

VON NEUMANN ON MEASURE AND ERGODIC THEORY

PAUL R. HALMOS

According to a currently popular principle of classification, mathematics is the study of various "categories." A category consists of certain "objects" (e.g., groups, topological spaces) and certain "mappings" (e.g., homomorphisms, continuous functions). One possible category has measure spaces for its objects and, correspondingly, measure-preserving transformations for its mappings. The usual distinction between pure measure theory on the one hand and ergodic theory on the other hand is merely the distinction between the study of the objects and the study of the mappings of this particular category. The purpose of the following pages is to give a descriptive summary of von Neumann's contributions to this category.

Pure measure theory consists of two parts whose motivations, methods, and results are radically different in both spirit and detail; one part treats finitely additive measures and the other part insists on assuming countable additivity. A corresponding split in ergodic theory is perfectly conceivable, but it just does not happen to exist; up to now ergodic theory has been built on a countably additive foundation only. Von Neumann's most spectacular contribution to this whole circle of ideas is in ergodic theory. This is not to say that he left no mark on pure measure theory; the discovery of the relation of the problem of (finitely additive) measure to group theory, and the proof of the uniqueness of (countably additive) Haar measure in locally compact groups are mathematical accomplishments of considerable importance. There are also a couple of isolated measure-theoretic results, one pretty and startling new proof of an old theorem, and some lecture notes of expository value. Let us proceed to a slightly more technical discussion of these matters, in the following order: finitely additive measures, countably additive measures, and measure-preserving transformations.

The "problem of measure" for n -dimensional Euclidean space R^n may be stated as follows: does there exist a positive, normalized, invariant, and additive set-function on the class of all subsets of R^n ? ("Positive" means non-negative, "normalized" means that the measure of the unit cube is 1, "invariant" means invariant under rigid motions, and "additive" means finitely additive.) The work of

Received by the editors October 28, 1957.

Hausdorff and Banach implies that the problem of measure has a positive solution if $n=1$ or $n=2$ and a negative solution in all other cases. This fact has caused some mathematicians to say that the character of space changes in a fundamental and slightly mysterious manner in the passage from two dimensions to three. The purpose of von Neumann's memoir [5] on the subject is to put the problem into a proper general context and, within that context, to analyze and to extend the known (positive and negative) results. The profound insight to be gained from the paper is that the whole problem is essentially group-theoretic in character, and that, in particular, for the solvability of the problem of measure the ordinary algebraic concept of solvability of a group is relevant. Thus, according to von Neumann, it is the change of group that makes a difference, not the change of space; replacing the group of rigid motions by other perfectly reasonable groups we can produce unsolvable problems in R^2 and solvable ones in R^3 .

The right way to generalize the problem of measure is to replace R^n by an arbitrary set X , to replace the unit cube (mentioned in the normalization) by an arbitrary subset C of X , and to replace the group of rigid motions by an arbitrary group G of transformations acting on X ; let us call the generalization so obtained the (X, C, G) problem. A useful special case is the (G, G, G) problem, where G is an arbitrary group considered as acting on itself by, say, left multiplication. If this special problem has a solution for a particular group G , then that group G is called "measurable." The general problem is reduced to the special problem in this sense: if G is a measurable group, then there is a relatively simple condition on X , C , and G that is necessary and sufficient for the solvability of the (X, C, G) problem. (We do not need to state the condition here, but, for purposes of reference, let us call it the condition K .) An abelian group is always measurable; if N is a normal subgroup of G such that both N and G/N are measurable, then G is measurable. It follows that every solvable group is measurable. Since the condition K turns out to be satisfied in the classical cases (rigid motions on Euclidean spaces), and since the group of rigid motions on R^n is solvable exactly when $n=1$ or $n=2$, the positive solution of the classical problem is subsumed under the generalization.

The classical negative results need different treatment; the crucial condition this time is the representability in G of a free group on two generators. The group of rigid motions on R^n satisfies this condition exactly when $n > 2$, and the unimodular group on R^2 also satisfies it. One of the things we can conclude from all this has a curious sound.

The rigid motions on R^2 are the length-preserving maps and the unimodular transformations are the area-preserving maps, so that the length-preserving maps preserve some generalized area whereas the area-preserving maps do not.

If the negative results are viewed as paradoxical, then there is even a paradox in R^1 , but it is of a metric (not measure-theoretic) character. Briefly: if E and F are subsets of R^1 , write $E < F$ if each of E and F can be decomposed into the same finite number of pieces so that there is a one-to-one and distance-decreasing mapping from each piece of F onto the corresponding piece of E . Assertion: if E and F are arbitrary nondegenerate intervals, then $E < F$.

