HISTORIA MATHEMATICA 4 (1977), 43-65

THE BOREL LAW OF NORMAL NUMBERS, THE BOREL ZERO-ONE LAW, AND THE WORK OF VAN VLECK

BY ALBERT NOVIKOFF, NEW YORK UNIVERSITY, N.Y. 10012
AND JACK BARONE, BARUCH COLLEGE (CUNY), N.Y. 10010

Summaries

A discussion is given of a 1908 paper by the American E. Van Vleck. It is argued that Van Vleck proved the first Zero-One law, anticipating the Zero-One law of Borel and, more strikingly, that of Kolmogorov. A brief description of the evolution of the link between measure theory and probability theory is given. By following Van Vleck's own steps in deriving consequences of his Zero-One Law, a result ("the Extended Van Vleck Theorem") is given which is directly comparable to Borel's Law of Normal Numbers. Finally, it is shown that the Van Vleck Zero-One Law, which in generality falls between that of Borel and that of Kolmogorov, is further distinguished in that it provides the key step in establishing what may be the earliest example in Ergodic Theory of a metrically transitive transformation.

Nous présentons une discussion d'un article de l'Americain E. Van Vleck publié en 1908. Nous montrons que Van Vleck démontra la première loi "zero-un", devançant la loi zero-un de Borel, et surtout celle de Kolmogorov. Nous donnons une brêve description de l'évolution des liens entre la théorie de la mesure et celle des probabilités. En reprenant les propres idées de Van Vleck nous deduisons de sa loi zero-un résultat ("le théorème étendu de Van Vleck") directement comparable à la Loi des Nombres Normaux de Borel. Finalement, nous montrons que la loi zero-un de Van Vleck, qui du point du vue de la généralité est intermédiare entre celle de Borel et celle de Kolmogorov, est aussi remarquable en ce qu'elle est une étape essentialle pour établir ce qui est problement le premier example d'un transformation métriquement transitive.

1.a. PRELIMINARIES

We begin with a precise description of the theorem referred to in the title.

Let B_0 be the set of dyadic expansions (or fractions) of a number x in the unit interval, $x = .x_1x_2...x_n...$, $x_i = 0$ or 1, $i = 1,2,3,...$ satisfying

$$\lim_{n\to\infty} (\nu_n(x)/\mu_n(x)) = 1 \tag{1}$$

where $\nu_n(x)$ = number of 1's in the first n digits of the dyadic expansion of x, and $\mu_n(x)$ = number of 0's in the first n digits of the dyadic expansion of x. Such numbers x are called "normal" in the terminology introduced by Borel. Each such dyadic expansion can be considered as the result of an infinite sequence of independent coin tosses (or trials) with a probability 1/2 of "heads" (meaning digit "zero") and probability 1/2 of "tails" (meaning digit "one").

The Borel Law of Normal Numbers (1909) (also, and perhaps more commonly, called The Borel Strong Law of Large Numbers) asserts

$$P(B_0) = 1, \tag{2}$$

where $P(E)$ denotes in general the probability (suitable defined) of any set E of outcomes.

Consider now a sequence of zeros and ones

$$.x_1x_2x_3... , \; x_n = 1 \text{ or } 0$$

where x_n is 1 or 0 according as a sequence of *biased* coins turns up heads (probability p_n) or tails (probability $q_n = 1 - p_n$). Then the Borel Zero-One Law states that the set E of those expansions with infinitely many ones satisfy

$$P(E) = 0 \text{ or } 1, \tag{3}$$

and indeed

$$P(E) = 0 \text{ if } \Sigma\, p_n \text{ converges,}$$

$$P(E) = 1 \text{ if } \Sigma\, p_n \text{ diverges} \quad [1].$$

Limited to the context of digits in the binary expansion, Borel's Zero-One Law asserts that the probability of infinitely many zeros (and/or ones) is 1. This is, of course, much weaker than the Strong Law, equation (2), above.

