it can be computed easily from the complete past of almost any sample function. Consider first the case in which G is the additive group of the integers and apply the ergodic theorem to the function $f([s]n)f(s)$ and the action $s, n \to [s](-n)$. We find that

$$\int f([s]n) f(s)\, ds = V_n(f) \cdot f$$

$$= \lim_{k \to \infty} \frac{f([s](n-1))f([s](-1)) + f([s](n-2))f([s][-2] + \cdots + f([s](n-k))f([s][-k])}{k},$$

where the limit exists for almost all s. In other words, for almost all sample sequences (sample functions) $n \to f([s]n) = f_s(n)$, we find that

$$\frac{f_s(n-1)f_s(-1) + f_s(n-2)f_s(-2) + \cdots + f_s(n-k)f_s(-k)}{k}$$

has a limit as $k \to \infty$ for all n and that the resulting function of n is independent of the sample function chosen and equal to the desired covariance function. The limit of

$$\frac{f_s(n-1)f_s(-1) + f_s(n-2)f_s(-2) \cdots f_s(n-k)f_s(-k)}{k}$$

is called the autocorrelation sequence of the sample sequence. One can clearly drop any finite number of terms without changing the value of the limit so that one really needs to know $f_s(m)$ only for $m < 0$; that is, one needs only the complete past of almost any sample sequence. An analogous result holds when G is the additive group of the real line, but there are additional technical complications in the proof.

7. The Harmonic Analysis of Sample Functions

In this section we return to the applications of ergodic theory to probability theory and relate the spectral analysis of the group representation V associated with an ergodic action to the behavior of the sample functions of the stationary stochastic processes built upon the action. We shall concentrate attention upon continuous stochastic processes so that the group G will be the additive group of the real line. The situation is especially transparent in the special case in which the action has a pure point spectrum so that $\mathscr{L}^2(S, \mu)$ admits an orthonormal basis $\varphi_1, \varphi_2, \ldots$ with $V_x(\varphi_j) = e^{ix\lambda_j}\varphi_j$, where the λ_j are real numbers. Then if the function f defining the process is in $\mathscr{L}^2(S, \mu)$, we have $f(s) = \sum_{j=1}^{\infty} c_j\varphi_j(s)$, where $c_j = \int f(s)\,\bar{\varphi}_j(s)\,d\mu(s)$ and convergence is in the Hilbert space norm. For f in a suitable dense subspace of $\mathscr{L}^2(S, \mu)$, we have $\sum_{j=1}^{\infty} |c_j| < \infty$ so that the series $\sum_{j=1}^{\infty} c_j\varphi_j(s)$ converges uniformly to $f(s)$. In this case, the sample function $f_s(x) = f([s]x) = \sum_{j=1}^{\infty} c_j\varphi_j[[s]x] = \sum_{j=1}^{\infty} c_j e^{i\lambda_j x}\varphi_j(s) = \sum_{j=1}^{\infty} c_j\varphi_j(s)\,e^{i\lambda_j x}$ where the series converges uniformly. In other words, almost every sample function can be expanded in an infinite series of the form $\sum_{j=1}^{\infty} c_j^1 e^{i\lambda_j x}$ with one term for each irreducible component of V and with $|c_j^1| = |c_j|$ and independent of the particular

sample function chosen. In particular, almost all sample functions are "almost periodic" in the sense of Bohr. We recall in this connection that, generalizing the original definition of Bohr, we can define almost periodicity for continuous functions on any topological group and that when the group is commutative, being almost periodic is equivalent to being a uniform limit of finite linear combinations of characters. Moreover, if A is the vector space of all almost periodic functions on the commutative topological group G, then there exists a unique continuous linear functional M on A which takes the function identically 1 into 1 and is translation invariant. One refers to $M(f)$ as the mean value of f. When G is the additive group of the real line,

$$M(f) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(t)\, dt$$

and when G is the additive group of the integers

$$M(f) = \lim_{n\to\infty} \frac{f(1) + f(2) + \cdots + f(n)}{n}.$$

For each character χ of G and each almost-periodic function f, $\bar{\chi}f$ is also almost periodic and $M(\bar{\chi}f)$ is zero except for a countable subset $\chi_1, \chi_2, \ldots$. The formal series $\sum M(\bar{\chi}_n f)\, \chi_n(x)$ is called the Bohr–Fourier series for f and converges to f when f is sufficiently regular. In any case, f can be uniformly approximated arbitrarily closely by finite linear combinations of $\chi_1, \chi_2, \ldots$. It is interesting to compare the theory of almost-periodic functions just described with a certain property of the sample functions of our process deducible from the ergodic theorem. Indeed the coefficient $c_j = \int f(s)\, \bar{\varphi}_j(s)\, d\mu(s)$, and by the ergodic theorem, we may compute this by choosing almost any s and setting

$$c_j = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(sx)\, \bar{\varphi}_j(sx)\, dx = \lim_{T\to\infty} \frac{1}{T} \int_0^T f_s(x)\, \bar{\varphi}_j(s)\, e^{-i\lambda_j x}\, dx.$$

In other words, the almost-periodic function mean of $f_s e^{-i\lambda_j x}$ exists and equals $c_j \varphi_j(s)$. Thus, even when $\sum_{j=1}^{\infty} | c_j | = \infty$, almost all sample functions have a formal Bohr–Fourier series equal to $\sum_{j=1}^{\infty} c_j \varphi_j(s)\, e^{i\lambda_j s}$. In fact, one can show that they are almost periodic in a generalized sense introduced by Besicovitch and related to the Bohr almost-periodic functions just as the square summable periodic functions are related to the continuous periodic functions. Note that any almost periodic function, whether Bohn or Besicovitch, is zero everywhere if it is zero on a half line. Thus, when the sample functions of a process are almost

periodic, the future is completely determined by the past and no real randomness is involved. In other words, the processes whose underlying actions have pure point spectra are of rather limited interest in probability theory.

In the general case, however, the unitary representation V may have irreducible subrepresentations even if the spectrum is not entirely discrete, and the analysis given above may be adapted to analyze the process into a sum of a deterministic and a more random process. Let $V = V^d \oplus V^c$, where V^d is a discrete sum of irreducibles and V^c has no irreducible subrepresentations. Then the function f defining the process may be written uniquely in the form $f = f_d + f_c$, where $f_d \in \mathscr{H}(V^d)$ and $f_c \in \mathscr{H}(V^c)$. Of course, f_d and f_c define processes themselves, and each sample function f_s will be the sum $(f_d)_s + (f_c)_s$ of corresponding sample functions of these two auxiliary processes. Just as in the case of a purely discrete spectrum, the sample function $(f_d)_s$ may be expanded into a Bohr–Fourier series whose coefficients may be computed from the past of f_s in the same way. It is possible to use f_s instead of $(f_d)_s$ because the contribution of $(f_c)_s$ is always zero. The resulting analysis of the sample function into a sum of periodic components and a function having no such components is called "determining the hidden periodicities" or "periodogram analysis". If $(f_d)_s$ has the Bohr–Fourier series $\sum_{j=1}^{\infty} c_j \varphi_j(s)\, e^{i\lambda_j s}$, then the measure supported by the countable set $\{\lambda_1\} \cup \{\lambda_2\} \cup \cdots$ and assigning $\{\lambda_j\}$ the measure $|\, c_j \,|^2$ is precisely the atomic part of the spectrum of the process as defined in the last section.

In 1930, Norbert Wiener published an important and influential memoir entitled "Generalized Harmonic Analysis" whose explicit aim was to find a generalization of periodogram analysis which would deal effectively with the continuous components $(f_c)_s$ of sample functions and the nonatomic part of what we have called the spectrum of the process. Although what Wiener did is best understood by thinking of the functions concerned as the sample functions of a stationary stochastic process, this was not the point of view be took at the time. Indeed, the theory of continuous parameter stochastic processes had yet to be developed and the ergodic theorem had not yet been proved. Wiener's paper is presented entirely from the point of view of the harmonic analysis of a single function on the line. Wiener was motivated by the needs of electrical engineers working on telephones, radios, and other communication devices. They concerned themselves with rapidly fluctuating currents and voltages and the effects on these of being passed through various kinds of circuits. Like so many problems in mathematics and

physics, this problem of the engineers was one for which Fourier analysis was the appropriate tool. However, the functions with which they had to deal were neither periodic nor rapidly decreasing at ∞, so that neither the theory of Fourier series not the theory of Fourier transforms was an adequate tool. The applied mathematician Oliver Heaviside had filled the breach after a fashion with his celebrated "operational calculus" which gave usable answers to many practical problems but involved a large amount of rather mysterious heuristic reasoning. Wiener had set himself the task of rigorizing and understanding the Heaviside rules and thus was led to the generalized Fourier analysis of the memoir alluded to above. On the other hand, Wiener was well aware that there were connections of his work with the mathematics of randomness. In fact, he obtained examples of functions which had "continuous spectra" by using his celebrated theory of "Brownian motion". This theory (which will be described in the next section) was, in fact, an anticipation of an important part of the theory of continuous parameter stochastic processes, and Wiener showed in effect that almost all sample functions of the corresponding process had the continuous spectra he was seeking. Moreover, the second volume of Wiener's autobiography contains the following statement (on page 79): "In other words I already began to detect a statistical element in the theory of the continuous spectrum, and through that, in communication theory. Now, almost thirty years later, communication theory is thoroughly statistical and this can be traced directly back to my work of that time." What Wiener means by this last sentence is that the fluctuating voltages and currents with which the communication engineer deals can fruitfully and meaningfully be regarded as sample functions of a stationary stochastic process and that it is better to work with the whole ensemble of possible sample functions than to deal with any one of them. This point will be elaborated further in a later section.

A key role in Wiener's theory is played by the concept of the "spectrum of a function". This is a measure on the real line which coincides with the spectrum of the underlying stochastic process whenever the function is a sample function, and one can motivate Wiener's definition by looking at how one would compute the spectrum from almost any sample function. If γ is the spectrum, then, as we have seen, the Fourier transform φ of γ can be obtained from almost any sample function by means of the formula

$$\varphi(x) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(x-t) f(-t)\, dt.$$

If γ is absolutely continuous with respect to Lebesgue measure so that $\gamma(E) = \int_E \rho(x)\,dx$ for some summable positive density ρ, then φ is just the ordinary Fourier transform of ρ and $\rho(x) = 1/2\pi \int_{-\infty}^{\infty} \varphi(t)\, e^{ixt}\,dt$, at least when ρ is bounded. Of course, the integral may have to be interpreted as a limit in the mean as $A \to \infty$ of the function $1/2\pi \int_{-A}^{A} \varphi(t)\, e^{ixt}\,dt$. Now $\gamma([0, x]) = \int_0^x \rho(y)\,dy = 1/2\pi \int_0^x \int_{-\infty}^{\infty} \varphi(t)\, e^{ity}\,dt\,dy$. If the interchange of integrals can be justified, then this may be rewritten as

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) \int_0^x e^{iyt}\,dy\,dt,$$

and, evaluating the inner integral, one has the formula:

$$\gamma([0, x]) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{t} \varphi(t)\,dt.$$

Wiener considers the class of all complex-valued measurable functions for which $\varphi(x) = \lim_{T\to\infty} 1/T \int_{-T}^{T} f(x + t)\bar{f}(t)\,dt$ exists for all x and shows by a long argument that the function S defined by

$$S(u) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{iux} - 1}{u} \varphi(x)\,dx$$

exists for all u. He calls it the *integrated periodogram* of f. It is a monotone function of u and the measure it defines coincides with the spectrum of the process when f is almost any sample function of a stationary stochastic process.

Let us write down the purely formal expression

$$g(x) = \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{e^{ity} - 1}{t} f(t)\,dt,$$

where f is as above. Differentiating formally, we get

$$g'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} f(t)\,dt.$$

Moreover, changing the integrand from $(e^{itx} - 1)/t$ to e^{itx}/t on the interval $-1 \leqslant t \leqslant 1$ will change the original expression by $1/2\pi i \int_{-1}^{1} f(t)/t\,dt$ which does not depend on x. Hence, formally speaking, if we define

$$g(x) = \frac{1}{2\pi i} \int_{-\infty}^{-1} \frac{e^{itx} - 1}{t} f(t)\,dt + \frac{1}{2\pi i} \int_{-1}^{1} \frac{e^{itx}}{t} f(t)\,dt + \frac{1}{2\pi i} \int_{1}^{\infty} \frac{e^{itx} - 1}{t} f(t)\,dt,$$

we get a function whose derivative is the (nonexistent) Fourier transform of f. One of the easy results of Wiener's generalized harmonic analysis is that the second expression for g actually makes sense when the infinite integrals are interpreted as limits in the mean. In other words, there is an actual function g whose nonexistent derivative may be looked upon as the nonexistent Fourier transform of f. In modern terminology, of course, g' exists as a distribution and this distribution is the Fourier transform of f.

If g' existed and were the actual Fourier transform of f, then it would follow from the Plancherel formula that $\int_{-\infty}^{\infty} |g'(x)|^2\,dx = \int_{-\infty}^{\infty} |f(x)|^2\,dx$. As a substitute for this, Wiener was able to prove that

$$\lim_{\epsilon\to 0} \frac{1}{2\epsilon}\int_{-\infty}^{\infty} |g(u+\epsilon) - g(u-\epsilon)|^2\,du = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} |f(t)|^2\,dt,$$

a result of which he was so proud that he had it printed on the cover of the second volume of his autobiography. It turned out to be quite easy to reduce the problem to showing that

$$\lim_{\epsilon\to 0} \frac{2}{\pi\epsilon}\int_{0}^{\infty} F(t)\frac{\sin^2 \epsilon t}{t^2}\,dt = \lim_{T\to\infty} \frac{1}{T}\int_{0}^{T} F(t)\,dt,$$

where $F(t) = |f(t)|^2 + |f(-t)|^2$, but Wiener had considerable difficulty in doing the latter. When it was pointed out to him that he needed a so-called "Tauberian theorem", he proceeded to prove a very general such theorem which not only enabled him to complete the proof of the above but included as corollaries all previously known Tauberian theorems.

The only obvious examples of functions f which satisfy Wiener's condition that

$$\varphi(x) \lim_{T\to\infty} \frac{1}{T}\int_{0}^{T} f(x-t)f(t)\,dt$$

should exist for all x are functions which are almost periodic in some sense or else such that $\varphi(x) \equiv 0$. On the other hand, for any stationary stochastic process with finite variance, it is clear that almost every sample function satisfies the Wiener condition—at least for almost all x. Moreover, the spectrum of the sample functions coincides with the spectrum of the process. These sample functions provide a rich source of examples; and others are difficult to come by and almost unknown. From a practical point of view, Wiener's theory is a theory of sample

functions. To get an idea of what a sample function might look like, consider a function f taking on only the values 1 and -1 and constant on each interval $n < x \leqslant n - 1$, where n is an integer. To decide whether f takes on the value 1 or -1 in each interval of constancy, toss a coin countably many times and label the tosses with the integers. Choose 1 or -1 for $n < x \leqslant n + 1$ according as the nth toss is heads or tails. As shown by Wiener, one thus obtains a function with a nontrivial continuous spectrum for almost every possible sequence of tosses. In fact, the spectrum is absolutely continuous with respect to Lebesgue measure, and its Radin–Nikodym derivative can be explicitly computed.

We have seen the meaning of Wiener's spectrum from the stochastic process point of view. What about the actual harmonic analysis of the sample functions? The spectrum tells us "what harmonic constituents are there" and with what "weights" but not how to decompose the functions. The answer lies in looking at the spectral theorem for the underlying group representation V and the projection-valued measure on the line $E \to P_E$ which it assigns to V. When there is a discrete spectrum $\{\lambda_1, \lambda_2, ...\}$, the projections $P_{\{\lambda_j\}}$ are one dimensional and the functions φ_j generate their ranges. In the general case, the P_E are the natural substitute for the φ_j. Let $f \in \mathscr{L}^2(S, \mu)$ define a stationary stochastic process with sample functions $f_s(x) = f([s]x)$, and let $\hat{G} = E_1 \cup E_2 \cdots \cup E_n$, where the E_j are disjoint Borel sets. If $g_j = P_{E_j}(f)$, then for almost all s, $f_s(x) = \sum_{j=1}^{n} (g_j)_s(x)$, and each g_j defines a stationary stochastic process whose spectrum is supported by the set E_j. In this manner, the projection-valued measure P may be used to define a sort of harmonic analysis of the sample functions which reduces to that provided by the Bohr–Fourier analysis when the sample functions are almost periodic. In particular, it makes sense to ask for the component of a sample function whose spectrum lies in a particular Borel set.

More generally, let ψ be any real-valued Borel function defined on $\hat{G}$, and let $P_F^{\psi} = P_{\psi^{-1}(F)}$ for each Borel subset of the real line. Then P^{ψ} is a projection-valued measure on the real line and, by the spectral theorem for self-adjoint operators, is canonically associated with a self-adjoint operator H_{ψ}. Whenever f is in the domain of H_{ψ}, the sample functions of the process defined by $H_{\psi}(f)$ may be thought of as the result of multiplying the harmonic constituents of the sample functions f_s by the values of ψ. If f_s had a Fourier transform, then $(H_{\psi}(f))_s$ would be the result of multiplying this Fourier transform by ψ and then taking the inverse Fourier transform. In the particular case in which ψ has a summable Fourier transform $\hat{\psi}$, one verifies, as might be expected, that

$(H_\psi(f))_s$ may be obtained directly from f_s by forming the convolution of f_s and $\hat{\psi}$.