A related paradox is von Neumann's sharpening of the usual (Vitali) argument for proving the existence of a nonmeasurable set. (The existence of such a set, by the way, yields a conclusive negative answer to the problem of countably additive measure in R^n for all values of n .) The usual construction constructs a countable partition of the perimeter of a circle (or of the entire line) into pairwise congruent sets. Is there such a partition for a bounded interval? The question is of a technical, gymnastic kind, and von Neumann's positive answer [4] uses the set-theoretic and epsilontic trickery appropriate to this domain.

Some of the methods of these papers on pathological measure theory are of greater significance than the results. In anticipation of his later study of dimension theory in algebras of operators, von Neumann made use of Banach's results on equivalence by finite decomposition, and, in anticipation of his later work on almost periodic functions, he had occasion to reformulate the problem of measure in terms of certain "means" of functions.

There is a more or less measure-theoretic question that arises naturally at the beginning of the subject; this is perhaps the appropriate time to mention von Neumann's treatment of it. Is there a way of selecting a collection of measurable sets in, say, the line, so that every measurable set is equivalent to exactly one selected set, and so that the process of selection preserves the finite set-theoretic operations? The answer [7] is yes, and the result is generalizable to the related selection problem for measurable functions. Later, in a joint work with Stone [15], the problem is placed into its appropriate algebraic setting. The general question is this: if A is a Boolean algebra and M is an ideal in A , when does there exist a subalgebra of A such that the restriction to that subalgebra of the projection from A to A/M is an isomorphism? Various sufficient conditions are given;

they are concrete enough to be applicable to the original (set-theoretic) problem.

We turn now to von Neumann's work on analytic, nonpathological measure theory. The most original item here is the proof of the uniqueness of Haar measure. Compact groups are treated in [14]. The method (later used systematically in the study of almost periodic functions) involves the construction of invariant means of continuous functions; the key fact is the compactness of the convex closure of the set of all left translates of each continuous function. The extension of the result to locally compact groups [16] calls for completely different methods. The idea now is to replace a given *left* invariant measure m by a smoothed version m' that has all the assumed properties of m , plus the new and highly desirable property of *right* zero-invariance. The smoothed measure is defined by $m'(E) = \int w(x)m(Ex)dm(x)$, with an appropriate weight function w ; to say that it is right zero-invariant means that $m'(Ex)$ vanishes exactly when $m'(E)$ does. The possibility of passing from m to m' is used to reduce the general uniqueness problem for left invariant measures to the special case of right zero-invariant measures. After the reduction, the technique is a moderately straightforward application of Fubini's theorem; the secret of success is that the group, acting on itself by left multiplications, acts ergodically.

In a footnote in his 1936 paper on Haar measure von Neumann indicated that he knew a new proof of the Radon-Nikodym differentiation theorem, but he did not publish the details of that proof till four years later. The article [17] is devoted to the construction of algebras of operators with peculiar properties. When in the course of the construction von Neumann found that he needed some measure theory, he cheerfully waded in and started writing a quick textbook on the subject. In six pages we get definitions of all the basic concepts (e.g., measure and measurable function) and a motivation, statement, and proof of the Radon-Nikodym theorem. Since the proof tidies up a corner of analysis in a completely satisfactory way, and since it deserves to be better known, we shall look at the situation in a little more detail.

At the center of the stage there are three deservedly celebrated theorems: the Riesz-Fischer theorem (F) on the completeness of L_2 , the Radon-Nikodym theorem (N) on the differentiability of absolutely continuous set-functions, and the Riesz representation theorem (R) for bounded linear functionals on L_2 by means of inner products. None of these theorems by itself is trivial to prove; the

pleasant and perhaps surprising fact is that the passage from any one of them to the others is, comparatively speaking, child's play. Thus, in particular, (R) is a very easy consequence of (N) ; von Neumann's proof proceeds in the reverse direction and shows how to derive (N) from (R) .

The derivation (for finite positive measures) runs as follows. Suppose that m is absolutely continuous with respect to m' , and write $m'' = m + m'$. If $I(f) = \int f dm$, then I is a bounded linear functional on $L_2(m'')$ and consequently, by (R) , $I(f)$ is representable in the form $\int f g dm''$. The Radon-Nikodym derivative of m with respect to m' is $g/(1-g)$.

In his lectures at the Institute for Advanced Study in 1934 von Neumann discussed measure theory. The mimeographed notes of these lectures, despite their limited circulation, were for a long time one of the major sources of measure-theoretic information in the United States. (They became more widely available fifteen years later, when they were reproduced and republished in more permanent form [19].) They contain an extremely thorough and detailed exposition of the classical theory of Lebesgue measure in Euclidean spaces, and they contain also the generalization of the theory to abstract measure spaces. Except for a few technical details (e.g., the concept of a semiring) the work is not original. It does, however, contain results that it is difficult to find anywhere else. It contains, for instance, a study of the connection between set-functions and point-functions (i.e., the theory of "cumulative distribution functions") in R^n , which is an important and complicated subject that most books avoid, or, at best, treat in the trivial one-dimensional case only.

