Introduction

Ergodic theory and functional analysis

Ergodic theory, broadly defined, is the study of group actions on measure spaces. Its origins trace back to the foundations of statistical mechanics in the work of Boltzmann, whose “ergodic hypothesis” famously postulated the equality of time and space averages in the Hamiltonian dynamics of a system of particles at constant energy. Another critical source was Poincaré’s work on celestial mechanics, in which he developed qualitative methods for analyzing solutions to differential equations. In accord with these roots, ergodic theory has mainly concentrated on actions preserving a probability measure. This has led to a rich theory based around recurrence and mixing properties with a remarkable variety of applications to Riemannian geometry, topological and smooth dynamics, operator algebras, Lie theory, harmonic analysis, number theory, and additive combinatorics.

Ergodic theory has benefited in a particularly fundamental way from its connection to functional analysis through the process of linearization, which replaces the measure space with certain spaces of functions on it, most commonly $L^1$, $L^2$, and $L^\infty$. In the case of $L^2$ the action induces a unitary representation of the group, called the Koopman representation, in terms of which the basic dynamical properties of ergodicity, weak mixing, and compactness can all be expressed. This representation was used by von Neumann in formulating and proving his mean ergodic theorem, which identified ergodicity as the condition under which Boltzmann’s hypothesis holds. Together with Birkhoff’s pointwise ergodic theorem, which was established immediately afterward and relied in contrast on $L^1$ techniques, this marked the starting point of ergodic theory as a formal branch of mathematics in the early 1930s.

Once these foundations were laid, attention shifted to the problem of classifying measure-preserving transformations up to conjugacy, a goal which remains hopeless in general but has been achieved in two cases at opposite ends of the stochastic spectrum, namely discrete spectrum transformations and Bernoulli shifts. The Halmos-von Neumann classification of discrete spectrum transformations (or “compact” transformations in terminology appropriate to actions of general groups) was an early success of the Koopman representation viewpoint. It also underscored the limitations of the Hilbert space approach. Most notably, nontrivial Bernoulli shifts cannot be distinguished by spectral
means, as their Koopman representation is always the left regular representation with infinite multiplicity, along with the trivial representation on the constant functions as a direct summand. Their proper analysis had to wait until the late 1950s when Kolmogorov introduced the concept of dynamical entropy, which was ultimately and spectacularly shown by Ornstein in the early 1970s to be a complete invariant for this class of transformations.

While the Koopman representation faithfully translates the probabilistic notion of independence into the geometric relation of orthogonality, it cuts out the higher-order information that accrues under the iterative application of Boolean operations on measurable subsets and their images under the dynamics. It is entropy that provides an asymptotic numerical measure of these higher-order statistics. As such it is not surprising that entropy is most finely attuned to situations where the Koopman representation, as an invariant, says nothing at all. Its algebraic nature is evident in the join operation on partitions in the original Kolmogorov-Sinai definition and in the use of approximately equivariant homomorphisms in the more recent framework of sofic groups.

At the function level, the passage to higher-order phenomena is reflected in the multiplicative structure that one gains in trading $L^2$ for the von Neumann algebra $L^\infty$. Although linearity is constitutive of the Koopman representation, whose power derives from the rotational symmetry of $L^2$ as a Banach space, there is typically no technical advantage in replacing the $\sigma$-algebra of measurable subsets with $L^\infty$. As a consequence, the overt application of functional-analytic ideas took a back seat in the 1960s and 1970s as ergodic theory shifted towards a more substantial use of probability, particularly in its discrete form, along with the associated development of combinatorial methods that one can see embryonically in the Rokhlin lemma and more extensively in entropy theory.

At the same time, this shift towards probability and combinatorics could equally well describe the course of functional analysis over the same period, especially in the core area of Banach spaces, but also in the theory of operator algebras, which was transformed through the emergence of such tools as K-theory, Bratteli diagrams, and free probability. That this is not a coincidence is one of the points we aim to highlight in this book. The title of the book itself speaks directly to Rosenthal’s $\ell_1$ theorem from the 1970s and its quantitative Elton counterpart from the 1980s, the first keyed to weak mixing and compactness and the second to entropy, both by way of the combinatorial notion of independence, which they directly connect to the isomorphic presence of $\ell_1$ structure in a Banach space. This explains the division of the book into two parts based on weak mixing and entropy, which represent the two most salient dynamical properties associated with randomness. While weak mixing and entropy each possess their own very distinct theory, together they share a common technical footing as reflections of the two basic regimes in which combinatorial independence can occur across orbits of subsets. The division is furthermore underscored by the dichotomies of weak mixing vs. compactness and positive vs. zero
entropy. The first of these is a linear version of the set-theoretic gap between the infinite and the finite and admits both $\ell_2$ and $\ell_1$ interpretations, while entropy is fundamentally an $\ell_1$ phenomena due to its higher-order nature.