An important case in which ψ is not the Fourier transform of a summable function is that in which $\psi(x) \equiv x$. In this case, one verifies that H_ψ is the operator which takes $f(s)$ into $i(d/dx)f[(sx)]_{x=0}$ so that $(1/i)(H_\psi(f))_s$ is just the derivative of f_s whenever f is in the domain of H_ψ. It follows that when ψ is a polynomial P, then $(H_\psi f)_s$ is the differential operator $P(i(d/dx))$ applied to f. This suggests that, quite generally, one can interpret the operator $f_s \to (H_\psi f)_s$ discussed above as $\psi(i(d/dx))$ and thus assign a meaning to $\psi(i(d/dx))$ when ψ is an arbitrary Borel function. The problem in rigorizing the Heaviside operational calculus was precisely to make sense of expressions such as $\psi(i(d/dx))$ when the functions to be operated on were those arising in communication engineering. The above discussion provides a solution for the sample functions of stationary stochastic processes and hence for the functions arising in communication engineering.

The notion of the spectrum of a stationary stochastic process is of importance when there is a question of separating the sum of two processes back into their components. Let f and g be two members of $\mathscr{L}^1(S, \mu)$, and let $h = f + g$. The question is this: Given a sample function h_s of the process defined by h, can be find the sample functions f_s and g_s at least approximately? The answer depends upon the relationship between the spectra of the process defined by f and by g. Let P be the projection-valued measure associated with the unitary group representation V defined by the action. Then $(V_t(\varphi) \cdot \varphi) = \int e^{ixt}\, d(P_t(\varphi) \cdot \varphi)$ for all φ in $\mathscr{L}^2(S, \mu)$, so $E \to (P_E(f) \cdot f)$ is the spectrum of the process defined by f and $E \to (P_E(g) \cdot g)$ is that defined by g. If there exists a finite interval I such that $(P_I(f) \cdot f)$ is near to $(f \cdot f)$ and $(P_I(g) \cdot g)$ is close to zero, then $P_I(f + g)$ will be a good approximation to f, and for almost all s, $(P_I(f + g))_s$ will be a good approximation to f_s. Such an I will exist, of course, if and only if the two spectra have a "small overlap" in the sense that every interval is assigned a "small" value by one spectrum or the other.

When I does exist, its characteristic function may be approximated by a smooth function $\varphi_I{}^0$ having a summable Fourier transform ψ_I. Then $P_I(f + g)$ will be approximated by $\varphi_I{}^0(H)(f + g)$, and so f will be approximated by $\varphi_I{}^0(H)(f + g)$. The convolution of $h_s = (f + g)_s$ with ψ_I will then be an approximation to f_s, and we have a usable way of computing an approximation to f_s from a knowledge of $(f + g)_s$. One speaks of *filtering* g_s out of $(f + g)_s$. In communication engineering,

f_s is the message and g_s is a contaminating "noise". Evidently, one must try to send messages whose spectra are more or less disjoint from that of the expected noise.

Notice that multiplying $f(t)$ by $e^{i\lambda t}$ simply translates the spectrum of f by an amount λ. Thus, given sample functions f and g, one may always filter g out of $e^{i\lambda t}f + g$ by taking λ large enough. This principle is made use of in radio broadcasting and in transmitting several telephone messages simultaneously along the same wire. Instead of transmitting the desired messages $f_1, f_2, \ldots, f_n$ directly, one sends purely periodic "messages" $e^{i\lambda_1 t}, \ldots, e^{i\lambda_n t}$ and "modulates the amplitudes" by replacing them by $e^{i\lambda_1 t}f_1(t), \ldots, e^{i\lambda_n t}f_n(t)$. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are sufficiently far apart, then the spectra of these functions will have very little overlap and one may send the message $e^{i\lambda_1 t}f_1(t) + \cdots e^{i\lambda_n t}f_n(t)$ and separate the components at the receiving end. When one tunes a radio set, one is adjusting an electronic filter to filter out everything except the message of the form $e^{i\lambda_j t}f_j(t)$ for some fixed j.

The spectrum of a process also plays a central role in the linear prediction theory independently worked out by Wiener and A. N. Kolmogoroff in 1940 and 1941. The fundamental problem of prediction theory (not necessarily linear) is the following. Given the past of a sample function of a stationary stochastic process, what can one conclude (probabilistically speaking) about the future? As explained in Section 4, it is possible to reconstruct the whole process (more precisely, the relevant ergodic component) from the past of almost any sample function. If one is dealing with a discrete process (as will now be assumed for simplicity), this means that one can find the appropriate probability measure μ on the space S of all functions from G to the real or complex numbers. Let S^- denote the set of all "pasts"; that is, the set of all functions from the negative integers to the real complex numbers, and let S^+ denote the set of all "futures". There are natural Borel maps π and φ of S onto S^- and S^+, respectively, where $\pi(s)$ is the restriction of s to the negative integers and $\varphi(s)$ is its restriction to the nonnegative integers. Clearly, $\pi^{-1}(p)$ is the set of all sample functions with the same past p, and the fibering of S defined by the $\pi^{-1}(p)$ is associated (as explained in Section 3) with a corresponding fibering of the measure μ into measures μ_p concentrated in the fibers. For each Borel set E in S^+, let $\tilde{\mu}_p(E) = \mu_p(\varphi^{-1}(E))$. Then for almost all p, $\tilde{\mu}_p$ will be a well-defined probability measure in S^+ which tells one the probabilities of various futures *given that the past was* p. When the random variables of the process are independent, the $\tilde{\mu}_p$ will be independent of p and will

coincide with the image $\tilde{\mu}$ of μ in S^+. More generally, the $\tilde{\mu}_p$ will depend upon p, and $\tilde{\mu}$ will be a weighted average of all μ_p. Thus the $\tilde{\mu}_p$ will be less "spread out" than $\tilde{\mu}$, and knowing the past p will enable one to make more precise predictions than just knowing the process and hence $\tilde{\mu}$. At the other extreme from independence, it may happen that almost every $\tilde{\mu}_p$ is a point measure in S^+. When this happens, the future is uniquely determined when the complete past is known, and the process is said to be *deterministic*. We have already seen that any process whose underlying action has a pure point spectrum is *deterministic* in this sense and other examples will be found in the discussion to follow.

The linear prediction theory of Wiener and Kolmogoroff is not concerned with finding the $\tilde{\mu}_p$ but with finding the "best" *linear* prediction for $s(0)$, given $s(-1)$, $s(-2)$, $s(-3)$,... . More precisely, they sought to find complex numbers $c_{-1}, c_{-2}, \ldots$ such that predicting $s(0)$ to be $\sum_{j=1}^{\infty} c_{-j}s(-j)$ minimizes the square of the expected error. Let $f(s) = s(0)$ and let V be the unitary group representation defined by the action of G on S. The problem is then easily seen to be equivalent to the following. Find $c_{-1}, c_{-2}, \ldots$ so as to minimize $\| \sum_{j=1}^{\infty} c_{-j}V_{-j}(f) - f \|$, where $\| \ \|$ is the norm in $\mathscr{L}^2(S, \mu)$. Let M denote the closed linear span of $V_{-1}(f)$, $V_{-2}(f)$,... . There is clearly a fundamental dichotomy depending upon whether or not f is in M. If f is in M, then there is no limit to how accurately $s(0)$ may be predicted by suitable choices of $c_{-1}, c_{-2}, \ldots$, and one can show accordingly that the process must be deterministic in the sense defined above. If f is not in M, then f may be written uniquely in the form $g + g'$, where $g \in M$ and $g' \in M^{\perp}$, and g will be the closest possible approximation to f among elements in M. The problem becomes that of expressing g in the form $c_{-1}V_{-1}(f) + c_{-2}V_{-2}(f) + \cdots$.

It is obvious that which side of the dichotomy one is on and, in fact, the whole problem depends only on the unitary equivalence class of the pair f, V—indeed only on that of the pair f, V', where V' is the subrepresentation generated by f. But V' is multiplicity free and so defined by a measure class in $\hat{G}$ which is here the circle group. In fact, if γ is the spectrum of the process or, equivalently, the measure associated with the positive definite function $n \to (V_n(f) \cdot (f)) = (V_n'(f) \cdot (f))$ by the Bochner–Herglotz theorem, then γ is in the measure class and one can realize the Hilbert space as $\mathscr{L}^2(\hat{G}, \gamma)$ in such a manner that f becomes the function identically one. The problem then becomes that of approximating 1 by $c_{-1}e^{-i\theta} + c_{-2}e^{-2i\theta} + \cdots$ on the unit circle $\hat{G}$ with respect to the norm in $\mathscr{L}^2(\hat{G}, \gamma)$ and so depends only on knowing γ, the spectrum of the process.

This problem had been considered and solved by *G.* Szegö in 1920 in a different context, but this was not known by Wiener and Kolmogoroff who solved it again. (Actually, Wiener dealt with the continuous case.). Let $\gamma = \gamma_s + \gamma_a$, where γ_a is absolutely continuous with respect to Lebesgue measure with Radin–Nikodym derivative ρ; and γ_s is singular with respect to Lebesgue measure. Of course, ρ and γ_s determine γ and are uniquely determined by it. The solution to our problem can then be expressed rather elegantly as follows. The function 1 is in the closed linear span of $e^{-i\theta}, e^{-2i\theta},\ldots$ if and only if $\int \log \rho \, d\theta = -\infty$. If $\int \log \rho \, d\theta > -\infty$ and $\gamma_s = 0$, then the coefficients $c_{-1}, c_{-2},\ldots$ may be computed from ρ by the following algorithm: Apply the Poisson formula to $\log \rho$ to obtain a harmonic function in the unit disk with the values of $\log \rho$ as boundary values. Let h be an analytic function with this harmonic function as its real part. The function h is then unique up to an additive pure imaginary constant, and $e^{h/2} = \tilde{h}$ is uniquely determined up to a multiplicative constant of absolute value 1. Expand $1 - (\tilde{h}(0)/\tilde{h})$ in a power series about 0. The complex conjugates of the coefficients will be the desired c_{-j}. When $\int \log \rho \, d\theta > -\infty$ and $\gamma_s \neq 0$, one can write the process as a sum of two others, one of which has $\gamma = \gamma_s$ and so is deterministic and the other of which has $\gamma = \gamma_a$ and can be dealt with as just explained.

Observe that whenever the underlying action has a pure point spectrum, one has $\gamma = \gamma_s$ and so a deterministic process as has already been deduced from the properties of almost periodic functions. Actually having $\int \log \rho \, d\theta > -\infty$ implies more than being deterministic in the sense that $\tilde{\mu}_p$ is a point measure. It implies that arbitrarily good *linear* prediction schemes exist. Recall also that γ is computable from the autocorrelation function of the past of almost any sample function.

8. The Construction of Properly Ergodic Actions of Nondiscrete Groups other than those with Pure Point Spectra: Induced Actions, Homogeneous Chaos and Wiener Measure

Although the discussion in Section 7 alluded to the existence of ergodic actions of the real line having a continuous spectrum, no examples were actually produced. In fact, except for discrete groups, we have exhibited no examples of properly ergodic actions other than those with pure point spectra. The first example was produced by von Neumann using a general procedure for constructing ergodic actions

of the real line out of ergodic actions of the integers. Given an ergodic action of the integers on the standard Borel measure space S, μ, where μ is finite and invariant, let α denote the measure-preserving transformation $s \to [s]1$ and choose a positive real number λ. Let I_λ denote the interval $0 < x < \lambda$, and let G be the additive group of the real line. For each s, x in $S \times I_\lambda$ and each t in G, let $[s, x]t = s, x + t$ whenever $0 \leqslant x + t < \lambda$, and let $[s, x]t = \alpha(s), \ x + t - \lambda$ whenever $\lambda \leqslant x + t < 2\lambda$. More generally, let n_t be the unique integer such that $n_t\lambda \leqslant x + t < (n_t + 1)\lambda$, and let $[s, x]t = \alpha^{n_t}(s)$. One verifies that $S \times I_\lambda$ is a standard Borel G space and that $\mu \times \nu$ is invariant and ergodic, where ν is Lebesgue measure in I_λ. The resulting action of G is called the *flow of height* λ *built over the measure-preserving transformation* α. It will have pure point spectrum if and only if the integer action defined by α has a pure point spectrum. Thus, to obtain a properly ergodic action of G not having a pure point spectrum, it is only necessary to start with such an action of the integers. To compute the spectrum of the G action from the given action of the integers, first identify the integers with the subgroup Z_λ of G consisting of all integer multiples of λ and consider the subgroup $Z_\lambda^\perp$ of $\hat{G}$ consisting of all characters which are trivial on Z_λ. The quotient group $\hat{G}/Z_\lambda^\perp$ is canonically identifiable with the dual Z_λ, and the measure classes in $\hat{G}$ defining the spectrum of the G action are precisely the "liftings" to $\hat{G}$ of the measure classes in $\hat{G}/Z_\lambda^\perp$ defining the spectrum of the Z_λ action. In particular, if the spectrum of the Z_λ action is the identity plus countable Lebesgue as in the examples of Section 4, the spectrum of the G action will be the discrete set $Z_\lambda^\perp$ plus countable Lebesgue.

It is interesting to study the nature of the sample functions of the continuous stationary stochastic process defined by the action of G on $S \times I_\lambda$ and a certain kind of real-valued function f on $S \times I_\lambda$. Let $S = S_1 \cup S_2 \cdots \cup S_r$, where the S_j are disjoint Borel sets of positive measure, and let $f_1, f_2, \ldots, f_r$ be real-valued Borel functions on I_λ. Define $f(s, x)$ as $f_j(x)$ whenever s is in S_j. Then for each s, x, the function $t \to f([s, x]t)$ will be a translate of a function which coincides on each interval $n\lambda < t < (n + 1)\lambda$ (n an integer) with the function $t \to f_j(t - n\lambda)$ for some value of j, and in a certain obvious sense, j will "depend randomly on n" in a manner depending on the choice of the S_j and the nature of the given action of Z_λ on S. In particular, if $\lambda = 1$, $r = 2$, and f_1 and f_2 are ± 1, respectively, then the Z_λ action, S_1, and S_2 may be chosen so that all sample functions are translates of the "coin-toss functions" mentioned toward the end of Section 7.

In order to obtain an example of a properly ergodic action of the real line with a finite invariant measure and *no point spectrum* (except for the identity), von Neumann introduced a generalization of the above construction in which the constant λ is replaced by a positive real-valued Borel function on S and $S \times I_\lambda$ by the set of all pairs s, x with $0 \leqslant x < \lambda(s)$. One speaks of the *flow built under the function* λ. Von Neumann was able to show that λ could be chosen so that there is no (nontrivial) point spectrum. Ten years later, W. Ambrose proved that every properly ergodic action of the real line having a finite invariant measure is isomorphic to a flow built under a function.

If one takes the "virtual group" point of view alluded to in Section 3, one is led more or less immediately to a far-reaching generalization of the concept of the flow of height λ built under a measure-preserving transformation and somewhat less immediately to a generalization of the concept of a flow built under a function. Since a closed subgroup of a closed subgroup of a group is itself a closed subgroup of that same group, it follows that a transitive action of a closed subgroup is canonically associated with a transitive action of the whole group. The analogy between transitive and properly ergodic actions now suggests the existence of an ergodic action of a separable locally compact group G canonically associated with every ergodic action of every closed subgroup H of G. This suggestion is borne out by the facts but before describing the relevant construction, it is necessary to introduce an auxiliary notion for which we shall have further use later on. This is the notion of a "quotient action" of a given ergodic action. Let S be a standard Borel G space, and let C be an invariant measure class in S. Let there be given an equivalence relation in S, let $\tilde{S}$ denote the space of all equivalence classes, and let Ψ denote the canonical map of S on $\tilde{S}$. One says that this equivalence relation is invariant if G maps each equivalence class onto another so that setting $[\Psi(s)]x = [\Psi(sx)]$ unambiguously defines an action of G on $\tilde{S}$. One converts $\tilde{S}$ into a Borel space by defining its Borel sets to be the sets E such that $\Psi^{-1}(E)$ is a Borel subset of S and defines $\tilde{C}$ to be the unique measure class in $\tilde{S}$ which contains the measures $E \to \mu(\Psi^{-1}(E))$, where μ is in C. If $\tilde{S}$ is standard or can be made so by discarding an invariant set of measure zero, one obtains a new action of G called the *quotient action defined by the equivalence relation*, and one says that the equivalence relation is *measurable* as well as invariant. It is clear that a quotient action is ergodic whenever this is true of the original action. On the other hand, a quotient action can be transitive even when the original action is properly ergodic.

Now let H be a closed subgroup of the separable locally compact group G, and let S be a standard Borel H space equipped with an ergodic invariant measure class C. Form $S \times G$ and convert it into an $H \times G$ space by defining $(s, x)(h, y)$ to be $[s]h, y^{-1}xh$. If C_G is the measure class of Haar measure, then $C \times C_G$ will be an ergodic $H \times G$ invariant measure class in $S \times G$. Let us define two elements of $S \times G$ to be equivalent if there exists an element of $H \times e$ carrying one into the other. One verifies that this equivalence relation is invariant and measurable under the action of $e \times G$ on $S \times G$ and that the quotient action is ergodic. This is the way in which one associates an ergodic action of G with every ergodic action of H. It is called the ergodic action of G *induced* by the given action of H. In the special case in which the H action is transitive so that S may be identified with H/K for some closed subgroup K of H, it is easy to see that the induced action of G is also transitive and isomorphic to the natural action of G on G/K. In the special case in which G is the additive group of the real line, the most general possible H other than $\{0\}$ and G itself is the group Z_λ of all integer multiples of some positive real number λ. Given an ergodic action of Z_λ with a finite invariant measure, let α be the measure-preserving transformation associated with λ, the generator of Z_λ. Then the action of G induced by the given ergodic action of Z_λ is easily shown to be isomorphic to the flow of height λ built over α. Using the technique of inducing from discrete subgroups and known ergodic actions of the latter, one can exhibit a considerable supply of properly ergodic actions for most noncompact separable locally compact groups.