In 1940 von Neumann lectured on measure theory again. The title of the course was *Invariant measures*. The notes were typed, and at least one copy is in the library of the Institute for Advanced Study, but they were never published. The main emphasis of the course was on measures in locally compact spaces and groups. Four of the six chapters are an exposition of the by now well known results of Haar and Weil. The other two chapters digress from the main theme, and, incidentally, they contain original work. Chapter II treats generalized limits. The first purpose of this chapter was to serve as an auxiliary in the proof of the existence of Haar measure, but the treatment is much more comprehensive than that single application requires. Chapter VI contains a construction of Haar measure by means of "approximately equi-distributed" finite sets; some of the results of this chapter were obtained in collaboration with Kakutani.

We come now to von Neumann's work on ergodic theory. The major part of this work was done in the early 1930s; with one exception all his publications on the subject appeared in 1932.

Modern ergodic theory started early in 1931 with a most significant observation made by Koopman [3]. The observation is that the functional operator induced by a measure-preserving transformation is unitary. If, in other words, T is a measure-preserving transformation on some measure space, and if for every square-integrable function f on that space a new square-integrable function Uf is defined by $(Uf)(x) = f(Tx)$, then U is a unitary operator on L_2 . By this time von Neumann had written his epoch-making articles on operator theory and he knew everything about the subject that was then known. Koopman's observation was simultaneously a challenge and a hint. If there is an intimate connection between measure-preserving transformations and unitary operators, then the known analytic theory of such operators must surely give some information about the geometric behavior of the transformations. By October of 1931 von Neumann had the answer; the answer was the mean ergodic theorem.

As stated nowadays, the mean ergodic theorem says that if U is a unitary operator on a Hilbert space H , then the sequence of averages

$$\frac{1}{n} (f + Uf + \dots + U^{n-1}f)$$

is strongly convergent for every f in H . (The mean ergodic theorem also says something about what the limit has to be, and the convergence assertion can just as well be made for suitable one-parameter families $\{U_i\}$ in place of the sequences $\{U^n\}$. These refinements are irrelevant in a summary such as the present one and will be systematically ignored.) In von Neumann's first note [8] on the subject the theorem was not stated in this abstract form; he emphasized the measure-theoretic roots of the result throughout. A reader of the note could have followed the entire argument, involving continual references to the underlying measure space, without ever becoming aware that the proof proved a more abstract theorem than the statement stated.

Shortly after von Neumann proved the mean ergodic theorem he discussed it with G. D. Birkhoff, and shortly after that Birkhoff proved the "individual" ergodic theorem [1]. (The order of events as here stated would be impossible to guess from the dates appearing on the various publications; the priority situation is explained in a

subsequent note [2] of Birkhoff and Koopman.) The difference between von Neumann's result and Birkhoff's is easy to describe; the former states that for suitable functions the sequence of averages

$$\frac{1}{n} (f(x) + f(Tx) + \cdots + f(T^{n-1}x))$$

is convergent in the mean of order 2, whereas the latter states that the same sequence is convergent for almost every x . Thus Birkhoff's result is inextricably tied up with measure; von Neumann's can be (and repeatedly has been) generalized to Banach spaces and other axiomatic contexts. It is therefore curious, but true, that von Neumann always looked at ergodic theory as a part of measure theory; he never worked on the abstract versions. What fascinated him most was the delicate interplay between measure and spectrum. The ergodic theorem itself (mean or individual) was almost never needed in his later work; its main role was that of historical justification for studying measure-preserving transformations.

To understand the nature of von Neumann's profound insights into the structure of measure-preserving transformations, let us look at some of the most striking and typical ones. For simplicity of statement we restrict attention to measure spaces of total measure 1; with a little care many of the results can be extended to infinite spaces.

Associated with every measure space there is a measure algebra, namely the Boolean algebra of measurable sets modulo sets of measure zero. A measure-preserving transformation of the space induces in a natural way a measure-preserving automorphism of the algebra. In highly pathological measure spaces it can happen [18] that the space has fewer transformations than the algebra has automorphisms, i.e., that not every measure-preserving automorphism is induced by a measure-preserving transformation. If the underlying measure space is subjected to some rather reasonable topological and metric conditions, then the pathology cannot happen. Von Neumann's proof of this useful fact [11] is a frightfully complicated set-theoretic argument. The result has not yet been improved in any essential way.

The most interesting measure-preserving transformations are the ergodic ones; ergodicity is the natural measure-theoretic generalization of the concept of transitivity that occurs in the theory of permutation groups. (To be precise, T is ergodic if and only if the only measurable sets invariant under T are the sets of measure zero and their complements.) In his longest and most diversified paper on ergodic theory [12] von Neumann showed that on suitably restricted spaces every measure-preserving transformation can be decomposed

into ergodic pieces. The theorem is a forerunner of many such decomposition (or "direct integral") theorems.