A collection $\{(A_{i,1}, \ldots, A_{i,k})\}_{i \in I}$ of $k$-tuples of subsets of a set $X$ is said to be independent if $\bigcap_{i \in E} A_{i,f(k)}$ is nonempty for every finite set $E \subseteq I$ and function $f : E \to \{1, \ldots, k\}$. This phenomenon is prototypically exhibited by the cylinder sets $A_{i,k} = \{x \in \{1, \ldots, k\}^I : x_i = k\}$ in a product $\{1, \ldots, k\}^I$. For a group action $G \curvearrowright X$ one applies the notion of independence to orbits or partial orbits of an initial tuple $(A_1, \ldots, A_k)$ of subsets of $X$, in which case the prototype gets reinterpreted as the shift action $G \curvearrowright \{1, \ldots, k\}^G$ together with the cylinder sets $A_k = \{x \in \{1, \ldots, k\}^G : x_e = k\}$ over the identity element $e \in G$. This action is naturally viewed as a topological dynamical system under the product topology, and if we consider the action of $G$ on $C(X)$ induced via composition then the orbit of the function $1_{A_i} - 1_{A_j}$ for $i \neq j$ is isometrically equivalent to the standard basis of $\ell_1^G$ over $\mathbb{R}$, as any linear combination can be evaluated at a point which will cancel out the negative signs among the coefficients. As this shift example demonstrates, the relationships between $\ell_1$, combinatorial independence, and dynamics occur in their most basic form in the topological setting of actions on compact spaces, and that is ultimately where their utility lies, whether one is interested in topological dynamics per se or in measure-preserving actions that come with a canonical topological model as in statistical mechanics or algebraic dynamics. These relationships are however of intrinsic structural interest in measurable dynamics, where they mix with $L_2$-approximation in a way that is novel and intriguing from the Banach space perspective. Unlike in Banach spaces and topological dynamics, for example, the asymptotic Rosenthal analysis converges in this case with the local theory, as understood in the sense of studying of finite-dimensional linear subspaces: up to $L^2$-perturbations, if combinatorial independence occurs along arbitrarily large finite partial orbits then it occurs along an infinite partial orbit.

In the case of entropy, the analysis of combinatorial independence naturally forms part of what is called the local theory of entropy, which was initiated by Blanchard in the early 1990s with his introduction of the concept of entropy pair. Entropy pairs, and entropy tuples more generally, consist of points in the space identified by a positive entropy condition on certain associated open covers. They are the key tool for investigating the behaviour of positive entropy under taking factors. It was eventually discovered, by Huang and Ye for $G = \mathbb{Z}$ and by Kerr and Li more generally, that entropy tuples are the same as independence entropy tuples, or “IE-tuples”, which are defined in contrast by a purely local combinatorial independence condition. This yields in particular a structure theorem for positive entropy in analogy with the Elton $\ell_1$ theorem. While the significance of combinatorial independence for entropy has long been clear, as famously illustrated by
Smale’s horseshoe map, it is only with these more recent results in the context of the local theory that a more systematic understanding has emerged.

The independence which underpins the Rosenthal and Elton $\ell_1$ theorems resides up to scaling in the unit ball of the dual Banach space, which, with its weak* topology, plays the role of the space in a topological dynamical system. Instead of using the dual as a locus for combinatorial investigation, one might try to target the geometry of the Banach space itself as the object of a wider ranging asymptotic analysis. The discovery of probability as a powerful tool for this purpose triggered a profound transformation of Banach space theory in the 1970s and 1980s whose effects continue to reverberate today with connections to operator algebras, convex geometry, quantum information theory, and theoretical computer science. A seminal event was Milman’s use of measure concentration to give a new proof of Dvoretzky’s theorem in 1971. This theorem, in the sharp form obtained by Milman, asserts the existence of almost Hilbertian subspaces of logarithmic dimension in a finite-dimensional Banach space, providing a basic picture of the kind of regularity that is strikingly characteristic of high-dimensional structures. In another highly influential direction, the Johnson-Lindenstrauss dimension reduction lemma, established by randomization, asserts that subsets of a finite-dimensional Hilbert space can be projected onto a subspace of logarithmically smaller dimension with only a small distortion of distances.

Despite the common use of probabilistic tools, the structural relation to dynamics within the general scope of this asymptotic geometric analysis actually only appears in the setting of the Rosenthal and Elton theorems. In order to synchronize with the set algebra combinatorics which give expression to recurrence and mixing properties in dynamics, one must move away from the rotational symmetry of Hilbert space to the extreme cases of $\ell_1$ and $\ell_\infty$ (or the related $L_\infty$ and $C(X)$), where linear geometry rigidifies. This situation can be described under the physical rubric of first quantization, which replaces points in a topological or measure space with vectors in a Hilbert space (the “average” Banach space) for the purpose of expressing the probabilistic calculus of observables subject to the Heisenberg uncertainty principle.

It follows that Hilbert space plays a completely different role here than in the Koopman representation. On the other hand, the impact of this role in Banach space theory has been matched in ergodic theory over the last few decades with similarly dramatic consequences. The reformulation and proof of Szemerédi’s theorem by Furstenberg as a multiple recurrence theorem in the mid-1970s triggered an interest in nonconventional ergodic averages and their application to additive combinatorics that over the last several years has produced striking results, most notably that of Green and Tao on the existence of arbitrary long arithmetic progressions in the primes. Furstenberg’s multiple recurrence theorem, or more precisely the structure theorem for measure-preserving actions on which
it rests, is built around a relativized version of the dichotomy between weak mixing and compactness, and accordingly we will treat it as Chapter 2 in the first part of the book.