The word "induced" is used in describing the above construction to emphasize the fact that it is closely analogous to a construction in the theory of group representations. If L is a unitary representation of the closed subgroup H of the separable locally compact group G, then using L one can define a unitary representation U^L of G called the *representation of G induced by L*. In the special case in which the coset space G/H admits an invariant measure μ, the definition of U^L is quite simple. One considers the set $\mathscr{F}$ of all Borel functions f from G to $\mathscr{H}(L)$ which satisfy the identity $f(hx) = L_h f(x)$ for all h in H and all x in g. For each such function, $(f(x) \cdot f(x))$ is evidently a constant on the right H cosets and hence may be looked upon as a function on the coset space G/H. Let $\mathscr{F}^0$ denote the subset of $\mathscr{F}$ consisting of all f in $\mathscr{F}$ for which $\int_{G/H} (f(x) \cdot f(x))\, d\mu(\bar{x}) < \infty$, where $\bar{x}$ denotes the canonical image of x in G/H. Identifying members of $\mathscr{F}^0$ which are equal almost everywhere, one obtains a Hilbert space with $\|f\|^2 = \int_{G/H} (f(x) \cdot f(x))\, d\mu(\bar{x})$ and the

obvious addition and scalar multiplication. This Hilbert space is $\mathscr{H}(U^L)$ and $U_x^L(f)(y) = f(yx)$. When G/H does not admit an invariant measure, the definition of U^L is a bit more complicated and will be omitted here. In any case, as might be expected, one can show that the unitary representation V defining the spectrum of an induced ergodic action of G is equivalent to U^L, where L is the unitary representation of H defining the spectrum of the inducing ergodic action of H. When G is commutative and L is one dimensional and defined by a character χ of H, the structure of U^L is easily described. Let $H^\perp$ denote the subgroup of $\hat{G}$ consisting of all characters of G which are identically 1 on H. Let χ^0 be any extension of χ to G, and let ν be the measure on $\hat{G}$ defined by setting $\nu(E) = \nu_{H^\perp}(E\chi^{-1} \cap H^\perp)$, where $\nu_{H^\perp}$ is Haar measure in $H^\perp$. Then U^L is the multiplicity-free representation of G defined by the measure class of ν. It may be looked upon as the "direct integral" of the one-dimensional representations defined by the characters in the coset $H^\perp\chi$ with respect to the translate by χ of Haar measure in $H^\perp$. Since inducing commutes with the taking of "direct integrals", one sees, in the commutative case at least, how to find the spectrum of an induced action when one knows the spectrum of the inducing action.

As mentioned above, the virtual group point of view also leads to a generalization of the von Neumann "flow built under a function" construction. However, the function $s \to \lambda(s)$ in the von Neumann construction has to be replaced by a function π from $S \times H$ to G satisfying the cocycle identity $\pi(s, h_1h_2) = \pi(s, h_1)\,\pi([s]\,h_1, h_2)$. Here, H and G are separable locally compact groups and S is a standard Borel H space equipped with an ergodic invariant measure class C. When H is the additive group of the integers, π is uniquely determined by the function $s \to \pi(s, 1)$, and given any Borel function φ from S to G, there is a unique π such that $\pi(s, 1) \equiv \varphi(s)$. In von Neumann's case, H is the additive group of the integers and G is the additive group of the real line so that the cocycles π correspond one-to-one to the real-valued functions on S. To construct an action of G, given π and the action of H, one forms $S \times G$ and converts it into an $H \times G$ space by defining $(s, x)(h, y)$ to be $[s]h,\ y^{-1}x\pi(s, h)$. The product of C and the measure class of Haar measure in G is an ergodic invariant measure class as above, and one defines the desired action of G as a quotient action of the restriction to $e \times G$. However, unlike the special case considered above, the equivalence relation need not be that defined by the H action. It will be only if that equivalence relation is measurable. More generally, it is the (essentially unique) measurable equivalence relation which has

the smallest equivalence classes among all of those whose equivalence classes contain the $H \times e$ orbits. To obtain it, one fibers $S \times G$ into ergodic parts under the $H \times e$ action and takes the fibers as the equivalence classes.

From a virtual group point of view (the cohomology class of), a cocycle π is a homomorphism into G of the virtual subgroup of H associated with the given ergodic action of H on S. The virtual subgroup associated with the ergodic action of G just defined is then the closure of the range of this homomorphism (and, in some cases, simply the range). From this point of view, the theorem of Ambrose cited above may be regarded as the assertion that a large family of virtual subgroups of the additive group of the real line are homomorphic images of virtual subgroups of the integers. A number of other results in ergodic theory have recently been recognized to have closely related interpretations. A startling result of R. M. Belinskaya published in 1968 and based on results of Vershik states that any two properly ergodic actions of the integers which have finite invariant measures define equivalence relations which are isomorphic in a certain sense. This implies that the corresponding virtual subgroups are isomorphic as virtual groups so that there are many fewer distinct virtual groups than one might have believed. Actually, an equivalent result is contained in a paper of H. A. Dye published in 1959, but its significance was obscured for some time by the Boolean algebra viewpoint in terms of which it was formulated. Moreover, a generalization which Dye published four years later shows that the group of integers may be replaced by a more general countable discrete group and, in particular, by a direct product of finitely many replicas of the integers. As observed by Peter Forrest in his 1972 Harvard thesis, Ambrose's theorem may be shown to be equivalent to the stronger statement that the virtual subgroup of the additive group of the real line defined by an ergodic action with a finite invariant measure is actually isomorphic to such a virtual subgroup of the integers. Forrest then went on to generalize the theorems of both Belinskaya and Ambrose by proving an analogue of Belinskaya's theorem in which the integers are replaced by a finite product of replicas of the integers and replicas of the additive group of the real line.

The examples of properly ergodic actions given in Section 4 were of two radically different kinds. On the one hand, there were the actions with pure point spectra and on the other those defined by groups of automorphisms of countable products of compact commutative groups. Moreover, those in the latter category were defined only for discrete

groups. This is because the automorphisms were defined as permutations of the index set over which the infinite product was taken. We propose now to fulfill the promise made at the end of Section 4 and show how to define analogues of these actions for arbitrary separable locally compact groups. The definition can best be motivated by looking at a probabilistic interpretation of what was done in the discrete case. Let $\mathcal{M}$ be a countable index set, and let $m \to f_m$ denote an assignment of a random variable on a probability space S, μ to each m in M. Suppose that these random variables are independent, identically distributed and separate the points of S. For each s in S, let F_s be the function from $\mathcal{M}$ to the space A in which the f_m take their values defined by $F_s(m) = f_m(s)$. Then $s \to F_s$ allows one to identify S with a subset of the space $A^{\mathcal{M}}$ of all functions from $\mathcal{M}$ to A and to think of μ as a measure in $A^{\mathcal{M}}$ which assigns the measure zero to the complement of the image of S. The fact that this complement is of measure zero allows one to replace S by $A^{\mathcal{M}}$. Of course, $A^{\mathcal{M}}$ may be looked upon as a direct product over the index set $\mathcal{M}$ of replicas of the standard Borel space A. Moreover, the fact that the random variables are independent implies that μ is the direct product of replicas of ν, where ν is the probability measure in A defining the common distribution of the f_m. Evidently, every permutation of $\mathcal{M}$ defines an automorphism of $A^{\mathcal{M}}$ as a Borel space which preserves the measure μ. Now suppose that the measure ν has no points of finite measure or else that A is finite and all points have the same ν measure. Then, by the theorems quoted in Section 2, A, ν will be isomorphic as a standard Borel measure space, either to the circle group with Haar measure or else to a finite cyclic group with Haar measure. But replacing A, ν by a separable compact group with its Haar measure reduces our construction to that of the last part of Section 4. In short, whenever a group G permutes a countable family of identically distributed independent random variables among themselves, there is a canonically associated action of G on the underlying probability measure space, and to within isomorphism, the actions constructed in the last part of Section 4 are the special cases of those in which the common distribution either has no atoms or else is atomic with all atoms of equal probability. It follows as a corollary of this discussion and the results of Section 4 that the action of the integers underlying a discrete stationary stochastic process is necessarily ergodic whenever the random variables are independent. While the above argument seems to demand a restriction on the common distribution, this restriction is easily removed by the following device. Any measurable equivalence relation in the space A

defines a G invariant equivalence relation in $A^{\mathscr{M}}$, and the corresponding quotient action is the same as the action one gets by replacing A by a quotient space $\tilde{A}$. Taking A to be the circle group, one can choose the equivalence relation so that $\tilde{A}$, together with the image in it of Haar measure, becomes an arbitrary standard Borel probability space. Thus, taking suitable quotient actions of the actions defined in the last part of Section 4, one obtains all the actions that one can get by permuting identically distributed independent random variables. As already pointed out, quotient actions of ergodic actions are necessarily ergodic.

In order to find an analogue for continuous groups of the ergodic actions defined by independent random variables, one needs to be able to deal somehow with a continuum of independent random variables. The most obvious thing to do is to seek a continuous stationary stochastic process for which the random variables $f_t(s)$ are mutually independent. It is easy to see, however, that such a process cannot exist. If it did, then $f_t(s) = f_0([s]t)$ for some f_0. By a change of variable, it may be assumed that f^0 is in $\mathscr{L}^2(S, \mu)$, and it follows that $\int f_0([s]t)\bar{f}_0(s)\,d\mu(s) = (V_t(f_0)\cdot f_0)$ is continuous in t. On the other hand, independence demands that the f_t be mutually orthogonal and, hence, that $(V_t(f_0)\cdot f_0)$ be zero except when $t = 0$.

The way around this difficulty is suggested by looking at a fundamental difference between families of random variables parameterized on the one hand by a countable set and on the other by a continuum. Let $\mathscr{M}$ be a countable set, and let $m \to f_m$ be an assignment of a real- or complex-valued random variable to each m in $\mathscr{M}$. For each finite subset E of $\mathscr{M}$, let $f_E(s) = \sum_{m\in E} f_m(s)$. Then the f_E constitute a family of random variables parametrized by the finite subsets of $\mathscr{M}$ and having the further property that $f_{E_1\cup E_2} = f_{E_1} + f_{E_2}$ whenever $E_1 \cap E_2 = 0$. Conversely, any such assignment $E \to g_E$ may be obtained in this way from the point assignment $m \to g_m$, where $g_m = g_{\{m\}}$. Now suppose that $\mathscr{M}$ is a standard Borel G space and that ν is a measure in $\mathscr{M}$. Let $m \to f_m$ be an assignment of a real- or complex-valued random variable to each m in $\mathscr{M}$. If suitable measurability and integrability conditions are satisfied, one can assign a random variable f_E to each Borel subset E of $\mathscr{M}$ for which $\nu(E) < \infty$ by defining $f_E(s)$ to be $\int_E f_m(s)\,d\nu(m)$. This assignment will then have the property that $f_{E_1\cup E_2} = f_{E_1} + f_{E_2}$ whenever E_1 and E_2 are disjoint and also the property that $f_E = 0$ whenever $\nu(E) = 0$. However, unlike the discrete case, there will be in general no converse. Unless ν is atomic, there will exist assignments $E \to g_E$ having the above two properties which cannot be obtained by integrating a point assign-

ment $m \to g_m$; that is, which are not "differentiable". Now in the discrete case to say that the f_m are independent is equivalent to saying that f_{E_1}, f_{E_2},... and f_{E_j} are independent whenever the E_j are disjoint. This suggests that there might exist additive set assignments $E \to f_E$ in the continuous case which are not differentiable but are such that disjoint sets map into independent random variables. Such assignments do exist and provide the sought-for continuum substitute for countable sets of independent random variables.

Adopting some terminology introduced by Wiener (and since abandoned), let us define a *chaos* on the standard Borel measure space $\mathscr{M}$, ν to be an assignment of a real- or complex-valued random variable f_E on a standard probability space S, μ to each Borel subset E of $\mathscr{M}$ of finite measure in such a manner that $f_{E_1 \cup E_2} = f_{E_1} + f_{E_2}$ whenever E_1 and E_2 are disjoint and $f_E = 0$ whenever $\nu(E) = 0$. Now suppose that $\mathscr{M}$ is a transitive Borel G space for some separable locally compact group G and that the measure ν is G invariant. In that event, one says that the chaos is *homogeneous* (with respect to the G action) if the random variables $f_{(E_1)x}$, $f_{(E_2)x}$,..., $f_{(E_j)x}$ have the same joint probability distribution as the random variables f_{E_1}, f_{E_2},..., f_{E_j} for arbitrary x in G and arbitrary disjoint Borel subsets E_1, E_2,..., E_j of $\mathscr{M}$. In the particular case in which $\mathscr{M} = G =$ the additive group of the integers, the action is translation and ν is the counting measure, the notion of homogeneous chaos is equivalent to that of stationary stochastic process. An example which occurs in statistical mechanics is provided by a gas idealized to occupy all of space. In this case, $\mathscr{M}$ is Euclidean space, G is the group of all isometries of $\mathscr{M}$, and for each Borel subset E of $\mathscr{M}$, f_E is the random variable whose value at any observation is the number of gas molecules in the set E. Other examples occur in the theory of turbulence.

Given a homogeneous chaos (normalized so that the random variables separate the points of S), there is an essentially unique way of defining an action of G on S which is consistent with its action on $\mathscr{M}$ in the sense that $f_{[E]x^{-1}}(s) \equiv f([s]x)$ and μ is invariant under this G action. Indeed, a homogeneous chaos may be equivalently defined as the system consisting of a standard Borel G space S, an invariant probability measure μ in S, and a chaos with S, μ as probability space which satisfies the identity $f_{[E]x^{-1}}(s) \equiv f([s]x)$. The G action on S, μ of course plays the same role in the theory of a homogeneous chaos that the underlying action of the real line or integers plays in the theory of stationary stochastic processes.

To find examples of ergodic actions of a group G generalizing those

defined for discrete G at the end of Section 4, it is now clear what we must do. For each transitive G space $\mathcal{M}$ with an invariant measure ν, we must find a chaos homogeneous with respect to the G action such that f_{E_1} and f_{E_2} are independent whenever E_1 and E_2 are disjoint. It is quite easy to show that the associated action of G on S will often be ergodic. However, showing that the f_E exist is far from easy. It was first accomplished by Norbert Wiener in the early 1920's as the key step in his construction of a mathematical theory of Brownian motion. He considered only the case in which $\mathcal{M} = G =$ additive group of the real line. However, using the isomorphism theorems for measure spaces quoted in Section 2, one can deduce what one needs in the general case from Wiener's results. In the real-line case, a chaos is determined by its values on finite intervals and, hence, its values on intervals of the form $0 \leqslant x \leqslant a$. Setting $f_a = f_{[0,a]}$, one obtains a continuous stationary stochastic process $a \to f_a$ which uniquely determines the homogeneous chaos. Independence in these terms amounts to the independence of differences $f_{a_0} - f_{b_i}$ when the intervals $[a_i, b_i]$ are disjoint. One speaks of a *stationary stochastic process with independent increments* and this is what Wiener actually produced (he did not introduce the chaos notion until 1938).

Let $E \to f_E$ be any homogeneous chaos on the real line, and let I be a finite interval. For each $n = 1, 2,...,$ I can be written as a sum of n disjoint intervals of equal length $I = I_1 \cup I_2 \cup \cdots \cup I_n$ and, correspondingly, $f_I = f_{I_1} + f_{I_2} + \cdots + f_{I_n}$. Now, if f takes disjoint sets into independent random variables, the f_{I_j} will be independent and identically distributed. In other words, for fixed I, f_I will be a sum of n independent identically distributed random variables for all n—no matter how large. This condition strongly restricts the possibilities for the distribution of f_I, and (as suggested by the central limit theorem) one of the few possibilities is that f_I be "normally distributed with mean zero and variance proportional to the length of I"; that is, that the probability measure in the real line defining the distribution have a density proportional to e^{-x^2/σ^2}, where σ is proportional to the length of I. What Wiener proved in effect is that such a chaos always exists. It is unique and we shall refer to it as the *Wiener chaos*. Using the isomorphism theorems as indicated above, one can now prove more generally, that, given any standard Borel space $\mathcal{M}$ equipped with an atom-free invariant measure ν and any real $\lambda > 0$, there exists a chaos $E \to f_E$ defined on $\mathcal{M}$ such that (a) $f_{E_1}, f_{E_2},..., f_{E_i}$ are independent when the E_i are disjoint and (b) each f_E is normally distributed with variance $\lambda\nu(E)$. This chaos is unique up

to the obvious isomorphism and is homogeneous with respect to G whenever ν is invariant under a transitive action of G. The associated action of G can easily be shown to be ergodic in many special cases; in particular, when $\mathcal{M} = G$, the action is right translation, ν is Haar measure, and G is not compact. In the special case in which G is the additive group of the real line so that $a \to f_a = f_{[0,a]}$ is a stationary stochastic process, the measure μ defines a measure in the space of sample functions with respect to which almost all functions are continuous. This is the famous "Wiener measure" in the space of continuous functions.