The mixing theorem (obtained in collaboration with Koopman [9]) is a typical example of the measure-versus-spectrum type of result; it asserts that a geometric property of T (mixing) is equivalent to a spectral property of U (no nontrivial proper values). In somewhat informal but quite suggestive language, the mixing theorem says that a necessary and sufficient condition that each pair of sets be eventually stochastically independent is that the spectrum be essentially continuous.

There is also a beautiful result about the very opposite situation in which U has pure point spectrum (i.e., L_2 has an orthonormal basis consisting of proper vectors of U). For ergodic transformations of this kind it turns out [12] that the spectrum, which is always a subset of the set of complex numbers of modulus 1, is in fact a subgroup of the multiplicative group of such numbers, and every such subgroup is the spectrum of some such transformation. Even more is true; the Koopman program is really fulfilled here. If both S and T satisfy the conditions (ergodic, pure point spectrum) then a necessary and sufficient condition for the measure-theoretic isomorphism of S and T is the unitary equivalence of the corresponding unitary operators. In other words, for a special but large class of transformations the analytic (operator) methods give complete information about the geometric (transformation) questions.

This concludes our summary of von Neumann's contributions to measure and ergodic theory. In quantity these contributions amount roughly to one tenth of von Neumann's scientific publications. As to their quality, it seems to be safe to say that if von Neumann had never done anything else, they would have been sufficient to guarantee him mathematical immortality.

BIBLIOGRAPHY

[Except for three notes by Birkhoff and Koopman, added for historical reasons, what follows is a complete list of von Neumann's publications on measure theory and on the mathematical aspects of ergodic theory. Not every item in the list is referred to in the text.]

1. G. D. Birkhoff, *Proof of the ergodic theorem*, Proc. Nat. Acad. Sci. vol. 17 (1931) pp. 656-660.
2. G. D. Birkhoff and B. O. Koopman, *Recent contributions to the ergodic theory*, Proc. Nat. Acad. Sci. vol. 18 (1932) pp. 279-282.
3. B. O. Koopman, *Hamiltonian systems and transformations in Hilbert space*, Proc. Nat. Acad. Sci. vol. 17 (1931) pp. 315-318.
4. John von Neumann, *Die Zerlegung eines Intervalles in abzählbar viele kongruente Teilmengen*, Fund. Math. vol. 11 (1928) pp. 230-238.

5. ———, *Zur allgemeinen Theorie des Masses*, Fund. Math. vol. 13 (1929) pp. 73–116.
6. ———, *Zusatz zur Arbeit "Zur allgemeinen Theorie des Masses,"* Fund. Math. vol. 13 (1929) p. 333.
7. ———, *Algebraische Repräsentanten der Funktionen "bis auf eine Menge vom Masse Null,"* J. Reine Angew. Math. vol. 165 (1931) pp. 109–115.
8. ———, *Proof of the quasi-ergodic hypothesis*, Proc. Nat. Acad. Sci. vol. 18 (1932) pp. 70–82.
9. ——— and B. O. Koopman, *Dynamical systems of continuous spectra*, Proc. Nat. Acad. Sci. vol. 18 (1932) pp. 255–263.
10. ———, *Physical applications of the ergodic hypothesis*, Proc. Nat. Acad. Sci. vol. 18 (1932) pp. 263–266.
11. ———, *Einige Sätze über messbare Abbildungen*, Ann. of Math. vol. 33 (1932) pp. 574–586.
12. ———, *Zur Operatorenmethode in der klassischen Mechanik*, Ann. of Math. vol. 33 (1932) pp. 587–642.
13. ———, *Zusätze zur Arbeit "Zur Operatorenmethode . . . ,"* Ann. of Math. vol. 33 (1932) pp. 789–791.
14. ———, *Zum Haarschen Mass in Topologischen Gruppen*, Compositio Math. vol. 1 (1934) pp. 106–114.
15. ——— and M. H. Stone, *The determination of representative elements in the residual classes of a Boolean algebra*, Fund. Math. vol. 25 (1935) pp. 353–378.
16. ———, *The uniqueness of Haar's measure*, Mat. Sbornik vol. 1 (1936) pp. 721–734.
17. ———, *On rings of operators. III*, Ann. of Math. vol. 41 (1940) pp. 94–161.
18. ——— and P. R. Halmos, *Operator methods in classical mechanics, II*, Ann. of Math. vol. 43 (1942) pp. 332–350.
19. ———, *Functional operators. I. Measures and integrals*, Princeton, 1950.

UNIVERSITY OF CHICAGO AND
INSTITUTE FOR ADVANCED STUDY