A complementary line of investigation in the $L^2$ vein has evolved out of the general unitary representation theory of locally compact groups, going back to von Neumann’s introduction of amenability as a way of gaining some systematic understanding of the phenomenon underlying the Banach-Tarski paradox. A considerable amount of research over the last few decades has been driven by ideas surrounding amenability and its counterpoint, property (T), which was formulated by Kazhdan in the late 1960s as a tool for proving that lattices are finitely generated in many semisimple Lie groups. Margulis took up the study of semisimple Lie groups and their lattices in 1970s and proved several deep results on finiteness, arithmeticity, and rigidity. In the late 1970s Zimmer connected this framework to problems in ergodic theory concerning orbit equivalence and its relation to conjugacy by establishing his groundbreaking cocycle superrigidity theorem, which opened up an area that has come to be known as measurable group theory. More recently this area has witnessed the deployment of nonamenability and its strengthenings like property (T) in dynamical settings of a more abstract nature related to the rigidity theory of von Neumann algebras. This development is epitomized by Popa’s cocycle superrigidity theorems from the mid 2000s, which we treat in the first part of the book in Chapter 5. As a consequence of this superrigidity, if a countable group either satisfies property (T) or is a product of a nonamenable group and an infinite group, then the orbit equivalence relation of each of its Bernoulli actions faithfully encodes both the dynamics up to conjugacy and the group up to isomorphism. In contrast, the classical theorems of Dye and Connes-Feldman-Weiss together show that free ergodic actions of countable amenable groups are all orbit equivalent, as discussed in Chapter 3.

While the concept of amenability appears in an explicit and fundamental way in measurable group theory, its presence actually pervades ergodic theory as a whole. This is due in large part to the fact that it provides the proper general framework for the kind of averaging that ones sees in the von Neumann and Birkhoff ergodic theorems as well as in the Kolmogorov-Sinai formulation of entropy, which all assume their definitive classical form for actions of amenable groups. Amenability (and similarly property (T)) can be expressed by means of a perturbative version of the dichotomy between weak mixing and compactness, and thus its basic theory fits naturally in Part I. It also figures prominently in Part II on entropy and thus acts as a common thread throughout the book.

The various analytic notions of finiteness that one encounters in operator algebras are all rooted in amenability. In the case of discrete groups, to which we will restrict our attention in the book, amenability is fundamentally a combinatorial notion that interprets the set-theoretic dichotomy between the infinite and the finite in a direct and dynamical
way, as we will discuss next, with the relation to weak mixing and compactness being one of linearization via unitary representations.

**Infinite vs. finite:**
**weak mixing, compactness, and amenability**

Axiomatically speaking, functional analysis can be described as the study of structures that couple linearity with various mixtures of topology, measure, algebra, and combinatorics. In practice, much of the richness of the subject has derived from the tension between the infinite and the finite within the context of these structures. As a prime example, the notion of finite-dimensional approximation, while already constitutive of integration theory, became a central theme in Banach spaces after Grothendieck’s Résumé laid the foundations for the local theory, and also in operator algebras after Murray and von Neumann introduced and investigated hyperfiniteness in the last of their pioneering series of papers.

In the group context, the analytic gap between the infinite and the finite is captured by the equivalence of the following properties, which characterize amenability for a discrete group $G$:

(i) (Følner property) For every finite set $E \subseteq G$ and $\delta > 0$ there is a nonempty finite set $F \subseteq G$ such that $|EF \Delta F|/|F| < \delta$.

(ii) There exists a left invariant mean on $G$, i.e., a finitely additive probability measure on $G$ which is invariant under left translation.

(iii) (nonparadoxicality) There do not exist a finite partition

$$G = A_1 \sqcup \cdots \sqcup A_n \sqcup B_1 \sqcup \cdots \sqcup B_m$$

and $s_1, \ldots, s_n, t_1, \ldots t_m \in G$ such that $G = \bigsqcup_{i=1}^n s_i A_i = \bigsqcup_{i=1}^m t_i B_i$.

Condition (ii) is the usual definition of amenability, often expressed as the existence of a state on $\ell^\infty(G)$ which is invariant under the action induced from left translation. The Følner characterization (i) is the combinatorial analogue of finite-dimensional approximation in operator algebras, while the Tarski characterization (iii) at the other formal extreme motivates the operator-algebraic notions of finiteness, infiniteness, and proper infiniteness, the last of which is its precise analogue. Finiteness in this sense corresponds dynamically to the kind of incompressibility for a group action that is witnessed by the presence of invariant probability measures. It plays an implicit background role in the present book given that recurrence and independence are essentially predicated on it, which explains our focus on probability-measure-preserving actions in the measurable case and on actions on compact spaces in the topological case.
One source of the structural richness of operator algebra theory is the fact that the concepts of finiteness and finite-dimensional approximation are logically independent, in contrast to the equivalence of (iii) and (i) above. We can understand this philosophically through the kind of duality in groups that is absent in general operator algebras:

(i) paradoxicality translates, as a global spatial effect of the group acting on itself, into the proper infiniteness of the associated von Neumann algebra crossed product, which by definition twists together $\ell^\infty(G)$ and the group von Neumann algebra, while

(ii) the Følner property corresponds to the hyperfiniteness (local approximability by finite-dimensional subalgebras) of the group von Neumann algebra itself, which plays the role of the Pontrjagin dual in the general noncommutative setting and is always (operator-algebraically) finite.