1.b. THE STEINHAUS MAP

The link between probability theory and measure theory has today been forged, and its forging played a substantial role in winning acceptance for probability theory among mathematicians. In the time of Borel, this had not yet occurred, and Borel's

1909 paper represents, more than does any other single contribution, the initiation of this link. It was not until 1923 (Steinhaus 1923) that there appeared a complete formal identification between probability theory applied to repeated "tosses" (or trials) of an identical fair coin ($p_n = 1/2$, $n = 1,2,3,\ldots$ in the notation above) and the theory of measure in the unit interval. Describing the situation in modern terminology, Steinhaus accomplished this by observing that the mapping

$$S\colon x_1 x_2 \ldots x_n \ldots \to x_1/2 + x_2/2^2 + \ldots + x_n/2^n + \ldots$$

from sequences (called "trials" or outcomes") to dyadic expansions is a measure-preserving map. More precisely, it is a map betweeen a specified σ-field in the space of "trials," (on which probability is to be defined and to be σ-additive), and the already familiar σ-field of Lebesgue-measurable sets of the interval [0,1], with the familiar Lebesgue measure defined thereon. We shall refer to this mapping S as the *Steinhaus map.*

It will be observed that the correspondence between real numbers x and their dyadic expansions $.x_1x_2\ldots x_n\ldots$ is 1-1 except for those x of the form $m/2^n$, the so-called dyadic rationals. These numbers have two dyadic expansions, one terminating and the other with all x_n equal to 1 for sufficiently large n. The Steinhaus map from sequences to dyadic expansions can be interpreted as a map S from sequences to the real numbers in [0,1] which is "essentially" 1-1, since only the denumerable set of dyadic rationals have two pre-images. From the point of view of measure theory the presence of these dyadic rationals and the corresponding failure of the Steinhaus map to be 1-1 is merely an inconvenience, and we take the liberty of omitting specific references to them henceforth.

More specifically, the Steinhaus map S determines a σ-algebra $\mathcal{B}$ of sets in the space of sequences of all possible outcomes by $\mathcal{B} = \{S^{-1}(F) : F$ is a Lebesgue measurable subset of $[0,1]\}$ and determines an associated measure P on $\mathcal{B}$ defined by $P(E) = m(F)$ if $E = S^{-1}(F)$. The resulting $\mathcal{B}$ and its associated measure $P(\cdot)$ (or more generally the completion of $\mathcal{B}$ with respect to $P(\cdot)$) then give the space of sequences the structure of a probability space in the modern sense, fulfilling all the axioms introduced by Steinhaus.

1.c. THE KOLMOGOROV ZERO-ONE LAW

There are known today several theorems which assert that a wide class of sets, obtained from some probabilistic sequence of events, can have only the probability 0 or 1. Perhaps the best known, and one of the earliest, is that of Kolmogorov [1933, 69].

The full machinery of the Kolmogorov Zero-One Law is more general and more complex than the instances dealt with here. For the sake of completeness, we give below a formulation slightly less general than the one of Kolmogorov, in that we limit ourselves to product spaces with product measure defined on them.

Suppose that for each positive integer n, $\{\Omega_n, \mathcal{B}_n, \mu_n\}$ is a probability space Ω_n with a given σ-field $\mathcal{B}_n$ and a countably additive non-negative measure μ_n defined on the sets of $\mathcal{B}_n$ (normalized by $\mu_n(\Omega_n) = 1$). Then the product space $\Omega = \Pi_1^\infty \Omega_n$ becomes a probability space if we take for its associated σ-field $\mathcal{B}$ the σ-field generated by sets of the form $B_1 \times B_2 \times \ldots \times B_n \times \Omega_{n+1} \times \Omega_{n+2} \times \Omega_{n+3} \times \ldots$. (Sets of the above form are called "cylinder-sets" in Ω.) That is, there exists a unique σ-additive non-negative normalized measure on the sets of $\mathcal{B}$, which we shall denote by μ, determined by the condition that

$$\mu(B_1 \times B_2 \times \ldots \times B_n \times \Omega_{n+1} \times \Omega_{n+2} \times \ldots) = \mu(B_1)\ \mu(B_2) \ldots \mu(B_n).$$

This fact was established by Kolmogorov [1933]. The resulting probability space $\{\Omega, \mathcal{B}, \mu\}$ is called the product of the spaces $\{\Omega_n, \mathcal{B}_n, \mu_n\}$; μ is called the product measure, $\mathcal{B}$ is called the product σ-field.