Another possibility for the distribution of the random variables f_E is the Poisson distribution. This one-parameter family of distributions is concentrated in the positive integers and assigns to n the probability $e^{-\lambda}\lambda^n/n!$, where λ is a fixed positive real number. There exists a corresponding chaos which in three-dimensional space is the positive integer valued chaos describing the distribution of molecules in a perfect gas.

9. Thermodynamics, Statistical Mechanics, and Entropy

Consider a gas which (for simplicity) will be assumed to consist of identical molecules which do not interact chemically. If unit mass of this gas is enclosed in a container of volume v and allowed to "come to equilibrium" at a fixed absolute temperature T, there will be a pressure p exerted on the walls of the container which depends only on T, v, and the chemical nature of the gas molecules. When v and T are sufficiently large, one has $p = NRT/v$ to a high degree of approximation, where N is the number of molecules and R is a universal constant called the gas constant. In a so-called "perfect gas", the formula $p = NRT/v$ holds exactly, but in an actual gas one has the more complicated relationship

$$p = \frac{NRT}{v}\left(1 + \frac{B(T)}{v} + \frac{C(T)}{v^2} + \cdots\right),$$

where the functions B, C, etc., must be determined by experiment and are called the *virial coefficients*. In addition to determining the pressure as a function of volume and temperature, one can also measure the so-called *heat capacities* of the gas as a function of these same two variables. If $H(\Delta T, T, v)$ is the quantity of heat required to raise the temperature of the gas from T to $T+\Delta T$ while holding the volume fixed at v, then by

definition the *heat capacity at constant volume* C_v is $\lim_{\Delta T \to 0} H(\Delta T, T, v)/\Delta T$. The *heat capacity at constant pressure* C_p is defined analogously holding the pressure constant and varying the volume. For a perfect gas, C_v and C_p are different constants, but for an actual gas they depend on v and T much as the pressure does, and one must determine two new sets of functions of T analogous to the virial coefficients in order to know the thermal properties of the gas.

Until one takes account of the laws of thermodynamics, there is no *a priori* relationship between the three functions p, C_v, and C_p. They must be determined separately by independent experiments. However, the law of conservation of energy implies that there exists a positive constant λ and a fourth function U called the *internal energy* function from which C_v and C_p may be computed by means of the formulas

$$C_v = \frac{1}{\lambda}\frac{\partial U}{\partial T} \qquad C_p = C_v - \frac{1}{\lambda}\left(p + \frac{\partial U}{\partial v}\right)\left(\frac{\partial p}{\partial T}\Big/\frac{\partial p}{\partial v}\right).$$

Thus the functions p, C_v, and C_p are not in fact independent, and it suffices to know the two functions p and U and the constant λ in order to be able to compute C_v and C_p as well. Moreover, λ is a universal constant—the same for all gases. Solving the above equations for $\partial U/\partial T$ and $\partial U/\partial v$ one finds that

$$\frac{\partial U}{\partial v} = \frac{\lambda(C_v - C_p)\,\partial p/\partial v}{\partial p/\partial T} - p$$

and $\partial U/\partial T = \lambda C_v$, so that U is uniquely determined up to an additive constant by p, C_v, C_p, and λ. Of course, U and λ will not exist at all unless p, C_v, and C_p are properly related, and the fact that they are so related is the content of the first law of thermodynamics in the present context. A simple computation shows that the condition on C_v, C_p, and p is that

$$\frac{\partial p}{\partial T}\Big/\left(\frac{\partial}{\partial T}\left[\frac{(C_v - C_p)\,\partial p/\partial v}{\partial p/\partial T}\right] - \frac{\partial C_v}{\partial v}\right)$$

should be a constant and this constant is λ. Thus λ is uniquely determined by p, C_v, and C_p.

The physical idea behind these mathematical considerations is that heat is a form of energy and never appears or disappears without an equivalent amount of some other form of energy disappearing or appearing to compensate for it. Let λ be the conversion factor which

converts quantity of heat into a corresponding number of units of mechanical energy. Then taking into account the mechanical work done or absorbed by an expanding or contracting gas, one finds that the net total change in the energy of the gas when v and T change from v_1, T_1 to v_2, T_2 is the line integral

$$\int_L \left[\left(\frac{\lambda(C_v - C_p)\,\partial p/\partial v}{\partial p/\partial T} - p\right) dv + \lambda C_v\, dT\right]$$

taken over whatever path L that the gas follows in making the change. Conservation of energy requires that this line integral should be independent of the path and, hence, that U should exist. Of course, the value of the line integral will be $U(v_2, T_2) - U(v_1, T_1)$.

Just as the first law of thermodynamics implies a relationship between the three independently defined functions p, C_v, and C_p, so the second law implies a relationship between p and U so that these functions do not have to be independently measured either. The second law has many equivalent formulations all having to do with the closely related facts that turning mechanical energy into heat energy and letting heat flow from high to low temperatures are in a sense irreversible processes. One formulation says that it is impossible to transfer a nonzero quantity of heat from a body at one temperature to a body at a higher temperature without effecting changes in other bodies. In proving the equivalence of various formulations as well as in deducing their consequences for the functions p and U, one considers the energy changes that take place when one changes the volume and temperature of a gas along a so-called Carnot cycle; that is, a closed curve made up of two curves of constant temperature and two "adiabatic" curves. By definition, an adiabatic curve is one along which the change in U is exactly compensated by the mechanical work done so that no heat energy is given off or absorbed. In mathematical terms, it is a solution of the first-order differential equation

$$dv/dT = -\partial U/\partial T/((\partial U/\partial v) + p).$$

The net effect of transversing a Carnot cycle in the proper direction is to take a quantity of heat Q_1 at a temperature T_1, transfer Q_2 of it to a body of lower temperature T_2, and turn the difference $Q_1 - Q_2$ into mechanical energy. By transversing the cycle in the opposite direction, one can reverse the process. An easy argument using the second law now shows that the "efficiency" $(Q_1 - Q_2)/Q_1$ must be the same for all cycles and all

substances operating between the same two temperatures T_1 and T_2. Moreover, it is easy to show that up to a multiplicative constant there is one and only one way of parameterizing temperature so that this efficiency is equal to $(T_1 - T_2)/T_1$. The absolute temperature scale is defined by choosing one of these parameterizations.

Now given U and p, one can compute the efficiency $(Q_1 - Q_2)/Q_1$ for any Carnot cycle. When one does this one finds that $(Q_1 - Q_2)/Q_1 = (T_1 - T_2)/T_1$ for all Carnot cycles if and only if the line integral $\int_L (\partial U/\partial T)\, dT/T + [(\partial U/\partial v + p)\, dv]/T$ depends only upon the end points of the path L; that is, if and only if there exists a function S such that $\partial S/\partial T = (1/T)\, \partial U/\partial T$ and $\partial S/\partial v = (1/T)(p + (\partial U/\partial v))$. The function S is, of course, uniquely determined by U and p up to an additive constant. It is called the *entropy* function, and its existence is equivalent to the second law of thermodynamics in the present context. Because of the existence of the entropy function, there is a single function of v and T from which U, p, S, and, hence, C_v and C_p can all be computed. This is the *free energy* function F defined as $U - TS$. Indeed

$$\partial F/\partial v = (\partial U/\partial v) - T(\partial S/\partial v) = (\partial U/\partial v) - (p + (\partial U/\partial v)) = -p$$

and

$$\partial F/\partial T = (\partial U/\partial T) - S + T(\partial S/\partial T) = (\partial U/\partial T) - S - (\partial U/\partial T) = -S$$

so that

$$p = -\partial F/\partial v, \qquad S = -\partial F/\partial T, \qquad U = F - T(\partial F/\partial T).$$

Notice that knowing p determines F up to an additive function of T alone. Thus the two laws of thermodynamics imply that, in addition to the virial coefficients in the pressure function, one needs to measure just one more function of the temperature in order to know all the thermodynamic functions of v and T.

In thermodynamics, heat and temperature are accepted as fundamental concepts not reducible to anything else. In statistical mechanics, on the other hand, one attempts to explain heat, temperature, and the laws of thermodynamics as consequences of the laws of particle mechanics applied to the molecules of the substances concerned. In particular, it provides an algorithm for computing the free-energy function of a gas once one knows the differential equations of motion of the molecules

of which the gas is composed. Given a gas of N molecules moving in a fixed enclosure V of volume v, let Ω denote the corresponding phase space; that is, the set of all possible positions q_i and momenta p_i of the molecules, let H be the function on Ω giving the total energy in terms of the positions and momenta, and let $t \to \alpha_t$ denote the one-parameter group of automorphisms of Ω whose orbits are the curves obtained by integrating the equations of motion $dq_i/dt = \partial H/\partial p_i$, $dp_i/dt = -\partial H/\partial q_i$. Let ζ denote the Lebesgue measure in Ω associated with the coordinatization $q_1, q_2, \ldots, q_{3n}$ $p_1, p_2, \ldots, p_{3N}$. As noted by Liouville, ζ is invariant under the real line action on Ω defined by the α_t. Now a great many different points of Ω define states of motion of the gas which appear the same to a macroscopic observer. According to statistical mechanics, the phase space counterpart of our gas in equilibrium at temperature T is not some point of Ω but a probability measure μ_T in Ω. The idea is that if one could actually determine ω in Ω, one would get different answers at different times and with different ways of preparing the system. $\mu_T(A)$ is the probability that a determination of ω would give a result in the set $A \subseteq \Omega$. What probability measure in Ω is μ_T to be? The answer proposed by Gibbs and now more or less universally accepted is that

$$\mu_T(A) = \frac{\int_A e^{-H/kT}\, d\zeta}{\int_\Omega e^{-H/kT}\, d\zeta}$$

where k is a universal constant known as Boltzman's constant. With this choice for μ_T, the expected value of the kinetic energy of the gas molecules can be computed to be $(3/2)\, NkT$ so that kT is just $2/3$ of the mean kinetic energy per molecule. Since the absolute temperature thus turns out to be proportional to the mean kinetic energy per molecule, one has a "natural" choice for the arbitrary multiplicative constant in the definition of absolute temperature. Except for the $2/3$, Boltzman's constant is just a conversion factor from conventional units to the "natural" ones. Identifying $U(v, T)$ with the expected value $\int H\, d\mu_T$ of H, one has an algorithm for computing U once H is known. The volume v enters into the picture because H and Ω depend upon the enclosure. Actually, $\int H\, d\mu_T$ will depend on the shape as well as the volume of the enclosure, but the dependence on the former is very slight as soon as v is reasonably large. One can avoid the difficulty by confining attention to one particular shape—say, cubic. A simple argument shows that $p(v, T)$ may be identified with $\int \partial H/\partial v\, d\mu_T$, and given these identifications, one verifies that $(dU + p\, dv)/T$ is indeed exact, so that second

law is a consequence of statistical mechanics. An easy computation shows that the free-energy function F is given by the formula

$$F(v, T) = -kT \log \int e^{-H/kT}\, d\zeta.$$

Thus to compute F and, hence, all thermodynamic functions, one needs to know only $\int e^{-H/kT}\, d\zeta$, the so-called *partition function*. Using the formula $S = -\partial F/\partial T$ for the entropy leads at once to the formula

$$S(v, T) = k \log \int e^{-H/kT}\, d\zeta + \frac{1/T \int He^{-H/kT}\, d\zeta}{\int e^{-H/kT}\, d\zeta}$$

and setting

$$\rho_T = \frac{e^{-H/kT}}{\int e^{-H/kT}\, d\zeta},$$

this may be rewritten as

$$S(v, T) = -k \int \rho_T \log \rho_T\, d\zeta.$$

Thus the entropy depends only on the probability measure $\mu_T = \rho_T\, d\zeta$ and not on the particular Hamiltonian function H which gives rise to it. As will be explained more fully below, the expression $-\int \rho_T \log \rho_T\, d\zeta$ may be regarded as a measure of the "degree of uncertainty" associated with μ_T. It is small when μ_T is highly concentrated and large when it is "spread out". One has then a simple statistical interpretation of the entropy of a gas. It is just k multiplied by the degree of uncertainty in the microscopic state of the gas for that macroscopic state whose entropy is in question. The Gibbs measures μ_T can be characterized as those probability measures of the form $\rho\, d\zeta$ which maximize $-\int \rho \log \rho\, d\zeta$ subject to the side condition that $\int \rho H\, d\zeta$ should have a fixed value and this fact can be used as a sort of justification for their use in statistical mechanics.

The sense in which $-\int \rho_T \log \rho_T\, d\zeta$ may be regarded as a measure of the "degree of uncertainty" associated with the probability measure $\rho_T\, d\zeta$ can best be appreciated by considering a discrete analogue. Let $A_1, A_2, \ldots, A_n$ be a finite set of events exactly one of which will occur at any one time. Let p_j be the probability of occurrence of A_j so that $0 \leqslant p_j \leqslant 1$ and $p_1 + p_2 + \cdots + p_n = 1$. If $p_{j_0} = 1$ for some j_0, then A_{j_0} surely occurs and there is no uncertainty involved. If on the other hand $p_1 = p_2 = p_1 = p_2 = \cdots = p_n = 1/n$, so that all events are equally likely, there is as much uncertainty as possible. Now one's

intuition suggests that having $p_1 = p_2 = 1/2$, $p_3 = p_4 \cdots = p_n = 0$ involves less uncertainty say than having $p_1 = p_2 = p_3 = p_4 = 1/4$, $p_5 = p_6 = \cdots = p_n = 0$, but it is easy to construct examples in which it is not clear a priori which of two sets of probabilities for the A_j involves the greater uncertainty. However, in 1948, Claude Shannon in the course of developing a general theory of "information" (cf. Section 11) showed that if certain simple and natural axioms are to be satisfied, then there is to within a multiplicative constant one and only one way of assigning a real number to each finite system $p_1, p_2, \ldots, p_n$ which measures the "degree of undertainty" exemplified by such a probability distribution. The constant may be chosen so that the number is $-\sum_{j=1}^{n} p_j \log p_j$, where $0 \log 0$ is interpreted as 0. While the formula was not new, such a satisfying justification for it had not previously been given. The expression $-\int \rho \log \rho \, d\zeta$ is an obvious continuous analogue of $-\sum_{j=1}^{n} p_j \log p_j$—provided one is willing to accept ζ as a reference measure. Without a reference measure, one cannot distinguish one atom-free probability measure from another.

One may explain the fact that many different points ω in Ω appear the same to a macroscopic observer by supposing that those functions ψ on Ω which a macroscopic observer can observe are "nearly constant" on the curves $t \to \alpha_t(\omega)$ for almost all ω. To make this statement more precise, considering the fibering of Ω corresponding to a direct integral decomposition of the invariant measure ζ into ergodic components ζ_r. Let the fiber supporting ζ_r be Ω_r. One can show that $\zeta_r(\Omega_r)$ is finite for all r, and the ergodic theorem then implies that whenever ψ is a bounded Borel function, there is a mean value

$$\bar{\psi}(\omega) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \psi(\alpha_t(\omega))$$

defined for almost all ω and depending only on r. To say that ψ is "nearly constant" on the curves $t \to \alpha_t(\omega)$ is to say that $\overline{(\psi(\omega) - \bar{\psi}(\omega))^2}$ is very small compared to $\bar{\psi}(\omega)^2$; i.e., that for almost all ω, $\psi(\alpha_t(\omega))$ is relatively very near to $\bar{\psi}(\omega)$ for a very large fraction of the time. When this holds, ψ will appear to be a constant in time to a macroscopic observer, and, moreover, the constant will be the same for ζ_r almost all ω in Ω_r. In other words, to a macroscopic observer, the state of the system depends only on the (abstract) parameter r and is constant in time. Now the fibering of Ω provided by the constant energy hypersurfaces Ω_E decomposes ζ into measures ζ_E, and the decomposition defined by the Ω_r is clearly a refinement of this. Suppose that it were not

a proper refinement; that is, that the action of the α_t on the Ω_E were already ergodic with respect to the ζ_E. Then the macroscopic state of the gas would be completely determined by the two real variables v and E, and this is what one finds experimentally. Moreover, under this hypothesis of ergodicity for the ξ_E, it would also follow that the mean values $\bar{\psi}$ and $(\psi - \bar{\psi})^2$ could be computed without solving the equations of motion in order to find the α_t. One would have $\bar{\psi}(E) = \int \psi(\omega)\, d\zeta_E/\zeta_E(\Omega_E)$ and a similar formula for $(\psi - \bar{\psi})^2$. It was with this application in mind that von Neumann and G. D. Birkhoff proved their versions of the ergodic theorem and created the subject of ergodic theory. Unfortunately, it seems extremely difficult to decide whether the invariant measures ζ_E are in fact ergodic, and the problem is still unsolved although Ya. G. Sinai has made important progress in the last decade.

The question of "the approach to equilibrium" may be viewed in a similar spirit. Whether a point ω_0 in Ω represents "equilibrium" or not is not precisely defined but depends upon whether $\psi(\omega)$ is "near" to $\bar{\psi}(\omega_0)$ for all macroscopically observable functions ψ. Assuming that ψ is nearly constant in the sense described above, $\alpha_t(\omega_0)$ will be an equilibrium point for a portion of the time which is very close to one. Thus, even if ω_0 is not an equilibrium point, $\alpha_t(\omega_0)$ will be such for some $t > 0$ and will continue to be such except during rare intervals which are brief compared to the distances between them.