In fact, in a clash of terminology, finiteness and hyperfiniteness are independent properties, although the celebrated Connes embedding problem asks whether finiteness implies the existence of weaker kinds of finite-dimensional models. In the absence of finiteness, hyperfiniteness is closely intertwined with versions of amenability and paradoxicality for actions on topological and measure spaces, but since these carry us away from the theme of independence we do not treat them in the book. We also point out that finiteness and hyperfiniteness in von Neumann algebras can also be expressed (nontrivially) using traces and hypertraces, respectively, in the spirit of the invariant mean condition (2).

The equivalence for discrete groups of the Følner condition and nonparadoxicality reproduces a more primitive picture for sets. Consider the following conditions expressing finiteness for a set $X$:

(i) There exists a bijection from $X$ to $\{1, \ldots, n\}$ for some natural number $n$.

(ii) (Dedekind finiteness) Every injection function from $X$ to itself is surjective.

(iii) There does not exist a partition of $X$ into two sets each of which has the same cardinality as $X$.

Condition (i) is the standard definition of a finite set and gives us a built-in structure theorem that allows us to exploit its analytic analogues, the Følner property and hyperfiniteness, for quantitative ends, ranging from the averaging in the definition of entropy to the classification theory of von Neumann algebras. Condition (iii) mirrors paradoxicality for groups and proper infiniteness for operator algebras. No meaningful analytic analogue of (ii) exists however for groups, which can be attributed to the symmetry inherent in the structure of a group. On the other hand, the Hilbert hotel phenomenon that one gets by negating (ii) is the basis for operator index theory and corresponds to operator-algebraic infiniteness, as illustrated prototypically by the Toeplitz algebra.
Although one can derive (ii) and (iii) from (i) within Zermelo-Fraenkel set theory, the converse directions require the axiom of choice. The axiom of choice is similarly necessary for proving (1) or (2) from (3).

The reader will have noticed that a match for condition (2) is missing from conditions (i) to (iii). We could certainly write one down, but it would appear artificial here. This stands in curious contrast to the case of groups, where the existence of a left invariant mean prevails as the standard definition of amenability and naturally mediates between the formal extremes of nonparadoxicality and the Følner property.

When we linearize by passing to unitary representations, and hence also to actions through the medium of the Koopman representation, the sharp divergence in character between amenability and nonamenability can be translated into the language of weak mixing and compactness. A unitary representation \( \pi : G \to \mathcal{B}(\mathcal{H}) \) of a group is \textit{ergodic} if \( \mathcal{H} \) has non nonzero \( G \)-invariant vectors, and \textit{weakly mixing} if the representation \( \pi \otimes \bar{\pi} \) is ergodic, where \( \bar{\pi} \) is the conjugate representation. It is \textit{compact} if the norm closure of the \( G \)-orbit of every vector is compact. This terminology transfers to probability-measure-preserving actions by applying it to the restriction of the Koopman representation to the orthogonal complement of the constant functions.

Weakly mixing representations possess a surprisingly strong asymptotic orthogonality property, while compact ones decompose into finite-dimensional subrepresentations. For general unitary representations, a vector in \( \mathcal{H} \) either generates a weakly mixing subrepresentation or has a nonzero component that generates a compact subrepresentation (and hence also a nonzero component generating a finite-dimensional subrepresentation). The Furstenberg-Zimmer structure theorem leverages this dichotomy and its analogue for dynamical factors to achieve a portrait of a general probability-measure-preserving action as a tower of extensions, the top one satisfying a relativization of weak mixing and the remaining ones a relativization of compactness.

For unitary representations in which every group element permutes the members of a fixed orthonormal basis, weak mixing and compactness boil down to the question of whether the orbits of the basis vectors are all infinite or all finite, which reveals that the difference between weak mixing and compactness for representations is a geometrization, in the spirit of quantum mechanics, of the difference between the infinite and the finite. The effect of this geometrization, when applied to actions via the Koopman representation, is that the dichotomy between the infinite and the finite transforms into an arithmetic phenomenon that pits multiplicative structure against additive structure. Weak mixing is an expression of mean asymptotic independence (multiplicative structure), while compactness is equivalent to decomposability into a direct sum of finite-dimensional subrepresentations, which are the Hilbert space counterparts of actions on finite sets by permutations (additive structure).
The property of amenability, along with its cousin property (T), is what emerges when we view the boundary line between weak mixing and compactness through a perturbative lens. Finite-dimensional representations, which are the buildings blocks of compact representations, correspond to finite-rank orthogonal projections commuting with the image of the group. Let us say, in nonstandard but suggestive terminology that mimics an expression for single vectors commonly used in discussions of property (T), that a unitary representation \( \pi : G \to \mathcal{B}(\mathcal{H}) \) has almost invariant finite-dimensional subspaces if for every finite set \( F \subseteq G \) and \( \varepsilon > 0 \) there is a nonzero finite-rank orthogonal projection \( P \in \mathcal{B}(\mathcal{H}) \) such that \( \| \pi(s)P - P\pi(s) \|_2 \leq \varepsilon \|P\|_2 \) for all \( s \in F \), where \( \| \cdot \|_2 \) is the Hilbert-Schmidt norm. Consider then the following conditions for a discrete group \( G \):

(i) Every weakly mixing unitary representation of \( G \) has almost invariant finite-dimensional subspaces.