A set E of $\mathcal{B}$ is called a "tail event" if $\mu(E \cap C) = \mu(E)\mu(C)$ for every cylinder set C.

The Zero-One Law of Kolmogorov says that every tail event E satisfies $\mu(E) = 0$ or $\mu(E) = 1$. (It readily follows that the tail events are themselves a sub σ-field of $\mathcal{B}$ consisting precisely of the sets of μ measure 0 or 1.)

A condition on a set E of $\mathcal{B}$ that assures that E be a tail event is that

$$\text{(I)} \quad \begin{array}{c} \omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots) \in E \\ \text{implies} \\ \omega' = (\omega'_1, \omega'_2, \ldots, \omega'_n, \ldots) \in E \end{array}$$

whenever $\omega'_k = \omega_k$ for all but finitely many values of k. An equivalent formulation is that ω and ω' are both in E (or both in E^c) whenever ω and ω' coincide after deletion of a (common) initial segment. This condition is essentially equivalent to the more abstract definition of tail event given above.

Let us now specialize this to the case in which the probability spaces $\{\Omega_n, \mathcal{B}_n, \mu_n\}$ are as simple as possible: Ω_n consists of two points, called "success" and "failure," or 1 or 0, respectively; $\mathcal{B}_n$ consists of the four possible subsets of Ω_n; and μ_n is defined by $\mu_n(\Phi) = 0$, $\mu_n(\text{success}) = p_n$, $\mu_n(\text{failure}) = q_n = 1 - p_n$, $\mu_n(\Omega_n) = 1$.

In this special case, Ω is the set of infinite sequences of 0 and 1, and the above set E considered by Borel, consisting of those sequences with infinitely many 1's, is a tail event, since it clearly satisfies condition (I). Hence Kolmogorov's Zero-One Law contains Borel's Zero-One Law as a special case.

(The notation $\mu(E)$, rather than $P(E)$, is conventional in this context.)

Further, if each $p_n = 1/2$ (classically described as repeated tosses of an unbiased coin) the set B_0 defined by (1) also satisfies (I), so that from the viewpoint of the Kolmogorov Zero-One Law, $\mu(B_0)$ (or $P(B_0)$ in our earlier notation) is known to be 0 or 1, and the merit of the Borel Law of Normal Numbers is to specify further that it must be 1.

2.2. VAN VLECK AND BOREL

It is virtually unknown today that Borel (and a fortiori, Kolmogorov) had a precursor, the American E. Van Vleck. In 1908, he established a Zero-One Law vastly more general than Borel's (indeed, much closer to Kolmogorov's). In addition, Van Vleck had within his grasp, but failed to establish, a theorem which we will call the Extended Van Vleck Theorem and which establishes a result curiously similar to, but slightly weaker than, Borel's Law of Normal Numbers. In the language of probability theory, the Extended Van Vleck Theorem states

$$P(V_0) = 1 \tag{4}$$

where V_0 is the set of dyadic expansions satisfying

$$\left(\overline{\lim_{n\to\infty}} \frac{\nu_n(x)}{\mu_n(x)}\right)\left(\underline{\lim_{n\to\infty}} \frac{\nu_n(x)}{\mu_n(x)}\right) = 1. \tag{5}$$

Clearly, since $B_0 \subset V_0$ the result contained in (4) is indeed weaker than Borel's [2]. If we introduce the name "nearly normal" for those dyadic expansions lying in V_0, then the theorem Van Vleck narrowly missed proving asserts that the probability of a dyadic expansion being "nearly normal" is 1.