The above considerations suggest that μ_T should be taken to be $\zeta_E/\zeta_E(\Omega_E)$, where $E = U(v, T)$, rather than the Gibbs measure $\rho_T\, d\zeta$. However, for gases with a large number of particles, the two are actually very close together and the Gibbs measure is much easier to deal with technically. It is, in fact, a continuous weighted mean of all measures of the form $\zeta_E/\zeta_E(\Omega_E)$ with practically all of the weight going to a very short interval on the E axis.

All known arguments designed to establish the relevance of the Gibbs measure and its corollary the algorithm $F(v, T) = -kT \log \int e^{-H/kT}\, d\zeta$ are less than compelling and vulnerable to objections of one sort or another. Nevertheless, the algorithm and other consequences of assuming the Gibbs measure are believed to be valid (when the temperature is sufficiently high) and one of the main tasks of the specialist in (equilibrium) statistical mechanics is to compute $F(v, T)$ starting with hypotheses about the function H. The reason for the qualification about high temperatures is that quantum effects become serious when the temperature is low and a treatment based on classical mechanics is no longer accurate.

The main problem in computing F from H is, of course, that of computing the partition function $\int e^{-H/kT}\, d\zeta = P(v, T)$ which is just $\int e^{-x/kT}\, d\beta(x)$, where β is the measure in the real line defined by the formula $\beta(A) = \zeta(H^{-1}(A))$. Finding P from β is merely a matter of computing a Laplace transform and the difficult problem is that of finding β when H is given. In dealing with this difficult problem, it is useful to observe that it can be quickly reduced to a problem involving configurations only. The point is that H is a sum

$$\sum_{j=1}^{3N} \frac{1}{2m}(p_j^2) + W(q_1, q_2, \ldots, q_{3N})$$

of a very simple function of the p_j and a function W (the potential energy) of the q_j. Here, m is the mass of an individual molecule. One computes easily that each Gibbs measure μ_T is a direct product of measures $\mu_T{}^1$ and $\mu_T{}^2$, where $\mu_T{}^1$ is a measure in R^{3N}, the space of all possible momenta, and $\mu_T{}^2$ is a measure in V^N, the space of all possible configurations. Moreover, $\mu_T{}^1$ factors further as a $3N$-fold product of replicas of identical measures on the real line, the latter being constant multiples of $e^{-p^2/2mkT}\, dp$. It follows that the partition function is a constant times $(kT)^{3N/2}P_0(v, T)$, where $P_0(v, T)$ is a modified partition function defined as $\int_{V^N} e^{-W/kT}\, d\zeta_0$, where ζ_0 is Lebesgue measure in "configuration space" V^N. Correspondingly, $F(v, T)$ up to an additive constant is $-(3/2)\, NkT \log kT - kT \log P_0(v, T)$, and everything depends upon being able to compute the modified partition function $P_0(v, T)$ or, more or less equivalently, the measure β_0, where $\beta_0(A) = \zeta_0(W^{-1}(A))$.

Notice that omitting the term $-(3/2)\, NkT \log kT$ does not change the pressure function at all and changes the entropy and internal energy functions only by adding N-times known functions of T which are the same for all gases. Thus, nothing essential is lost if one ignores the momenta and kinetic energy and develops what might be called *configurational thermodynamics* in which one replaces Ω by V^N, H by W, and μ_T by $\mu_T{}^2$. Given V and α, let ρ be any positive Borel function on V^N such that $\int \rho\, d\alpha^N = 1$. Let $W = -k \log \rho$. Then the Gibbs measures for the potential energy W will be

$$\frac{e^{(\log\rho)/T}\, d\alpha^N}{\int_{V^N} e^{(\log\rho)/T}\, d\alpha^N} = \frac{\rho^{1/T}\, d\alpha^N}{\int_{V^N} \rho^{1/T}\, d\alpha^N},$$

and, in particular, the Gibbs measure for unit temperature will be $\rho\, d\alpha^N$.

In other words, *every* probability measure in V^N which is absolutely continuous with respect to α^N and has an everywhere positive Radon–Nikodym derivative is the Gibbs measure at unit temperature for some (uniquely determined) potential energy function W. Moreover, it is easy to see that one can allow ρ to be merely nonnegative by permitting W to take on the value ∞. From this point of view, configurational statistical mechanics (in a fixed volume) is just the theory of the one-parameter families of probability measures in V^N of the form $\rho^{1/T}\, d\alpha^N / \int \rho^{1/T}\, d\alpha^N$, where ρ is an arbitrary nonnegative Borel function for which the integrals exist. In particular, given any probability measure μ in V^N such that $d\mu/d\alpha^N$ exists and satisfies the indicated integrability conditions, it makes sense to ask for the associated free-energy and entropy functions.

In order to include volume dependence, one must have a way of passing from a W in one volume to a "corresponding" W in the other. In actual gases, this comes about naturally because W is a sum of contributions from "two body potentials". One has a function w defined on the nonnegative real axis and

$$W(x_1 y_1 z_1 , x_2 y_2 z_2 \cdots x_N y_N z_N) = \sum_{\substack{i \neq j \\ i,j=1}}^{N} w(\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2})$$

a function which is well-defined on V^N for all V as soon as w is given. Of course, only very special measures $\rho\, d\alpha^N$ will arise from such a W. On the other hand, those that do play a special role in probability theory quite apart from their significance for statistical mechanics (cf. Section 10)

Mathematically speaking, there is no reason why configurational statistical mechanics should be limited to the case in which V is a region in Euclidean space. All definitions continue to make sense when V, α is replaced by a quite general measure space which can even be discrete. Thus far, only one special case of his suggested generalization has been of actual interest to physicists. This is the case in which the measure space is the set of all points with integer coordinates in some volume V of Euclidean space and the measure is the counting measure. One speaks then of a *lattice gas*. The properties of lattice gases have been studied as a preliminary indication of what to expect in dealing with the more difficult case of actual gases. The theory of lattice gases is also interesting because it is mathematically equivalent to certain other (more realistic) problems in statistical mechanics.

10. Statistical Mechanics in the Infinite Volume Limit and Its Connections with Stationary Stochastic Processes and Homogeneous Chaoses

Given a gas of N molecules moving in a region V of space, what happens to the entropy, pressure, internal energy, etc., if one changes N to N' and at the same time replaces V by a region V' whose volume is N'/N times as great as that of V? Intuition suggests and experiment confirms that, if the temperature remains the same, then the pressure remains the same, while the entropy and internal energy are multiplied by N'/N. Accordingly, one speaks of pressure as an intensive quantity and of entropy and internal energy as extensive quantities. If one wants relationships which do not depend upon N or v, one looks at the entropy per unit volume, the internal energy per unit volume, and the pressure as functions of v/N and T. However, if one tries to compute these functions from the algorithms of statistical mechanics described in Section 9, one finds that there is a slight dependence on size as well. This dependence can be traced to the so-called "edge effect"—the fact that conditions become different near the walls of the container. On the other hand, it is clear that for a given v/N, the edge effect becomes relatively less and less important as N and v get large, and this, in turn, suggests that p, U/v should have definite limits as N and v tend to ∞. These limits, if they exist, are the quantities one is really interested in computing, since the failure to find edge effects in experiments suggests that gases in laboratory quantities already are extensive enough to make edge effects too small to measure. One speaks of them as the *thermodynamic limits* of the functions concerned. As might be expected, they tend to be rather simpler functions of T and v/N than the corresponding functions with edge effects included.

The first attempts to gave a rigorous mathematical proof of the existence of thermodynamic limits were made independently by van Hove in 1949 and by Yang and Lee in 1952. A thoroughgoing systematic attack on the problem was begun by D. Ruelle in 1963 and quickly led to quite complete results for rather general potentials. Ruelle's results together with those of other workers in the field may be found in a book entitled "Statistical Mechanics—rigorous results" which he published in 1968.

As a refinement of computing the thermodynamic functions for finite systems of constant density and increasing volume and then taking their limits as the volume becomes infinite, one can conceive of trying

to find a (hypothetical) infinite system for which these limit functions are the actual thermodynamic functions themselves. This turns out to be a feasible program which has been in the course of development since the publication of a fundamental paper by H. Araki and E. J. Woods in 1963. These authors dealt with the special case of a perfect "Bose gas" in quantum statistical mechanics. A more general attack was inaugurated by Ruelle in 1965 and continued by Ruelle and others, especially R. L. Dobrushin. In the approach of Ruelle and his coworkers, emphasis is placed on developing the classical and quantum cases simultaneously using the theory of noncommutative C^* algebras as a technical tool. This has the disadvantage of obscuring somewhat the connection with the theory of stochastic processes, which we wish to emphasize here.

Accordingly, the description to follow will be closer in spirit to that given in the papers of Dobrushin. Dobrushin, a probabilist by training and a specialist in information theory, confined himself to classical statistical mechanics—and in fact to the lattice-gas case.

Prior to introducing the infinite system limit, let us study the lattice gas notion in somewhat more detail. Let $\mathcal{M}$ denote the finite set of lattice points in our volume V, and let there be N gas molecules so that $\mathcal{M}^N$ is the configuration space of the gas. Let W be the real $\cup\ \infty$-valued function on $\mathcal{M}^N$ defining the potential energy of each configuration. The macroscopic state of the gas at temperature T is then defined by the probability measure μ_T which assigns a probability $e^{-W(s)/kT}/\sum_{s\in\mathcal{M}^N} e^{-W(s)/kT}$ to each $s = m_1, m_2, \ldots, m_n$ in $\mathcal{M}^N$: For each m in $\mathcal{M}$, let $N_m(s)$ be the number of indices j for which $m = m_j$; that is, the number of molecules at lattice point m when the macroscopic state is s. The functions N_m are random variables on the probability space $\mathcal{M}^N$, α_T, and these functions separate the points of $\mathcal{M}^N$. Thus using the mapping $s \to \omega$, where ω is the nonnegative integer-valued function on $\mathcal{M}$ taking m into $N_m(s)$, one may identify $\mathcal{M}^N$ with a subset of the space $(Z^+)^{\mathcal{M}}$ of all functions from $\mathcal{M}$ to the nonnegative integers. Since extra points do no harm if they are collectively of measure zero, one can replace $\mathcal{M}^N$ with the whole of $(Z^+)^{\mathcal{M}}$ and so obtain a configuration space which is the same for all N. Actually, it is customary to confine attention to the case in which it is impossible for two molecules to occupy the same lattice point so that Z^+ may be replaced by the two-element set F_2 consisting of 0 and 1. The system is completely defined then by giving $\mathcal{M}$ and the function W on $(F_2)^{\mathcal{M}}$ or, (almost) equivalently, by giving $\mathcal{M}$ and for any $T > 0$ the Gibbs measure μ_T which assigns the measure $e^{-W(\omega)/kT}/\sum_{\omega\in(F_2)^{\mathcal{M}}} e^{-W(\omega)/kT}$ to each one-point set $\{\omega\} \subseteq (F_2)^{\mathcal{M}}$. Note that changing W by adding a

constant to it leaves all the μ_T unaltered so that W is determined by the μ_T only up to an (inessential) additive constant.

In order to compare one set of lattice points with another and talk of an infinite limit, it is necessary to restrict W. As already mentioned in Section 9, physical experience suggests that it will suffice to consider sums of "two-body potentials". In other words, it will be assumed until further notice that there exists a real-valued function w defined on the additive group of *all* lattice points in space such that $w(-m) = w(m)$, $w(0) = 0$

$$W(\omega) = 1/2 \sum_{\substack{m_1 \in \mathscr{M} \\ m_2 \in \mathscr{M} \\ \omega(m_1)=\omega(m_2)=1}} w(m_1 - m_2) \qquad \text{whenever} \quad \sum_{m \in \mathscr{M}} \omega(m) = N,$$

and $W(\omega) = \infty$ whenever $\sum_{m \in \mathscr{M}} \omega(m) \neq N$. The function w will be called the interaction function. Let $|\mathscr{M}|$ denote the number of lattice points in $\mathscr{M}$. Given w and $\rho = N/|\mathscr{M}|$, the formula just written down assigns a unique energy function W to every $(F_2)^{\mathscr{M}}$ for which $|\mathscr{M}|\rho$ is an integer.

Whatever else is involved in defining the infinite volume limit of the systems specified by a fixed interaction function w and density ρ, it is clear that the configuration space must be $(F_2)^{Z^3}$ and that the macroscopic state of the system at temperature T must be defined by a probability measure μ_T in this configuration space. The problem is to generalize the construction of the μ_T in the $(F_2)^{\mathscr{M}}$ for finite $\mathscr{M}$ so that it makes sense in the infinite case and, correspondingly, to find appropriately generalized definitions of free energy, entropy, etc. Of course, an infinite homogeneous system will have infinite entropy, infinite free energy, etc., and one seeks instead to define entropy per lattice site, free energy per lattice site, etc. This circumstance is one of the difficulties standing in the way of a straightforward generalization from finite to infinite $\mathscr{M}$. Another difficulty lies in the fact that $(F_2)^{Z^3}$ is not a discrete space and that there is *no* measure with respect to which the μ_T are all absolutely continuous.

However one goes about the task of overcoming these difficulties, it is clear from homogeneity that each μ_T must be invariant with respect to the natural action of Z^3 on $(F_2)^{Z^3}$ so that (except for the replacement of Z by Z^3) the system consisting of the configurations space and the macroscopic state is identical with that defining a discrete stationary stochastic process in the special case of random variables which take on only the two values 0 and 1. Put slightly differently, each macroscopic

(equilibrium) state of an infinite lattice gas is canonically associated with a three-dimensional discrete parameter chaos whose random variables take on only the values 0 and 1. This suggests that the methods of ergodic theory may be applied to the problem and also raises the question of possibly finding analogues of thermodynamic concepts in other probabilistic contexts.

One application of ergodic theory is almost immediate and arises out of the attempt to assign a "density" to each microscopic state of our infinite lattice gas; that is, to each point of $(F_2)^{Z^3}$. An obvious way to do this is to consider for each $l = 1, 2, 3,...$, the set $\mathscr{M}_l$ of all n_1, n_2, n_3 in Z^3 with $-l \leqslant n_j \leqslant l$ for $j = 1, 2, 3$. Then $\sum_{m \in \mathscr{M}_l} \omega(m)$ will be the number of gas molecules in $\mathscr{M}_l$ when the state is ω and the average "density" or "number per lattice site" will be

$$\Big(\sum_{m \in \mathscr{M}_l} \omega(m)\Big) \Big/ |\mathscr{M}_l| = \Big(\sum_{m \in \mathscr{M}_l} \omega(m)\Big) \Big/ (2l+1)^3.$$

If this expression has a definite limit $\rho(\omega)$, as l tends to ∞, it is natural to define $\rho(\omega)$ to be the gas density in the microscopic state ω. Now if the gas were one instead of three dimensional, the ergodic theorem applied to the function $\omega \to \omega(0)$ would at once imply that the limit does indeed exist for μ almost all ω for every invariant probability measure μ in $(F_2)^Z$. In other words, the (Borel) set of all ω for which the limit does not exist is of measure zero simultaneously for all invariant probability measures and hence can be neglected once and for all. A corresponding result for the three-dimensional case would follow at once if the obvious generalization of the ergodic theorem from Z to Z^3 could be proved. Such a generalization (from Z to Z^n) was proved by Wiener in 1939. Thus, $\rho(\omega)$ is a well-defined Borel function on $(F_2)^{Z^3}$ except on a subset which is of measure zero for all possible μ_T. Evidently, the density ρ_0 to be assigned to any particular macroscopic state μ_T will be the expected value $\int \rho(\omega)\, d\mu_T(\omega)$ of the random variable ρ. Notice that if μ_T is *ergodic* (and we shall see that it is in many interesting cases), then $\rho(\omega)$, being invariant under the Z^3 action, will be equal to its expected value for μ_T almost all ω.

The three-dimensional ergodic theorem may be used in a similar manner to associate an internal energy per lattice site $\overline{W}(\omega)$ to almost every ω. However, the exceptional "universal" null set (and of course $\overline{W}$ as well) will depend upon the interaction function w; and $\overline{W}$ will only exist if w is suitably restricted. In most of the literature, w is assumed to

be zero outside of a small finite set, but it suffices for the present purposes to suppose that $\sum_{x\in Z^3} |\, w(x)| < \infty$. Then when the gas is in the microscopic state ω, for each $m \in Z^3$ the contribution to the potential energy of the interaction of a molecule at m with all other molecules will be $\omega(m)(\sum_{m'\in Z^3} w(m)'\, \omega(m + m'))$. The infinite sum clearly converges and defines a Borel function $\omega \to E_m(\omega)$, and it is evident that $E_m(\omega) = E_0([\omega]m)$ where $[\omega]m$ is the translate of ω by m. Applying the three-dimensional ergodic theorem to E_0, one finds that $\lim_{l\to\infty} \sum_{m\in\mathscr{M}_l} E_m(\omega)/|\,\mathscr{M}_l\,|$ exists except for a set of ω's which is of measure zero with respect to all invariant probability measures in $(F_2)^{Z^3}$. Now $\sum_{m\in\mathscr{M}_l} E_m(\omega)$ is *twice* the contribution to the potential energy of all pairs m_1, m_2 with m_1 and m_2 both in $\mathscr{M}_l$ plus the contribution to the potential energy of all pairs m_1, m_2 with m_1 in $\mathscr{M}_l$ and $\mathscr{M}_2$ not in $\mathscr{M}_l$. Moreover, an easy argument shows that the contribution of the pairs of the second kind divided by $|\,\mathscr{M}_l\,|$ goes to zero as l approaches infinity. Thus, defining $\overline{W}(\omega) = 1/2 \lim_{l\to\infty} \sum_{m\in\mathscr{M}_l} E_m(\omega)/|\,\mathscr{M}_l\,|$, one obtains a Borel function defined except on a set which is of measure zero for every possible μ_T and which may be interpreted as the internal energy per lattice site when the gas is in the microscopic state ω. Correspondingly, in the macroscopic state defined by an invariant probability measure μ, the internal energy of the gas will be $\int \overline{W}(\omega)\, d\mu(\omega)$.