(ii) No weakly mixing unitary representation of \( G \) has almost invariant finite-dimensional subspaces.

The first is amenability and the second property (T), appearing here in guises that enable us to compare them in a vivid way through the simple flip and negation of a quantifier. The usual definition of property (T) replaces “weakly mixing” with “ergodic” and “finite-dimensional subspaces” with “vectors” (with “has almost invariant vectors” meaning that for every finite set \( F \subseteq G \) and \( \varepsilon > 0 \) there is a unit vector \( \xi \in \mathcal{H} \) such that \( \| \pi(s)\xi - \xi \| < \varepsilon \) for all \( s \in F \)), and the equivalence of this with (2) is a theorem of Bekka and Valette. Note that one cannot make the same substitutions in (1), as every nontrivial one-dimensional unitary representation of \( \mathbb{Z} \) is ergodic but fails to almost have invariant vectors. Even weakly mixing unitary representations of \( \mathbb{Z} \) can fail to have almost invariant vectors, as illustrated by \( n \mapsto z^n \) where \( z^n \) acts on \( L^2(\mathbb{T}, \mu) \) by multiplication and \( \mu \) is any atomless Borel probability measure that does not contain 1 in its support. Moreover, although one can substitute “ergodic” for “weak mixing” in (1), one cannot do this in (2), since there exist property (T) groups that have nontrivial irreducible unitary representations which are finite-dimensional.

Despite its essentially linear nature as a property of unitary representations, weak mixing also has algebraic meaning in the dynamical context, as starkly expressed in its connection to Rosenthal’s \( \ell^1 \) theorem and combinatorial independence, which is explained in Chapter 7. Weak mixing is characterized, both measure-theoretically and combinatorially, by an independence condition across infinite subsets of orbits, which shows it to be a multiplicative version of Poincaré recurrence that considers two or more sets at a time instead of just one. This provides a clear picture of how weak mixing relates to entropy, which asks how frequently the multiplicative recurrence occurs on average and not merely whether it occurs infinitely often.
INTRODUCTION

With all of this discussion in mind, we collect together the following topics in Part I under the title “Weak Mixing and Compactness”. Chapter 1 introduces ergodicity, weak mixing, and compactness. In Chapter 2 we present the Furstenberg-Zimmer structure theorem and Furstenberg’s proof of Szemerédi’s theorem via multiple recurrence. Chapter 3 covers the basic theory of amenability and its role in ergodic theory, including the Rokhlin lemma of Ornstein and Weiss, von Neumann’s mean ergodic theorem, Lindenstrauss’s pointwise ergodic theorem, and the theorems of Dye and Connes-Feldman-Weiss on orbit equivalence. At the opposite end from amenability, we give an introduction to property (T) in Chapter 4. In Chapter 5 we continue to explore nonamenability and present Popa’s cocycle superrigidity theorems and their application to orbit equivalence, as well a result of Bowen in the opposite direction on the orbit equivalence of Bernoulli actions of free groups. Chapter 6 is an introduction to the basic themes in topological dynamics. Chapter 7 presents the notions of tameness and combinatorial independence and discusses their connections to Rosenthal’s $\ell_1$ theorem and weak mixing.

Entropy

From the perspective of combinatorial independence, positive entropy is the density counterpart of weak mixing. Entropy is an asymptotic measure of the average exponential size of the space of finite models for the dynamics at fixed scales, and positive entropy can be structurally characterized by the occurrence of combinatorial independence with positive density over these models. Classically the finite models are partial orbits of subsets or points, while in the more recent theory of sofic entropy they have a more general and abstract character. The analogue of the dichotomy between weak mixing and compactness is the dichotomy between positive and zero entropy.

While weak mixing asks for independence along partial orbits which satisfy the simple set-theoretic condition of being infinite, formulating a positive density condition involves taking a limit across finite partial orbits or other finite models for the dynamics. In order to meaningfully connect independence to the theory of entropy, this finite modeling must be carried out in ways that impose structural requirements on the acting group. The definition of entropy itself demands the same properties of the group for the purpose of taking limits of averages in a manner that produces a computable and nontrivial invariant. The finite modeling in question can be done either

(i) internally using Følner sets, which requires the group to be amenable, or
(ii) externally on finite sets on which the group acts in an approximate way, which requires the group to satisfy the much weaker property of soficity.