In the interests of historical accuracy, it must be observed that Van Vleck always used the language of measure, not probability. We shall use $m(\cdot)$ to indicate the Lebesgue measure of a set in the unit interval [0,1], and $P(\cdot)$ to indicate the probability of a set of sequences of 0's and 1's (or equivalently, of dyadic expansions) produced by a sequence of independent identical trials with an unbiased coin. That these are necessarily equal is a restatement of the fact that the Steinhaus map is measure-preserving as described above. It is entirely possible that the present-day obscurity of Van Vleck's work is an unfortunate consequence of the fact that it preceded the initial work of Borel and the subsequent developments which permit its probabilistic interpretation. To the extent that Van Vleck's work was read at all, it seems not to have been read by probabilists.

In what follows, we examine in more detail the assertions of Van Vleck, indicating their relations to classic results, and their historic role. We also complete the argument which

Van Vleck initiated, to show how the Extended Van Vleck Theorem follows readily from his main result, which we shall call the Van Vleck Zero-One Law. While the terminology "the Extended Van Vleck Theorem" is cumbersome and perhaps represents a slight elasticity in the art of attribution, the term "the Van Vleck Zero-One Law" seems to us undisputable, and an overdue recognization of deserved priority.

2.b. THE PROGRAM OF VAN VLECK

The Van Vleck paper is concerned, not with probability, but with a program to construct a non-measurable set without the intervention of the Axiom of Choice. To this end, he first established that every set E in $[0,1]$ with a certain geometric character, which he called "homogeneous," necessarily satisfies one of the alternatives: (i) $m(E) = 0$, (ii) $m(E) = 1$, (iii) $m^*(E) = 1$, $m_*(E) = 0$, where $m^*(\cdot)$ and $m_*(\cdot)$ represent Lebesgue outer and inner measure respectively. Alternative (iii) clearly implies that E is non-measurable. As it turns out, the class of *measurable* homogeneous sets (for which (iii) is excluded by hypothesis) is only slightly less general than the class of sets of the Kolmogorov Zero-One Law when applied to the case of repeated identical tosses of an unbiased coin. More specifically, when Kolmogorov's original condition (I) is translated, via the Steinhaus map, to an assertion concerning the unit interval, and probability $P(\cdot)$ is interpreted thereon as Lebesgue measure $m(\cdot)$, the result can be compared directly with Van Vleck's Zero-One Law, which we shall do below.

It was Van Vleck's purpose to exploit alternatives (i)-(iii) in order to construct a set E of "homogeneous" character possessing an additional (anti-) symmetry property, namely that E and its complement E^c were to be obtained from each other by reflection about the point $x = 1/2$. More explicity, E was to be "homogeneous" and constructed so that

$$E^c = \{1 - x : x \in E\}. \tag{6}$$

Condition (6) shows that E and E^c are congruent, and therefore, if measurable, have the same measure. It follows that a "homogeneous" set E satisfying (6) is necessarily non-measurable. (The precise definition of "homogeneous" will be given in a later section.)

Van Vleck showed that any set whose rule for membership depended only on the "ultimate form" of the digits in the binary expansion for the points x of $[0,1]$ is necessarily homogeneous. To say that the rule for membership depends on the "ultimate form" of $x = .x_1x_2\ldots$ means specifically, for Van Vleck, that if $y = .y_1y_2\ldots$, and if for some N and p (depending on x and y), $y_n = x_{n+p}$, $n > N$, then x and y are either both in the set or both in its complement. Here N is a positive integer, and p an integer such that $N + p > 0$. This defines an equivalence relation:

x and y are equivalent if their expansions differ only by a shift, except for finite initial segments. An equivalent formulation is that x and y have the same "ultimate form" (i.e., belong to the same equivalence class) if their dyadic expansions coincide after a suitable initial segment has been dropped from each. The reader will note the similarity to condition (I) above.