At this stage, the entropy per lattice site is best thought of as a property of macroscopic states rather than of almost all microscopic states. (However, as will be shown below, one can assign an entropy to almost all microscopic states as well.) Recall that when $\mathscr{M}$ is finite, the entropy of the macroscopic state defined by the probability measure μ in $(F_2)^{\mathscr{M}}$ is $S(\mu) = -\sum_{\omega\in(F_2)^{\mathscr{M}}} \mu(\{\omega\}) \log \mu(\{\omega\})$. Now each ω in $(F_2)^{Z^3}$ is a function on Z^3, and by restriction ot $\mathscr{M}_l$, defines a member $\theta_l(\omega)$ of $(F_2)^{\mathscr{M}_l}$ for each $l = 1, 2, \ldots$. The function θ_l maps $(F_2)^{Z^3}$ onto $(F_2)^{\mathscr{M}_l}$ and associates a probability measure μ_l in $(F_2)^{\mathscr{M}_l}$ to each probability measure μ in $(F_2)^{Z^3}$: $\mu_l(A) = \mu(\theta_l^{-1}(A))$. The probability measure μ_l describes the macroscopic state of that part of the gas whose molecules lie in $\mathscr{M}_l$ when μ describes the macroscopic state of the infinite gas. Now $S(\mu_l)$ is well-defined by the formula given above and one may think of $S(\mu_l)/|\,\mathscr{M}_l\,|$ as the entropy per lattice site for that part of the gas. This suggests that it should be possible to prove that $\lim_{l\to\infty} S(\mu_l)/|\,\mathscr{M}_l\,|$ exists and that this limit is the appropriate definition of entropy per lattice site for the infinite gas. The analogous question with Z^3 replaced by Z (and F_2 replaced by any finite set) arises in information theory (cf. Section 11), and it turns out that in that case, the limit exists and is finite

for any invariant probability measure μ. Khinchin has given a short elementary proof of this important fact on pages 47–49 of his book "Mathematical Foundations of Information Theory" (1957). Since Khinchin's argument adapts easily to the n-dimensional case, it follows that $\lim_{l\to\infty} S(\mu_l)/|\mathscr{M}_l|$ exists and is finite; and one has a well-defined "entropy per lattice site" $\bar{S}(\mu)$ for every invariant probability measure μ in $(F_2)^{Z^3}$.

There remains the problem of defining the particular invariant measure μ_{T,ρ_0} which describes the macroscopic state when the temperature is T and the density (in molecules per lattice site) is ρ_0. A natural approach to this problem proceeds (as in the case of entropy) through a study of the measures $(\mu_{T,\rho_0})_l$ in the finite space $(F_2)^{\mathscr{M}_l}$ obtained from μ_{T,ρ_0} by the mappings θ_l. It is easy to see that μ_{T,ρ_0} is uniquely determined by the $(\mu_{T,\rho_0})_l$ and can be described by describing the latter. One's first thought is that $(\mu_{T,\rho_0})_l$ must be just the Gibbs measure μ_T for $\mathscr{M}_l$ with $N = \rho_0(2l+1)$. However, this is wrong—even when ρ_0 is such that N is always an integer. The probability measures in the $\mathscr{M}_l$ that one gets in this way are inconsistent in the sense that the natural map of $\mathscr{M}_{l+l'}$ onto $\mathscr{M}_l$ does not map one into the other. This is because a portion of an infinite gas in equilibrium with the rest of the gas is not the same as that same portion when isolated from the rest. In particular, the number of molecules can vary.

To overcome this and other difficulties, one must do several things. First of all, in considering finite systems, one must give up the idea of a fixed number of particles. Instead of fixing the number of particles (or the number per lattice site), one must fix its expected value. The Gibbs measure μ_T which assigns a probability proportional to $e^{-W(\omega)/kT}$ to each ω with $\sum \omega(m) = N$ and zero to all other ω must be replaced by a probability measure which assigns a probability proportional to $e^{-[W(\omega)-cN(\omega)]/kT}$ to each ω in $(F_2)^{\mathscr{M}}$, where $N(\omega) = \sum_{m\in\mathscr{M}} \omega(m)$ and c is a real number chosen so as to make the expected value of $N(\omega)$ take on the desired value. (Note that this may now be nonintegral). One is led to an expression of this form if one attempts to maximize the "spreadoutness" of μ subject to the conditions that $W(\omega)$ and $N(\omega)$ have fixed expected values. The parameter c is called the *chemical potential* and c/T is related to the particle number much as temperature is related to energy. In classical statistical mechanics, replacing the Gibbs measures μ_T by the two-parameter family $\mu_{T,c}$ just defined is referred to as replacing the Gibbs "canonical ensemble" by the Gibbs "grand canonical ensemble". In cases which occur in practice, the probability distribution

which $\mu_{T,c}$ assigns to N is highly concentrated about its expected value. Thus, whether one uses $\mu_{T,c}$ or the μ_T corresponding to $\int N(\omega)\, d\mu_{T,c}(\omega)$ makes very little difference to computed results. However, the "grand" Gibbs measure $\mu_{T,c}$ has theoretical advantages.

Given an interaction function w, one has a well-defined family of grand Gibbs measures $\mu^l_{T,c}$ in each $F_2^{\mathcal{M}_l}$. Now fix l and consider the natural mapping $\theta^{l'}_l$ of $(F_2)^{\mathcal{M}_{l+l'}}$ on $(F_2)^{\mathcal{M}_l}$ for each $l' = 1, 2, \ldots$. Via the usual formula $\tilde{\mu}(A) = \mu(\theta^{-1}(A))$, these mappings carry the $\mu^{l'}_{T,c}$ into probability measures $\tilde{\mu}^{l',l}_{T,c}$ in $(F_2)^{\mathcal{M}_l}$, and, the larger l' is, the more nearly the $\tilde{\mu}^{l',l}_{T,c}$ are like the images in $(F_2)^{\mathcal{M}_l}$ of measures in $(F_2)^{Z^3}$. Moreover, it is easy to see that if $\lim_{l\to\infty} \tilde{\mu}^{l',l}_{T,c}$ exists for all l, T, and c, then the limiting measures $\tilde{\mu}^{\infty,l}_{T,c}$ are consistent and there exists a unique probability measure $\tilde{\mu}_{T,c}$ in $(F_2)^{Z^3}$ whose images in the $(F_2)^{\mathcal{M}_l}$ are the $\tilde{\mu}^{\infty,l}_{T,c}$. The measure $\tilde{\mu}_{T,c}$ (when it exists) is the (grand) Gibbs measure for the infinite lattice gas at temperature T and chemical potential c. To find the Gibbs measure μ_T for the gas when the density is ρ_0 , one chooses c_0 so that

$$\int \bar{\rho}(\omega)\, d\tilde{\mu}_{T,c_0}(\omega) = \rho_0$$

and sets $\mu_T = \tilde{\mu}_{T,c_0}$. Of course, it has to be proved that c_0 is unique.

While physical intuition suggests that the limiting measures $\tilde{\mu}^{\infty,l}_{T,c}$ indeed exist at least whenever w is a "physically realistic" interaction function, the mathematical question remains of proving the existence under suitable hypotheses on w. Moreover, the interesting question to the practicing physicist is that of finding explicit procedures for computing the values of $\tilde{\mu}_{T,c}$ and the integrals $\int \overline{W}(\omega)\, d\tilde{\mu}_{T,c}(\omega)$ and $\int \bar{S}(\omega)\, d\tilde{\mu}_{T,c}(\omega)$ for a given interaction function w. The simplest case of course is that in which $w \equiv 0$ so that the gas molecules do not interact at all. Here, an easy calculation shows that $\mu^{l',l}_{T,c}$ is independent of l' so that the limit $\tilde{\mu}^{\infty,l}_{T,c}$ exists trivially. The corresponding probability measure $\tilde{\mu}_{T,c}$ in $(F_2)^{Z^3}$ is a direct product over the lattice sites of that probability measure in F_2 which assigns the probability $e^{c/kT}/1 + e^{c/kT}$ to 1 and the probability $1/1 + e^{c/kT}$ to 0. The functions $\omega \to \omega(m)$ are thus independent random variables with a distribution depending on c/kT, and the density of the gas is $e^{c/kT}/1 + e^{c/kT}$ molecules per lattice site. Thus the infinite system limit of a "perfect" lattice gas in any given macroscopic state is a three-dimensional analogue of a stochastic process with two-valued independent identically distributed random variables. In particular, the invariant probability measures $\tilde{\mu}_{T,c}$ are all ergodic. This ergodicity is in sharp contrast to the behavior of the ζ_E under the time

translation group as discussed in the last section. Here, although one hopes to prove ergodicity for suitable interactions, it is easy to see that it fails for a perfect gas; that is, when there is no interaction. In general, the ergodicity of the $\tilde{\mu}_{T,c}$ is much easier to come by than that of the ζ_E.

Another interesting special case that can be explicitly and exactly worked out is that of a *one-dimensional* lattice gas for which the interaction w is zero except at ± 1. This means that two gas molecules interact only when they are on adjacent lattice sites and one speaks of a "nearest neighbor" interaction. In this case, one can show that, for any choice of $A = w(1) = w(-1)$ the measures $\tilde{\mu}_{T,c}^{l',l}$ converge for all T and c as l' tends to ∞ and that the limiting measure defines a stationary stochastic process whose underlying action is ergodic. It is interesting that the stationary process in question is in a sense the next simplest kind after the processes whose random variables are independent. Given a discrete two-valued process and a finite sequence $a_1, a_2, \ldots, a_{n-1}$ of zeros and ones, one can ask for the probability that $\omega(n)$ will be one when it is known that $\omega(1) = a_1$, $\omega(2) = a_2 \cdots \omega(n-1) = a_{n-1}$. When the random variables are independent, this probability is independent of $a_1, a_2, \ldots, a_{n-1}$ and n. By definition, the process is a *Markov process* if this probability is independent of $a_1, a_2, \ldots, a_{n-2}$ and n but depends on a_{n-1}. A two-valued Markov process is uniquely determined by its matrix of transition probabilities $\begin{pmatrix} p & q \\ 1-p & 1-q \end{pmatrix}$, where p is the probability that $\omega(n)$ will be zero when $\omega(n-1)$ is zero and q is the probability that $\omega(n)$ will be one when $\omega(n-1)$ is one. Of course, when $p = 1 - q$, the random variables are independent so that $|p + q - 1|$ measures the degree of dependence. The measures $\tilde{\mu}_{T,c}$ in $(F_2)^Z$ defined by a one-dimensional lattice gas with $w(1) = w(-1) = A$ define stationary Markov processes whose transition probabilities p and q are the unique real numbers on the interval $0 < x < 1$ which satisfy the equations:

$$(p/(1-p))(q/(1-q)) = e^{-A/kT} \qquad \text{and} \qquad q/p = e^{(c-A)/kT}.$$

To understand the relationship between two-valued Markov processes and one-dimensional nearest-neighbor lattice gases including the above formula for p and q, it suffices to compare the formula for computing the probability of a fixed finite configuration on a Markov process with the corresponding formula for a finite portion of a lattice gas. Consider the expression

$$P(a, b) = p^{(1-a)(1-b)}q^{ab}(1-p)^{(1-a)b}(1-q)^{a(1-b)},$$

where a and b are 0 or 1. Evidently,

$$P(1, 1) = q, \quad P(0, 0) = p \quad P(1, 0) = 1 - q, \quad \text{and} \quad P(0, 1) = 1 - p$$

so that $P(a, b)$ is the probability that the $k + 1$-st random variable will take on the value b given that the kth took on the value a. Moreover, $P(a, b)$ may clearly be written in the form $e^{\lambda_1 ab + \lambda_2 a + \lambda_3 b + \lambda_4}$, where

$$\lambda_1 = \log(pq/(1-p)(1-q)), \qquad \lambda_2 = \log((1-q)/p),$$
$$\lambda_3 = \log((1-p)/p), \qquad \text{and} \qquad \lambda_4 = \log p.$$

One simply sets $p = e^{\log p}$, $q = e^{\log q}$, etc., and rearranges the result in an obvious way. It follows that the probability that $b_1, b_2, \ldots, b_r$ will occur in that order immediately after b_0 has occurred is just

$$P(b_0, b_1)\, P(b_1, b_2) \cdots P(b_{r-1}, b_r)$$
$$= e^{r\lambda_4} e^{\lambda_1(b_0 b_1 + b_1 b_2 + \cdots b_{r-1} b_r) + \lambda_2(b_0 + b_1 + \cdots b_{r-1}) + \lambda_3(b_1 + b_2 + \cdots + b_r)}.$$

In particular, if ω is an arbitrary function from $-j \leqslant k \leqslant j$ to $\{0, 1\}$, then the probability that the sequence $\omega(-j), \omega(-j+1), \ldots, \omega(j)$ occurs given that $\omega(-j-1)$ was zero is

$$e^{\lambda_4(2j+2)} e^{\lambda_1(\sum_{k=-j}^{j-1} \omega(k)\omega(k+1)) + (\lambda_2+\lambda_3)(\sum_{k=-j}^{j} \omega(k)) - \lambda_2 \omega(j)}.$$

Except for the term $-\lambda_2\omega(j)$ (which becomes relatively less and less important as j increases), the probabilities in questions as functions of ω are proportional to

$$e^{\lambda_1 \sum_{k=-j}^{j-1} \omega(k)\omega(k+1) + (\lambda_2+\lambda_3)\sum_{k=-j}^{j} \omega(k)}.$$

The resemblance of this formula to that provided by the Gibbs measure is striking indeed:

$$W(\omega) = A \sum_{k=-j}^{j-1} \omega(k)\, \omega(k+1) \qquad \text{and} \qquad N(\omega) = \sum_{k=j}^{j} \omega(k)$$

so

$$\frac{-W(\omega) - cN(\omega)}{kT} = -\frac{A}{kT} \sum_{k=-j}^{j-1} \omega(k)\, \omega(k+1) + \frac{c}{kT} \sum_{k=-j}^{j} \omega(k).$$

Thus the formulas coincide if $-A/kT = \lambda_1$ and $c/kT = \lambda_2 + \lambda_3$; that is,

if $\log(pq/(1-p)(1-q)) = -A/kT$ and $\log((1-q)(1-p)/p^2) = c/kT$. Exponentiating both sides and multiplying the resulting equations together leads at once to the relations between p, q, A, c, and T announced above. Of course, when $A = 0$, $p = 1 - q$ and one is back to the case of independent random variables.

In the general case, it is possible to prove that the $\tilde{\mu}^{\infty,1}_{T,c}$ exist and define an ergodic invariant measure in $(F_2)^{Z^3}$ provided that $1/T$ and c/T are sufficiently small; that is, provided that the gas is sufficiently near to being perfect. Existence can be proved under much broader conditions but as things stand at the moment, some of the restrictions are rather artificial and presumably have more to do with inadequate methods of proof than with the mathematical facts. On the other hand, one does not expect ergodicity except at high temperatures and low densities because of the phenomenon of condensation from the gas to the liquid state and other so-called "phase transitions". The idea is that the system behaves like a gas when T and c are such that ergodicity does hold and that failure of ergodicity occurs just when a phase transition sets in. The nature of the separation into ergodic parts is no doubt related to the character of the phases that result, but the precise nature of the relationship is still obscure. Traditionally, one has sought to correlate the occurrence of phase transitions with discontinuities and other singularities in the thermodynamic limit of the free energy and other thermodynamic functions. Failure of ergodicity seems to go more deeply into the essence of the matter and this is one reason for being interested in infinite system limits.

When conditions are such that $\mu_{T,c}$ is ergodic, one has a very clear sense in which the macroscopic state defined by $\mu_{T,c}$ is uniquely determined by any one of a large number of microscopic states (i.e., points ω in $(F_2)^{Z^3}$). In fact, the relationship between macroscopic states and microscopic ones is completely analogous to that between stationary stochastic processes and their sample functions. In particular, $\mu_{T,c}$ can be uniquely reconstructed from $\mu_{T,c}$ almost any point ω in $(F_2)^{Z^3}$ by an obvious adaptation of the argument given at the end of Section 5. In other words (barring events of probability zero), it suffices to know the actual distribution of molecules on lattice sites in an (infinite) sample of the gas in order to know the associated $\mu_{T,c}$ and hence all thermodynamic quantities including the entropy. Of course, many different distributions will give the same results. The fact that the entropy of $\mu_{T,c}$ can be computed from $\mu_{T,c}$ almost any ω allows one to assign a unique entropy to each member of a family Ω of points ω of $(F_2)^{Z^3}$ which is such that

$\mu((F_2)^{Z^3} - \Omega)$ is zero for all invariant probability measures μ. Thus, as promised earlier, entropy can be regarded as a function defined on the points of $(F_2)^{Z^3}$ as well as one defined on the probability measures on $(F_2)^{Z^3}$.