Soficity for a group $G$ means that for every finite set $F \subseteq G$ and $\delta > 0$ there is a $d \in \mathbb{N}$ and a map $\sigma : G \to \text{Sym}(d)$ such that
(i) \( |\{ v \in \{1, \ldots, d \} : \sigma_s \sigma_t(v) = \sigma_{st}(v) \}| < \varepsilon d \) for all \( s, t \in F \), and
(ii) \( |\{ v \in \{1, \ldots, d \} : \sigma_s(v) = \sigma_t(v) \}| < \varepsilon d \) for all distinct \( s, t \in F \).

This property originates in work of Gromov and is satisfied by amenable groups, for which one can build the above finite models by patching up the left translation action on Følner sets, as well as residually finite groups, for which one can use left translation on finite quotients, in which case \( \sigma \) is a genuine homomorphism. While nonamenable groups are not hard to come by and include free groups on two or more generators, it is not known whether there exist groups that fail to be sofic. The difficulty in finding obstructions for soficity can be pinned to the fact that it is a purely local property, unlike amenability. Indeed one only needs to find external permutation models for the multiplication table of a given finite subset, instead of having to search around inside the group for suitably good Følner sets.

The distinction between internal and external finite modeling is a fundamental one that has long been pervasive in operator algebra theory, where one might try to locally approximate an algebra either (i) by mapping finite-dimensional algebras into it or (ii) mapping it into finite-dimensional algebras. Amenability and soficity are in fact directly analogous to the von-Neumann-algebraic properties of hyperfiniteness and \( R_\omega \)-embeddability, respectively. As general principle that applies as much to groups and dynamics as to operator algebras, internal modeling trades a loss of generality in the objects of study with a gain in leverage that often leads to structure theorems (e.g., the Rokhlin lemma for actions of amenable groups) or classification (e.g., of hyperfinite von Neumann algebras or certain classes of amenable \( C^* \)-algebras), while external modeling typically applies in broader and more flexible settings where one does not expect to obtain a complete structural understanding but may nevertheless be able to formulate useful invariants like entropy.

What is remarkable is that the internal/external dualism is already inherent in the basic theory of entropy for finite partitions of a probability space, and that this dualism harmonizes with the amenability/soficity alternative when one injects dynamics into the picture. Although they ultimately output the same values, the internal and external approaches to partition entropy differ fundamentally from each other in their motivation and technical set-up. This is confirmed by the significant difference in generality (amenable vs. sofic) with which they may be put to dynamical use. In a curious historical twist that inverts the chronological relation between amenability and soficity, it is the external viewpoint that arose first in this setting, originating in Boltzmann’s work on statistical mechanics in the 1870s. If one has \( d \) indistinguishable particles each of which is assigned one of \( n \) possible values (representing say momentum or position), and the proportion \( c_i \) of particles assigned the \( i \)th value is fixed, then the total number of ways of assigning values to
particles subject to this constraint (the total number of “microstates”) is given by

\[ \frac{d!}{(c_1 d)! \cdots (c_n d)!} \]

Given that we cannot distinguish particles and thus cannot know which particle was assigned which value, this provides a measure of our uncertainty about the actual configuration of the system given the distribution of values into the proportions \( c_1, \ldots, c_n \). If we measure the average exponential size of this quantity by taking a logarithm and dividing by \( d \), then Stirling’s formula tells us that as \( d \to \infty \) with \( c_1, \ldots, c_n \) fixed we obtain

\( -\sum_{i=1}^{n} c_i \log c_i \)  

as the limiting value. One can think of these finite systems of particles as models for a finite partition \( \mathcal{P} \) of a possibly atomless probability space \( (X, \mu) \), where the \( c_i \) will need to vary slightly with \( d \) for arithmetic reasons but should converge to the measures of the atoms of \( \mathcal{P} \) as \( d \to \infty \). Note also that idea of finite approximation is also present in Boltzmann’s physical set-up at the distribution level, as each of the discrete values assigned to particles actually represents a range of microscopic parameters that are indistinguishable at the given scale of observation. The explicit conjunction of these two types of finite approximation is a distinctive feature of the general definition of sofic measure entropy, where one partition controls the finite modeling of the system while another determines the scale at which we are able to distinguish between models.

One can describe the finite models for the partition \( \mathcal{P} \) as homomorphisms from the algebra of subsets generated by \( \mathcal{P} \) to the algebra of subsets of \( \{1, \ldots, d\} \) which approximately pull back the uniform probability measure to \( \mu \). It is in this algebraic sense that we understand the model to be external. Because we are dealing here with spaces rather than groups or operator algebras, we could dualize and talk about (approximate) factors and subsystems instead of subalgebras and their images, thereby transposing externality and internality, but it is worth adhering to the algebraic perspective in this discussion to be consistent with the relation to amenability and soficity. The spatial duality between subsets and points does however play an important role in entropy theory.