The two may be compared directly as follows: condition (I) for a "tail event" E is

$$x = .x_1x_2\ldots x_n\ldots \in E \text{ if and only if} \qquad (7)$$
$$y = .y_1y_2\ldots y_n x_{n+1}x_{n+2} \in E$$

for every $n = 0,1,2,\ldots$ and where $y_1,y_2,\ldots,y_n$ are any choice of n zeros and ones. (When $n = 0$), y is to be interpreted as x.) By contrast, Van Vleck's definition of homogeneity leads to the condition that a set E is homogeneous if it enjoys the "ultimate form" property:

$$x \in E \text{ if and only if } .y_1y_2\ldots y_p x_{n+1}\ldots \in E \qquad (8)$$

for any choice of the positive integers p and n and where $y_1,\ldots,y_p$ are any choice of p zeros and ones. Thus Van Vleck's condition concerning "ultimate form" is more stringent than Kolmogorov's formulation of tail-event, and so Van Vleck's theorem is correspondingly less general. It is readily seen that an equivalent form of Van Vleck's condition (8) is

$$x \in E \text{ if and only if } .x_{n+1}x_{n+2}\ldots \in E, \qquad (8')$$
$$\text{for every } n = 0,1,2,\ldots .$$

A more recent, equivalent formulation will be given in the concluding section.

The construction which Van Vleck proposed for a non-measurable homogeneous set E had the following rule for membership.

$$\text{Let } A(x) = \overline{\lim}\, \frac{\nu_n(x)}{\mu_n(x)}, \quad B(x) = \underline{\lim}\, \frac{\nu_n(x)}{\mu_n(x)} .$$

From this it follows at once that $A(1 - x) = 1/B(x)$, $B(1 - x) = 1/A(x)$. (It should be remarked that the set for which $A(x) = B(x) = 1$ is the Borel set B_0 of "normal" numbers.) The set $V_+ = \{x : A(x) > A(1 - x)\} = \{x : A(x) > 1/B(x)\}$ is homogeneous because it depends only on the "ultimate form" of the dyadic expansion of its members. The set V_+ is entirely assigned to E. Further, let $V_- = \{x : A(x) < A(1 - x)\} = \{x : A(x) < 1/B(x)\}$. Then V_- is also homogeneous and V_+ and V_- are images of each other by reflection about $x = 1/2$ (so that $x \in V_+$ if and only if $(1 - x) \in V_-$). The set V_- is entirely assigned to E^c.

The criterion for membership of the points x in V_+ or V_-

makes no appeal to the Axiom of Choice, but rather to the "ultimate form" of the binary expansion of x, in terms of the behavior of the above defined $A(x)$ and $B(x)$. However, the rule for member ship in E proposed by Van Vleck leaves thus far undecided the membership in E or E^c for those x in neither V_+ nor V_-. This is the set denoted V_0 above in (5). One can complete the definition of E by assigning the elements of V_0 arbitrarily to E or E^c so long as elements of the same ultimate form are assigned consistently. This can be done by an Axiom of Choice argument, sending the equivalence class $[x]$ containing x to E (or E^c) and the equivalence class $[1 - x]$ to E^c (or E, respectively). Only for rational numbers x do x and $1-x$ fail to have distinct equivalence classes, as was observed by Van Vleck. The result is a construction of a non-measurable homogeneous set, employing the Axiom of Choice [3]. But not content with this construction, elegant as it was, Van Vleck hoped the appeal to the Axiom of Choice could be avoided entirely.

Van Vleck's hoped-for example depended on showing that $(V_+ \oplus V_-)^c = V_0$ is of outer measure 0 where by its definition

$$V_0 = \{x : A(x) \cdot B(x) = 1\}. \tag{9}$$

The notation $A \oplus B$ means the union of the disjoint sets A and B. Van Vleck was unable to verify this assertion, but observed that *if it were true*, then V_+ and V_- would be (apart from the set V_0 which would have outer measure 0, and would therefore be negligible) complements and homogeneous without recourse to the Axiom of Choice. They would then of necessity be non-measurable. Although not precisely satisfying the original motivating relationship (6), V_+ and V_- are reflections of each other around the point $x = 1/2$, and are "almost" complements as indicated by the relation

$$V_+^c = V_0 \oplus V_- , \quad V_- = \{1 - x : x \in V_+\}. \tag{10}$$

Thus V_+ and V_+^c would be explicitly constructable, complementary, non-measurable sets, as desired. The key unverified assertion

$$m^*(V_0) = 0 \tag{11}$$

was the only missing piece of the puzzle.