The relationship between the infinite system limit of a gas and an actual gas is of course analogous to the relationship between an actual finite time series in statistics and the hypothetical stationary process of which it is a segment. However, the number of molecules in an ordinary macroscopic piece of matter is so enormously large that the approximation to the hypothetical infinite system is much much better than the corresponding approximation in statistics.

In the very special case of a one-dimensional lattice gas with nearest-neighbor interaction, the problem of finding explicit procedures for computing the values of $\tilde{\mu}_{T,c}$ and the associated thermodynamic functions reduces to problems about two-valued stationary Markov processes in probability theory. More general interactions in the one-dimensional case can be studied in a similar fashion using n-valued Markov processes. In higher dimensions, the problem becomes much more difficult. For the two-dimensional nearest-neighbor lattice gas (with c suitably related to the two interaction constants), L. Onsager created a sensation in 1944 by finding an exact formula for the thermodynamic limit of the free energy. No one has ever been able to find such an exact formula in the three-dimensional case or to eliminate the restriction on c in Onsager's result. On the other hand, much work has been done (and is being done) in simplifying Onsager's difficult and ingenious arguments and in using similar methods to deal with other two-dimensional problems. One can define higher-dimensional Markov processes and relate them to higher-dimensional nearest-neighbor lattice gases in a manner that has been analyzed by Dobrushin, Averintsev, and Spitzer. However, the theory of higher-dimensional Markov processes is both much more difficult and much less developed than the one-dimensional theory. In other words, the infinite system approach to statistical mechanics suggests a natural, interesting, and difficult extension of the classical theory of Markov processes which is still in a primitive state of development.

Of course, physicists do not insist on exact solutions of their problems and the literature of statistical mechanics is filled with descriptions of more or less successful approximate methods for treating the problems outlined above. In particular, series expansions have been worked out which are valid for quite general interactions as long as the temperature is high or the density low. It would be interesting to examine the implica-

tions for ergodic theory of some of these methods in view of the connection between ergodic theory and the underlying problems outlined above.

We have concentrated on lattice gases because of their greater technical simplicity. One can develop an analogous theory for the physically more realistic continuum models but we shall not give details here. Suffice it to say that the discrete family of two-valued random variables defined on Z^3 must be replaced by a homogeneous, nonnegative, integer-valued chaos defined on the Borel subsets of Euclidean space and that the theory of a one-dimensional *perfect* gas is equivalent to the classical theory of the Poisson chaos.

As indicated earlier, the results of classical statistical mechanics are in accord with experiment only at relatively high temperatures. At lower temperatures, quantum effects become manifest and one has to replace classical statistical mechanics by a new theory based on the laws of quantum mechanics instead of those of classical mechanics. The relationship between classical and quantum statistical mechanics is best appreciated by thinking of classical phase space Ω not as a point set but as the orthocomplemented partially ordered set $\mathscr{L}_\Omega$ of all of its Borel subsets. Its analogue in quantum mechanics is the orthocomplemented partially ordered set $\mathscr{L}_\mathscr{H}$ of all closed subspaces of some separable complex Hilbert space. One can define measures and probability measures on any orthocomplemented partially ordered set and the analogue of the Liouville measure on $\mathscr{L}_\Omega$ is the measure $M \to \dim M$. The analogue of a real-valued Borel function H on Ω is a self-adjoint operator A in the Hilbert space $\mathscr{H}$, and in the case in which A has a point spectrum, the analogue of the set mapping $E \to H^{-1}(E)$ is $E \to M_E$, where M_E is the closed linear span of all eigenvectors of A whose eigenvalues are in E. (More generally, M_E is the range of the projection $P_E{}^A$, where $E \to P_E{}^A$ is the projection-valued measure associated with A by the spectral theorem.) Thus the analogue of the measure $E \to \beta(E) = \zeta(H^{-1}(E))$ on the real line, which defines the classical partition function via the formula

$$p(T) = \int e^{-x/kT}\, d\beta(x),$$

is the measure $E \to \dim(M_E) = \beta_q(E)$. Thus the quantum mechanical partition function p_q is defined by

$$p_q(T) = \int e^{-x/kT}\, d\beta_q(x) = \sum_{j=1}^{\infty} e^{-E_j/kT} m_j = \operatorname{Tr} e^{-A/kT}$$

whenever A has a pure point spectrum. Here, the E_j are the distinct eigenvalues of A and m_j is the multiplicity with which E_j occurs. Just as $\zeta(H^{-1}(E))$ is finite when H is the Hamiltonian function of a gas in a finite volume, so A has a pure point spectrum and $e^{-A/kT}$ a finite trace when A is the Hamiltonian operator for such a gas. The analogue of the Gibbs measure $e^{-H/kT}\, d\zeta/p(T)$ is the probability measure in $\mathscr{L}_{\mathscr{H}}$; $M \to \mathrm{Tr}(e^{-A/kT}P_M)/p_q(T)$, where P_M is the projection operator whose range is M. The operator $e^{-A/kT}/p_q(T) = e^{-A/kT}/\mathrm{Tr}(e^{-A/kT})$ is called the von Neumann density operator.

In sum, quantum statistical mechanics differs from classical statistical mechanics in replacing the continuous measure β by the atomic measure β_q. It turns out that $\beta_q([0, x])$ is asymptotically a constant multiple of $\beta([0, x])$ for large x. Using this and the elementary theory of the Laplace transform, it follows that the classical and quantum mechanical free-energy functions agree at high temperatures.

In order to have a quantum analogue of an infinite lattice gas, one needs to define an infinite product of orthocomplemented partially ordered sets of the form $\mathscr{L}_{\mathscr{H}}$ and this is not easy to do. Tensor products are not well-defined for partially ordered sets, and while one can consistently think of $\mathscr{L}_{\mathscr{H}_1} \otimes \mathscr{L}_{\mathscr{H}_2}$ as $\mathscr{L}_{\mathscr{H}_1 \otimes \mathscr{H}_2}$, there are difficulties with infinite products connected with the fact that $\mathscr{H}_1 \otimes \mathscr{H}_2 \otimes \mathscr{H}_3 \cdots$ does not have a clear and unambiguous meaning. There seems to be no analogue of the fact that a countable product of standard Borel spaces is again such. The situation is even worse when one tries to do infinite limit quantum statistical mechanics with gas molecules moving in a continuum. Recent work of Araki, Woods, Streater, Parthasarathy, and Schmidt connecting continuous tensor products of Hilbert spaces with "infinitely divisible" probability distributions may be regarded as contributions to the problem of dealing with the continuum case. In the approach taken by the French school and treated in Ruelle's book, one replaces $\mathscr{L}_{\mathscr{H}}$ by $\mathscr{B}(\mathscr{H})$, the algebra of all bounded operators on $\mathscr{H}$, and then uses a C^* algebra which is an inductive limit of C^* algebras of the form $\mathscr{B}(\mathscr{H})$ is a substitute for an infinite product of $\mathscr{L}_{\mathscr{H}}$'s. Their method can be used in the classical case as well and it may throw some light on what is going on to recast the lattice gas set up in C^* algebra language. The standard Borel space $(F_2)^{Z^3}$ is a product of infinitely many replicas of F_2 and as such is a compact Hausdorff space. The C^* algebra in this case is the algebra $\mathscr{A}$ of all continuous complex-valued functions on $(F_2)^{Z^3}$, and the action of Z^3 on $(F_2)^{Z^3}$ defines and is defined by a homomorphism of Z^3 into the group of automorphisms of $\mathscr{A}$. Each

probability measure μ on $(F_2)^{Z^3}$ defines a linear functional l_μ on $\mathcal{A}$ vie $l_\mu(f) = \int f(\omega)\, d\mu(\omega)$, and the l_μ are precisely the linear functionals l such that $l(ff^-) \geqslant 0$ and $l(1) = 1$. The invariant probability measures on $(L_2)^{Z^3}$ thus correspond one-to-one to the l_μ which are invariant under the automorphisms of $\mathcal{A}$ defined by the members of Z^3.

In the quantum case, $\mathcal{A}$ is a noncommutative C^* algebra but one still has a homomorphism of Z^3 (the Euclidean group in the continuum case) into the group of automorphisms of $\mathcal{A}$ and the macroscopic states of the system are described by invariant linear functionals l such that $l(ff^*) \geqslant 0$ and $l(1) = 1$. This suggests the desirability of developing a sort of "noncommutative ergodic theory" in which one begins with a homomorphism of a group G into the group of automorphisms of a noncommutative C^* algebra $\mathcal{A}$ instead of a standard Borel G space and studies invariant nonnegative linear functionals instead of invariant measures. Such a theory has been in the course of development since 1966, beginning with a paper of Kastler, Doplicher, and Robinson. The idea of using C^* algebras to study systems with infinitely many degrees of freedom stems from work of I. E. Segal beginning in the 1940's.

11. Entropy, Information Theory, and the Classification of Ergodic Actions

As indicated in the last section, a one-dimensional lattice gas in the infinite volume limit is mathematically equivalent to a family of two-valued stationary stochastic processes—there being one member for each value of the temperature T and the chemical potential c. The entropy (per lattice site) of the lattice gas at any T and c depends only upon the probability measure in $(F_2)^Z$ defining the process, and the definition makes sense for any invariant probability measure in $(F_2)^Z$. Thus one can speak of the entropy of any two-valued stationary stochastic process. Moreover, the restriction to two-valuedness is quite inessential. There is no difficulty in working out a theory of lattice gases with several kinds of molecules, and a lattice gas with n molecule types is related to an $(n + 1)$-valued stationary stochastic process in just the way that a gas with a single molecule type is related to a two-valued process. Thus every finite-valued stationary stochastic process has a well-defined entropy which is a nonnegative real number.

While we have chosen to define the entropy of a process in such a way as to emphasize its essential identity with the classical entropy of thermo-

dynamics, it can also be defined directly and the entropy of a process was so defined by Claude Shannon in the fundamental memoir with which he founded the subject of information theory in 1948. While Shannon was of course aware of the analogy of his notion with the entropy of thermodynamics and statistical mechanics, the theory of infinite volume limits of lattice gases was almost 20 years in the future when his paper appeared. The direct definition is an obvious adaptation of the definition for a lattice gas given in Section 10. If F is the finite set of values taken on by the process, one considers for each $N = 1, 2,...$ and each function f from $\{1, 2,..., N\}$ to F the probability that the sequence $f(1), f(2),..., f(N)$ will occur. Call this $p_{f,N}$. One shows that $\lim_{N\to\infty} -\sum_f p_{f,N} \log p_{f,N}/N$ exists an defines this limit to be the entropy of the process.

From the point of view of stochastic process theory, the entropy notion has a fairly straightforward interpretation. It represents the rate at which uncertainty is removed (or information gained) as one observes the actual values of the random variables one after the other. *A priori*, the sequence $f(1), f(2),..., f(N)$ could have any value in $(F)^N$ so that before one has made any observations, there is a degree of uncertainty about what will happen which (as explained in Section 9) is measured by the number $-\sum_f p_{f,N} \log p_{f,N}$. When the N observations have been made, all this uncertainty is gone and one has received a corresponding amount of information. The amount per observation is $-\sum_f p_{f,N} \log p_{f,N}/N$, and the entropy H is just the limit of this amount per observation as $N \to \infty$.

Let $H_N = -\sum_f p_{f,N} \log p_{f,N}$. A simple calculation shows that $H_N = NH_1$ wherever the random variables are independent and that in any case $H_N \leqslant NH_1$. Hence the limit H_N/N obviously exists and equals H_1 in the independent case and the general case $H \leqslant H_1$. Clearly, $(H_1 - H)/H_1$ is a sort of measure of the degree of dependence which takes on its maximum value of one when the entropy is zero. In this connection, it is interesting to look at the entropy notion from the point of view of general prediction theory. As explained early, one has an (almost everywhere defined) mapping $q \to \alpha_q$ of one-sided infinite sequences $q = \cdots s_{-2}, s_{-1}, s_0$ of elements of F into probability measures α_q in F such that $\alpha_q(s)$ is the probability of getting s immediately after an infinite past represented by $q = \cdots s_{-2}, s_{-1}, s_0$. The process is deterministic or perfectly predictable if and only if α_q is concentrated in a point for $\tilde{\mu}$ almost all q, where $\tilde{\mu}$ is the image in the space of all pasts of the invariant measure μ in F^Z which describes the process. But to say that α_q is concentrated in a point for almost all q is precisely to say that $\sum_{s\in F} \alpha_q(s) \log \alpha_q(s) = 0$ for almost all q. Moreover, it can be proved

that the entropy H of a process is equal to the expected value $-\int \sum_{s \in F} \alpha_q(s) \log \alpha_q(s)\, d\tilde{\mu}(q)$ of the degree of uncertainty in α_q. Hence the entropy is zero if and only if the process is deterministic.

Shannon introduced the entropy concept as an essential tool in developing a mathematical theory of the efficient "coding" of messages. As everyone knows, ordinary language is redundant in the sense that if one alters or omits letters at random (but not too frequently), one can guess what was meant. Obviously, one could communicate the same message with fewer symbols by eliminating such redundancy. However, if one eliminates *all* redundancy, the smallest error will completely destroy the sense of the message, and since errors are impossible to eliminate, one has lost more than one has gained. Since one can obviously have much more redundancy than is necessary to guard against error, the question arises of just how much redundancy is optimal. How does one maximize efficiency of transmission while keeping the probability of misunderstanding at a tolerable level? Shannon's "information theory" attacks and solves this problem.

Before describing Shannon's solution, it is necessary to explain the sense in which one can think of a message as (a segment of) a sample functions of a stationary stochastic process. As mentioned in Section 7 in relating Wiener's generalized harmonic analysis to the harmonic analysis of sample functions, the essential identity of messages and sample functions is a notion which goes back to early work of Wiener and one which Wiener did much to develop and emphasize. While Wiener was concerned with continuous-parameter stochastic processes, the basic idea is the same and is technically easier to explain in the discrete case.

Consider a page of English text. It consists of a sequence of letters, spaces, and punctuation signs. If one counts the number of occurrences of any particular letter and divides by the total number of symbols, one obtains a number which is very nearly the same for one page as it is for another and in any event does not fluctuate any more than one would expect of different measurements of sample means in probability theory. The same is true if one counts two-letter sequences, three-letter sequences, etc., so long as the length of the sequences considered is short compared to the total number of symbols in the text. Actually, one finds that one can confine oneself to short sequences as the probability of getting a particular letter tends to be independent of all but a few letters immediately preceding it. Once the frequencies of occurrence have been determined, one can construct a stationary stochastic process the segments of whose sample sequences will include almost all meaningful

English texts. The value space of the random variables will of course be the finite set consisting of all letters and other symbols that occur. This process will have a definite entropy or rate of information transmission which can be regarded as a characteristic of the English language. One can make the same construction with other languages and one finds that different languages have different entropies.

One of the key ideas of Shannon is that there is a very precise sense in which the entropy of a language measures the degree of redundancy in it. This relationship depends upon a nontrivial theorem about stochastic processes called MacMillan's theorem. In essence, it was discovered by Shannon but MacMillan was the first to give a rigorous and general proof. Let F be a finite set as above, and let μ be an invariant probability measure in $(F)^Z$ so that the functions f_j: $\omega \to \omega(j)$ are the F-valued random variables of a stationary stochastic process. Let $\tilde{\mu}_N$ be the natural image of μ in $(F)^N$ defined by the map $\omega \to \omega(1), \omega(2),..., \omega(N)$, and let $\varphi_N(\omega) = -(1/N) \log \tilde{\mu}_N(f_1(\omega), f_2(\omega),..., f_N(\omega))$. Then each φ_N is a random variable on $(F)^Z$ whose expected value is easily seen to be H_N/N. Thus $\lim_{N\to\infty} \int \varphi_N(\omega)\, d\mu(\omega)$ exists and equals the entropy H of the process. MacMillan's theorem states that much more is true; that there exists an invariant measurable function h on $(F)^Z$ such that $\int | \varphi_N(\omega) - h(\omega)|\, d\mu(\omega) \to 0$. In particular, if the action is ergodic, then $h(\omega) = H$ for μ almost all ω. Of course, the function h must be essentially the pointwise entropy introduced in the last section. What is actually used in information theory is the following corollary of MacMillan's theorem. Let the underlying action be ergodic, and let $\epsilon > 0$ and $\delta > 0$ be given. Then there exists N_0 such that $N > N_0$ implies that $| \varphi_N(\omega) - H | \leqslant \epsilon$ except on a set of measure less than δ. In other words, when N is sufficiently large, it is possible to divide all sequences $s_1, s_2, ..., s_N$ in $(F)^N$ into two subsets such that

(1) the sum of the probabilities of the first set is less than δ, and
(2) for every $s_1, s_2, ..., s_N$ in the second set,

$$\left| \frac{H - (-\log \tilde{\mu}_N(s_1, ..., s_N))}{N} \right| < \epsilon.$$

In still other (somewhat less precise) terms, when N is sufficiently large, there will be a set $\mathscr{S}_N$ of N-term sequences all having a probability of occurrence approximately equal to e^{-NH} and such that the probability that at least one occurs is greater than $1 - \delta$. Of course, the number of sequences in this high-probability set $\mathscr{S}_N$ will be approximately e^{NH}.