The internal approach to partition entropy was introduced by Shannon in the 1940s as a cornerstone of his theory of information. Given a finite partition \( \mathcal{P} \) of a probability space \( (X, \mu) \), one aims to define the amount of information gained in learning that an a priori unknown point \( x \) lies within a specified atom \( A \) of \( \mathcal{P} \). This information should reflect the probability that we can distinguish \( x \) from a random point \( y \), which we can do precisely when \( y \notin A \), and so it is reasonable to assign it the value \( \mu(A)^{-1} \). Since we want to measure the average amount of information gained for a random point \( x \), this quantity should behave additively so that we can integrate it over \( X \), and so we define on \( X \) the
**information function**

\[ I = - \sum_{A \in \mathcal{P}} 1_A \log \mu(A) \]

and for the partition \( \mathcal{P} \) its Shannon entropy

\[ H(\mathcal{P}) = \int I(x) \, d\mu(x) = - \sum_{A \in \mathcal{P}} \mu(A) \log \mu(A). \]

We thereby reproduce formula (1) on the basis of an entirely internal probabilistic heuristic that does not involve finite models. Khinchin demonstrated in fact that the quantity (2) is uniquely determined by a short list of natural axioms.

As studied by Khinchin in the language of stationary stochastic processes, the Shannon entropy can be applied to a measure-preserving transformation \( T : X \to X \) in a straightforward way through the quantities

\[ \frac{1}{n} H(\mathcal{P} \vee T^{-1} \mathcal{P} \vee \cdots \vee T^{-n+1} \mathcal{P}) \]

which measure the average amount of information gained in learning that the trajectory of a random point visits a certain sequence of atoms in the partition \( \mathcal{P} \). The limit of (3) as \( n \to \infty \) exists as a consequence of subadditivity. It was Kolmogorov who realized in the late 1950s that this limit can be used to define a conjugacy invariant \( h(T) \) by showing that it takes a common value among all generating partitions. In response to the problem of how this invariant should be meaningfully formulated in the absence of a generating partition (as for the trivial transformation of an atomless space), Sinai soon after proposed the now standard definition

\[ h(T) = \sup_{\mathcal{P}} \lim_{n \to \infty} \frac{1}{n} H(\mathcal{P} \vee T^{-1} \mathcal{P} \vee \cdots \vee T^{-n+1} \mathcal{P}) \]

where the supremum is taken over all finite measurable partitions. The Kolmogorov-Sinai theorem asserts that this reduces to Kolmogorov’s definition in the presence of a finite generating partition.

For a general measure-preserving group action \( G \curvearrowright (X, \mu) \) we can take averages as in (3) over arbitrary finite subsets of the group, but it is only by doing this asymptotically across Følner sets that limiting values exist and are comparable for different partitions, which enables us to establish a Kolmogorov-Sinai theorem and hence end up with a meaningful and computable invariant. The same scenario applies to topological entropy, which was originally defined by Adler, Konheim, and McAndrew for a homeomorphism \( T : X \to X \) of a compact space by

\[ h_{\text{top}}(T) = \sup_{\mathcal{U}} \lim_{n \to \infty} \frac{1}{n} N(\mathcal{U} \vee T^{-1} \mathcal{U} \vee \cdots \vee \mathcal{U}) \]
where $N(\cdot)$ denotes the minimal cardinality of a subcover and $U$ ranges over the finite open covers of $X$. Thus we arrive at amenability as the ultimate scope of this classical internal approach to entropy, whose basic theory in this generality was developed by Ornstein and Weiss in the 1970s and 1980s. Ornstein and Weiss proved a Rokhlin lemma and extended the Ornstein entropy classification of Bernoulli shifts to this setting, while Kieffer established the Shannon-McMillan theorem and Moulin Ollganier and Pinchon extended the variational principle, which expresses topological entropy as the supremum of the measure entropies over all invariant probability measures. In the late 1990s, Lindenstrauss proved the pointwise ergodic theorem as well as the Shannon-McMillan-Breiman theorem, which replaces the $L^1$ convergence in the Shannon-McMillan theorem with pointwise convergence. Around the same time Rudolph and Weiss showed that complete positive entropy is equivalent to uniform mixing by means of an unexpected orbit equivalence technique that permits one to transfer phenomena from $\mathbb{Z}$-actions to actions of general countable amenable groups. The orbit equivalence method was subsequently used by several authors to extend other classical entropy results to the amenable case.

For some time it appeared that the theory of dynamical entropy could not be pushed in any way beyond amenability, but then in the late 2000s Bowen made the surprising discovery that the Boltzmann idea of finite models could be exploited in the framework of sofic groups to define a more general invariant. To be sure, the Boltzmann picture has long played an important role in the classical theory of entropy and is manifest whenever one replaces partitions and open covers with points, as in the $\varepsilon$-separated set formulation of topological entropy. Until Bowen’s breakthrough, however, the potential of this external viewpoint remained unrealized beyond its utility as a dual interpretation of amenable entropy that tracks orbits of points instead of sets. The most remarkable consequence of Bowen’s entropy is an Ornstein-type classification of Bernoulli actions for a wide class of sofic groups, including those which are torsion-free.

Bowen’s definition of entropy for a measure-preserving action $G \curvearrowright (X, \mu)$ of a countable sofic group starts with a sequence $\Sigma = \{\sigma : G \to \text{Sym}(d_i)\}_{i=1}^{\infty}$ of maps into finite permutation groups which are asymptotically multiplicative and free in the proportional sense of the definition of soficity given above. The entropy $h_\Sigma(\mathcal{P})$ of a finite partition is then a measure of the exponential growth as $i \to \infty$ of the number of models for $\mathcal{P}$ which are dynamically compatible with the sofic approximations for $G$ on the sets $\{1, \ldots, d_i\}$. In sync with the definition of soficity, when taking the limit as $i \to \infty$ this dynamical compatibility is understood locally relative to a fixed finite subset $F$ of $G$ and tolerance $\delta$, over which we then take infima to produce $h_\Sigma(\mathcal{P})$. In general $h_\Sigma(\mathcal{P})$ depends on the choice of sofic approximation sequence $\Sigma$, and so we may potentially get a spectrum of entropy values.
In historical parallel with Kolmogorov, Bowen showed that $h_\Sigma(P)$ takes a common value over generating partitions, yielding an invariant $h_\Sigma(X, G)$ for the action when such a partition exists. However, one cannot simply follow Sinai here and extend this to actions lacking generators by taking a supremum over all finite partitions, as this would yield infinite values for nontrivial Bernoulli actions of a free group $F_r$ with $r \geq 2$ given that such actions all factor onto one another. To circumvent this difficulty, Kerr and Li broadened the notion of generator from partitions to functions, and then finally Kerr gave a generator-free formulation in the conventional ergodic-theoretic language of finite partitions.

So far our discussion has centred on the abstract mechanisms that make the basic theory of dynamical entropy function, with Bernoulli actions as a successful but solitary test case. An astonishing discovery of the 1960s and 1970s was the prevalence of Bernoulli structure among $\mathbb{Z}$-actions deriving from geometry and classical mechanics, as exhibited by geodesic flows on compact hyperbolic surfaces. Given this backdrop it might come as a bit of a shock that the applications of entropy to smooth dynamics that one sees in the case $G = \mathbb{Z}$ completely vanish when one passes, for example, to actions of $\mathbb{Z}^d$ for $d \geq 2$. Indeed smooth $\mathbb{Z}^d$-actions for $d \geq 2$ always have zero topological entropy, since (i) positive entropy for any $\mathbb{Z}^d$-action in the case $d \geq 2$ implies that every nontrivial element of the group has infinite entropy as a single homeomorphism, and (ii) diffeomorphisms always have finite entropy by a compactness argument. It turns out on the other hand that the amenable and sofic theories of entropy are tailor-made for algebraic dynamics, which studies actions by automorphisms on compact Abelian groups and is distinguished by a rich blend of Fourier analysis, algebra, and operator algebra theory with connections to algebraic geometry and number theory.

The study of entropy for algebraic actions traces back to the work of Rokhlin and Yuzvinski on single transformations in the 1960s. In the late 1980s, Lind, Schmidt, and Ward obtained a general entropy formula for algebraic $\mathbb{Z}^d$-actions by automorphisms of compact metrizable groups in terms of the Mahler measure. More recently, Deninger suggested that such a formula might hold for principal algebraic actions of general amenable groups and established it, under a hyperbolicity-like invertibility assumption, for a certain class of groups including those with polynomial growth. In this case it is the Fuglede-Kadison determinant in the von Neumann algebra of the acting group that substitutes for the Mahler measure. Li extended Deninger’s result to the general amenable setting, and subsequently Li and Thom removed the invertibility assumption and also generalized the determinant formula beyond the principal case to include a class of algebraic actions satisfying a homological finiteness condition. For such nonprincipal actions the entropy is no longer expressible as a single determinant but rather as the $L^2$-torsion of the associated
module over the integral group ring. Determinant formulas for the sofic entropy of principal algebraic actions of residually finite groups were also established by Bowen and by Kerr and Li.

For expansive algebraic actions of certain types of amenable groups, entropy possesses in addition a structural significance beyond these determinant formulas by virtue of a duality relation with homoclinicity. This was explored by Lind, Schmidt, and Verbitskiy in the case $G = \mathbb{Z}^d$ using tools from commutative algebra. Chung and Li later developed a different approach using combinatorial independence that enabled them to broaden the scope of the theory to polycyclic-by-finite acting groups. The basis of Chung and Li’s investigation was their discovery that, for actions of a countable group on a compact group $X$ by automorphisms, the local analysis of combinatorial independence within the density regime of positive entropy is governed by a single closed invariant normal subgroup of $X$, called the IE group.

Part II of the book circles through all of these topics with the aim of providing a detailed introduction to the theory of topological and measure-theoretic entropy for actions of amenable and sofic groups. The title of the book is reflected in our emphasis on the role of combinatorial independence and the local theory of entropy. Chapter 8 presents the fundamentals of entropy for actions of amenable groups, and Chapter 9 does the same for sofic entropy. Chapter 10 explores the relationship between entropy and combinatorial independence. In Chapter 11 we present Bowen’s $f$-invariant for probability-measure-preserving actions of free groups and the formula that relates it to sofic entropy. Chapter 12 covers the Rudolph-Weiss orbit equivalence technique following the approach of Danilenko. The Rokhlin-Abramov addition formula from this chapter is used later in Chapter 14. In Chapters 13 and 14 we concentrate on algebraic dynamics. The work of Chung and Li on expansiveness, homoclinicity, entropy, and duality is covered in Chapter 13, while the entropy formulas of Li-Thom in the amenable case and of Bowen and Kerr-Li in the residually finite case are treated in Chapter 14.