In terms of messages, this means that if one considers texts of length N, then with very high probability (greater than $1 - \delta$) the message will be one of the set $\mathscr{S}_N$ and all of the messages in $\mathscr{S}_N$ will occur with approximately the same probability. Now the total possible number of messages of length N is of course $|F|^N$, where $|F|$ denotes the number of elements in the set F. Hence, $\mathscr{S}_N$ contains a fraction $(e^H/|F|)^N$ of all possible messages. When the symbols of F occur independently with equal probability, then $H = \log |F|$ and $\mathscr{S}_N$ contains all possible sequences. However, in actual languages this is far from the case so $H < \log |F|$ and $(e^H/|F|)^N$ is a small fraction. Let M be the smallest integer greater than $HN/\log |F|$. Then the number of sequences of length M will be greater than $HN/|F|^{\log |H|} = e^{HN}$ and, hence, greater than the number of sequences of length N in the set $\mathscr{S}_N$. Thus there will exist a one-to-one correspondence between $\mathscr{S}_N$ and a set of sequences of length M, and, by setting up a suitable code, every sequence of length N that is at all likely to occur may be transmitted using only M symbols. In other words, when the entropy H of a language is less than the logarithm of the cardinal number $|F|$ of its "alphabet" F, then by recoding one may compress the length of all messages by the factor $H/\log |F|$. In this sense, the entropy H gives a very direct measure of the magnitude of the redundancy. Of course, high entropy (relative to $\log |F|$) means low redundancy.

Now as mentioned earlier, low redundancy is by no means an unmitigated blessing and we are now prepared to see that this is so in a more quantitative manner. When $H < \log |F|$ so that $(e^H/|F|)^N$ is very small, a random error in a message is very likely to change the message from one in $\mathscr{S}_N$ to one in $F^N - \mathscr{S}_N$ which will then be recognized as an unlikely possibility. On the other hand, when one eliminates all or practically all of the redundancy by recoding, then all possible messages are approximately equally likely and errors cannot be recognized at all.

Before turning to a description of Shannon's quantitative theory of the relationship of probable error to optimal redundancy, we pause to emphasize the importance of the hypothesis of ergodicity in the corollary to MacMillan's theorem. The corollary is simply not true without it. For example, if the action decomposed into two ergodic parts, then there would be two groups $\mathscr{S}_N$ and $\mathscr{S}'_{N'}$ of approximately equally probable N-term sequences, but unless the entropies of the two parts happened to coincide, the probabilities of occurrence of the members of the two groups would be quite different. Of course, it would be the probability of being in $\mathscr{S}_N \cup \mathscr{S}'_{N'}$ that would have probability greater than $1 - \delta$. On the other hand, if all the ergodic parts into which the underlying

action decomposed had the same (or nearly the same) entropy, then one could argue as before, so the hypothesis is in a sense a bit stronger than necessary. As to how restrictive the ergodicity hypothesis is, one can make the same remarks made earlier in connection with stochastic processes in general.

It is also perhaps worth noting that the corollary to MacMillan's theorem provides a consequence of ergodicity that can be recognized in the finite truncations of a process.

The simplest way to consider the introduction of errors into a message being transmitted with an alphabet F is to suppose that for each f in F, there is a certain probability measure α_f in F with the property that f' will be received with probability $\alpha_f(\{f'\})$ when f was sent. If α_f gave equal probabilities to all members of F for all f in F, then what is received would be totally unrelated to what was sent and all the information in the original process would be lost. Of course, what happens realistically is that α_f gives a probability close to 1 to $\{f\}$ and low variable probabilities to other members of F. Consequently, only some information is lost and the next task is to measure how much.

To this end, one sets up a new stochastic process whose random variables are the pairs f, f', where f is the letter sent and f' is the letter received. This will be defined by an invariant measure $\bar{\mu}$ in $(F \times F)^Z$ which can be easily described in terms of the measure μ in F^Z defining the original process and the family $f \to \alpha_f$ of probability measures in F. Indeed, for each ω in $(F)^Z$, let μ_ω be the probability measure in F^Z which is the finite product over Z of the probability measures $\alpha_{(\omega)(n)}$ in F. Integrating the μ_ω in F^Z with respect to the measure μ in F^Z, one obtains a probability measure $\bar{\mu}$ in $F^Z \times F^Z$ and this may be regarded as a probability measure in $(F \times F)^Z$ via the canonical map of $F^Z \times F^Z$ on $(F \times F)^Z$. It is easy to see that $\bar{\mu}$ defines the desired process and is ergodic whenever μ is. Projecting $\bar{\mu}$ on the other of the two factors F^Z, one obtains a probability measure μ' which is also ergodic and invariant and describes the process whose random variables are the letters received. Each of the three processes has an entropy which may be denoted by the symbols H_μ, $H_{\bar{\mu}}$, and $H_{\mu'}$, respectively. Of course, $H_\mu \leqslant H_{\bar{\mu}}$ and $H_{\mu'} \leqslant H_{\bar{\mu}}$, and it is not difficult to show that $H_\mu + H_{\mu'} \geqslant H_{\bar{\mu}}$ so that $0 \leqslant H_{\bar{\mu}} - H_{\mu'} \leqslant H_\mu$ and $0 \leqslant H_{\bar{\mu}} - H_\mu \leqslant H_{\mu'}$. The information emitted by the process defined by μ' represents in part the information coming from the sent message and in part that coming from the random errors. One can argue that $H_{\bar{\mu}} - H_{\mu'}$ is the rate at which the sent information is lost so that $H_\mu - (H_{\bar{\mu}} - H_{\mu'}) = H_\mu + H_{\mu'} - H_{\bar{\mu}}$ repre-

sents the net rate at which the sent information is received. The difference between this and $H_{\mu'}$ represents (uninteresting) information about what errors were actually made.

Now $H_\mu + H_{\mu'} - H_{\bar{\mu}}$ depends both on the original process (i.e., on μ) and on the "channel" $f \to \alpha_f$. Moreover, it is bounded above by $\log |F|$. Hence it has a finite least upper bound C_α as μ varies over all ergodic invariant probability measures in F^Z. This number C_α depends only on the channel $f \to \alpha_f$ and is called the *ergodic capacity* of the channel.

For the special case in which errors are produced in the simple way described above, Shannon's fundamental theorems may be roughly stated as follows. If the entropy of the original message is less than the ergodic capacity of the channel, then given any $\epsilon > 0$ the message may be recoded into a different alphabet in such a way that the transmitted message may be guessed from the received message with a probability of error which is less than ϵ. Moreover, this recoding may be done so that the entropy of the received message is as close as one pleases to that of the transmitted message. For a more exact statement as well as a detailed proof, the reader is referred to the book of Khinchin cited in the last section.

Actually, errors can of course be produced in more complicated ways in which the mistake made in sending a particular letter may depend on several preceding letters. Moreover, the sending and receiving alphabets may be different. Quite generally, one starts with two finite sets F and G and defines a *channel* to be any mapping β assigning a probability measure β_ω on G^Z to each point ω of $(F)^Z$ in such a manner that $\beta_\omega(E)$ is measurable for every Borel set E in $(G)^Z$. The channel is said to be invariant if $\beta_{[\omega]n}(E) = \beta_\omega([E]n)$ for all n, E, and ω. Given an invariant channel, one may define $\bar{\mu}$, μ' and the ergodic capacity just as before. Shannon's theorems are proved in Khinchin's book for any channel which is "non-anticipating" and has a "finite memory". These concepts are, of course, just the translations into precise mathematical language of the condition that the probability of making any particular error should not depend upon future letters and should depend upon only a finite number of past letters. The exact formulation is left to the reader. Finding the codes whose existence is assured by Shannon's theorems is nontrivial and an elaborate and interesting theory has been developed to deal with the problem. We shall not attempt to describe this theory here.

As mentioned in the introduction, the notion of the entropy of a stochastic process has turned out to be of major importance in the

difficult problem of classifying the ergodic actions of the integers. Consider a fixed ergodic action of the integers on a standard Borel measure space Ω, μ and consider the different stochastic processes one gets by choosing different finite sets F and different Borel functions f from Ω to F. Each will have a well-defined entropy H_f which of course will depend only on the disjoint sets $f^{-1}(s)$ for $s \in F$; that is, on the partition $\Omega = \bigcup_{s \in F} f^{-1}(s)$. Considering all possible finite Borel partitions, one obtains a family of possible entropies and this subset of the nonnegative real numbers is of course an isomorphism invariant of the given ergodic action. Conceivably, every nonnegative real number will occur for every action but this turns out not to be the case and the invariant to be a useful one. The idea of investigating it occurred to Kolmogoroff who in 1958 proved the following important theorem. Let $E_1, E_2, \ldots, E_j$ and $F_1, F_2, \ldots, F_k$ define finite Borel partitions of Ω, and let $\mathscr{E}$ and $\mathscr{F}$ be the σ fields of sets generated respectively by the translates of the sets E_i and the translates of the sets F_i, respectively. Then if $\mathscr{E} \supseteq \mathscr{F}$, the entropy defined by $\Omega = E_1 \cup E_2 \cdots \cup E_j$ is greater than or equal to that defined by $\Omega = F_1 \cup F_2 \cdots \cup F_k$. It follows as a corollary that if $\mathscr{E}$ is the family of all Borel subsets of Ω, then the entropy defined by $\Omega = E_1 \cup E_2 \cup \cdots \cup E_j$ is equal to the least upper bound of the entropies defined by all possible finite partitions.

Altering somewhat an earlier definition of Kolmogoroff, Sinai then proposed that one define the *entropy of an action* to be the least upper bound of the entropies of the processes associated with all possible finite partitions. This is clearly an invariant of the action which may be infinite but is finite whenever there is a finite partition which generates the Borel sets.

The Kolmogoroff–Sinai entropy of an action can easily be computed in many cases of interest. For example, when the action has a pure point spectrum—or, more generally, has a spectrum with respect to which Lebesgue measure is not absolutely continuous—then every partition defines a perfectly predictable process and hence one with entropy zero. At the other extreme, consider the so-called "Bernoulli shifts"; that is, the actions underlying stochastic processes whose random variables are independent, separate points and take their values in a finite set F. Such a process and the underlying action is uniquely defined by a probability measure α in F. One computes at once that the entropy of the process is just $-\sum_{s \in F} \alpha(\{s\}) \log \alpha(\{s\})$, that is, the degree of uncertainty in the probability distribution defined by α. Moreover, since the random variables of the process separate points, the entropy of the

action coincides with the entropy of the process. In other words, the entropy of the Bernoulli shift defined by the probabilities $p_1, p_2, \ldots, p_k$ in a set with k elements is precisely $-\sum p_k \log p_k$. In particular, when all p_k are equal to $1/k$ so that one has the so-called k shift, then the entropy is just $\log k$.

At the time that Kolmogoroff proved his theorem, it was known that all Bernoulli shifts had the same spectrum and it was an open question of many years standing as to whether or not they were isomorphic. In particular, no one had been able to decide whether the two shift and the three shift were isomorphic or not. Kolmogoroff's result settled this and many similar questions. Since $\log n \neq \log m$ unless $n = m$, the n shift and the m shift are isomorphic if and only if $m = n$.

Since the Kolmogoroff–Sinai entropy of an action is only at least upper bound of an invariant set of nonnegative real numbers, one could hope to define other invariants by considering other properties of the set. A remarkable theorem announced by Sinai in 1962 and whose proof was published in detail in 1964 shows that this hope is essentially illusory and has several other interesting consequences. Here is the theorem. Let μ be an invariant ergodic probability measure in the standard Borel Z space Ω, and let $0 < h \leqslant \infty$ be less than or equal to the entropy of the action defined by μ. Let $p_1, p_2, \ldots$ be any finite or infinite sequence of positive real numbers such that

$$p_1 + p_2 + \cdots = 1 \qquad \text{and} \qquad -(p_1 \log p_1 + p_2 \log p_2 + \cdots) = h.$$

Then there exists a Borel function f from Ω to the integers j for which p_j is defined such that if $f_n(\omega)$ is $f([\omega]n)$, then the f_n are independent random variables such that $\mu(f_n^{-1}(j)) = p_j$. It follows immediately from this theorem that the entropies defined by the various finite partitions of Ω fill up the open interval $0 < x < H$, where H is the entropy of the action. Another consequence is that when an action has positive entropy H, then every Bernoulli shift with entropy H or less is isomorphic to some quotient action. Still another consequence is that the spectrum of an action with positive entropy always includes a countable Lebesgue component.

The zero greatest lower bound of the entropies of nontrivial quotient actions may or may not be attained. If it is attained, then any stochastic process defined by a function on the corresponding quotient space will be deterministic. This implies, in turn, that any stationary stochastic process defined on the original space whose random variables separate points will be "partially completely predictable". More precisely, certain

nontrivial functions of the random variables $f_1, f_2, f_3, \ldots$ can be predicted with probability one when the values of $\cdots f_{-2}, f_{-1}, f_0$ are known. Those for which no such quotient action exists may be correspondingly thought of as being "completely nondeterministic". For obvious reasons, the underlying action is then said to have completely positive entropy.

That actions with completely positive entropy exist can be seen by making use of the concept of K action (where K stands for Kolmogoroff). [In the bulk of the literature on the subject, one concentrates on the automorphism of Ω, μ which generates the action and refers to K automorphisms.] Given an action, let f be a Borel function from Ω to a finite set F, and for each $n = 1, 2, \ldots$, introduce the following equivalence relation

$$\omega \overset{n}{\sim} \omega' \quad \text{if} \quad f([\omega]k) = f([\omega']k) \quad \text{for} \quad k \leqslant -n.$$

The Borel subsets of Ω which contain all members of an equivalence class when they contain one define a Boolean σ subalgebra B_n of the Boolean algebra of all Borel sets mod μ null sets, and $B_{n+1} \subseteq B_n$. Let $B_\infty = \bigcap_{n=1}^{\infty} B_n$. The action is said to be a K action if B_∞ is trivial for some f whose translates separate points. Notice that if B_∞ is not trivial, then a generating countable family of Borel sets for B_∞ defines an invariant equivalence relation in Ω and hence a quotient action. This quotient action is easily seen to have zero entropy. Thus, if an action admits a stationary process whose random variables separate points and take values in a finite set, then to say that this action is a K action is to say that certain quotient actions which would have zero entropy do not in fact exist. It follows that such an action must be a K action if it has completely positive entropy, and in 1961, Pinsker proved the converse of this result. Every K action has completely positive entropy. It is trivial to verify that Bernoulli shifts are K actions and it follows that every Bernoulli shift has a completely positive entropy. In the same paper, Pinsker proved that any ergodic action which does not have completely positive entropy admits a unique quotient action with zero entropy which is *maximal* in the sense that all other zero-entropy quotient actions are quotients of this quotient. Using Pinsker's results, Rohlin and Sinai proved that any action with a completely positive entropy necessarily has a finite generator and hence is a K action. The K actions are thus identical with the actions having completely positive entropy; that is, which are such that every associated stationary stochastic process is completely nondeterministic.

It was conjectured for a while that every ergodic action with positive entropy is the product of one with zero entropy and one with completely positive entropy and that every K action is isomorphic to some Bernoulli shift. However, both conjectures have been shown to be false by D. S. Ornstein. On the positive side, Ornstein in 1971 proved the following very remarkable converse of Kolmogoroff's theorem on the entropy of Bernoulli shifts. Two Bernoulli shifts with the same entropy are isomorphic. He dealt at first with the case of a finite number of p_j, but managed the general case soon afterwards with the help of Smorodinsky. Since many ergodic actions which are not so defined can be shown to be isomorphic to Bernoulli shifts, this result makes it clear that the classification problem for ergodic actions is not so hopeless as was once supposed.

Except, of course, for the pure point spectrum case, next to nothing has been done toward classifying the ergodic actions of Z^k with $k \geqslant 2$. However, infinite system statistical mechanics makes it clear how to extend the entropy notion to this case, and is it natural to wonder whether the results of Kolmogoroff, Sinai, Ornstein, and others may be correspondingly extended. One thing seems clear at the outset and that is that entropy will not suffice to classify the k-dimensional analogues of the Bernoulli shifts. To see this, consider those ergodic actions of Z^2 of the form $(\omega_1, \omega_2)(n, m) = [\omega_1]n, [\omega_2]m$, where $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, and μ_1 and μ_2 are ergodic invariant probability measures in the standard Borel Z spaces Ω_1 and Ω_2. If the two component Z actions are Bernoulli shifts, the Z^2 action will be a generalized Bernoulli shift whose entropy can presumably be proved to be the sum $H_1 + H_2$ of the entropies of the two component actions. On the other hand, the two component actions are invariants of the Z^2 action and can be recovered from it by restriction to $Z \times e$ and $e \times Z$ and considering the action of each group on the space of ergodic parts of the action of the other. Since $H_1 + H_2 = (H_1 - t) + (H_2 + t)$, there will be many nonisomorphic product actions with the same entropy.

Given a general ergodic action of Z^2, one can associate two actions of Z with it by the restriction process just described, and the entropies of these two Z actions will be invariants of the Z^2 action. However, these Z actions are far from determining the Z^2 action. This is made quite clear by looking at the situation from the virtual group point of view. If H is a subgroup of a product group $G \times G$, one may consider the subgroups H_1 and H_2 of G consisting of all x in G such that x, y is in H for some y in G and all y in G such that x, y is in H for some x in G. Then $H \subseteq H_1 \times H_2$, but H_1 and H_2 are far from determining H. On the other

hand, the two Z actions are the direct analogues of the transitive actions of G on G/H_1 and G/H_2.

Classifying the ergodic actions of Z^2 (and a fortiori of Z^k) is clearly a considerably more difficult problem than that of classifying the ergodic actions of Z.