Indeed, Van Vleck observed that if V_0 was merely of inner measure less than 1, then the set $V_0^c = V_+ \oplus V_-$ is a subset of the unit interval, itself partitioned into non-measurable sets V_+ and V_-. However, Van Vleck admitted that he had not been able to show $\overline{m}_*(V_0) < 1$, though he remained hopeful: "... Thus it seems to me possible, and perhaps not difficult, to remove the arbitrary element of choice in my example by confining one's attention to a proper subset of the continuum, though as yet I

have not succeeded in proving that this is possible."
[Van Vleck 1908, 241].

From the very definition of Lebesgue measure m, and of Lebesgue outer measure m^*, (11) is equivalent to

(12) $$m(V_0) = 0.$$

It was for this reason that Van Vleck found himself, one year before Borel's landmark paper, directly addressing himself to the question whether, in the terminology introduced by Borel, almost no numbers were normal, and hoping to prove it to be true.

The authors have been unable, thus far, to settle whether or not Borel was aware of Van Vleck's 1908 paper in 1909, when he addressed a similar question. The only evidence available to us so far is from an informal diary kept by Van Vleck. This contains several entries for November 1905 during a sabbatical stay in Europe. During this month Van Vleck met Borel on several occasions, called on Borel and went to a party chez Borel. Much contact with the French mathematical circle including Painlevé, Frechet, and Hadamard continued through January 1906, culminating in a dinner pary in his honour given by the Hadamards. This certainly raises the possibility that Van Vleck communicated his results on "non-measurable sets" to Borel, perhaps around the period 29 February - 7 March, 1908, when the paper was first read and then received for publication. Had this occured it might have sown the seed of considering the decimal (or binary) digits in Borel's mind as a fruitful example of his emerging theory of "denumerable probability." The fact that the set V_0 occured in connection with non-measurable sets while B_0 arose in "denumerable probability" seemingly free of such entangling alliances would surely have appealed to Borel.

2.c. A CRUCIAL OVERSIGHT OF VAN VLECK

Several observations can be made concerning Van Vleck's partition of [0,1] into $V_+ \oplus V_- \oplus V_0$:

(a) V_+, V_-, V_0 *are clearly measurable:* $A(x)$ and $A(1 - x)$ are lim sups of measurable functions, hence themselves measurable, hence so are the sets where $A(x) > A(1-x)$, $A(x) = A(1-x)$, $A(x) < A(1-x)$ respectively.

(13) (b) $$m(V_+) = m(V_-)$$

since reflection about $x = 1/2$ is a congruence, and therefore a measure-preserving transformation.

(c) V_+, V_-, V_0 are all "homogeneous." Van Vleck himself observed that a set whose membership rule depends only on the "ultimate form" of its elements is necessarily "homogeneous." In particular

(14) $m(V_+)$, $m(V_-)$, $m(V_0)$ are each equal to 0 or 1,

by Van Vleck's alternatives (i)-(iii), and

(15) $$m(V_+) + m(V_-) + m(V_0) = 1$$

by the finite additivity of Lebesgue measure.

The only choices of 0 and 1 in (14) which satisfy (13) and (15) are

(16) $$m(V_+) = m(V_-) = 0,\ m(V_0) = 1.$$

All of these facts were available to Van Vleck. It was only his program, aimed as it was at producing an instance of non-measurability, that could have distracted him from noting, as above, that V_+, V_-, and V_0 are measurable by their very construction (See (a) above).

Thus (16), which we call the Extended Van Vleck Theorem, and and which states that almost all numbers are "nearly normal," could have been proved by Van Vleck himself had he simply observed that V_+, V_-, and V_0 were measurable sets.

To summarize thus far: Van Vleck established, in his alternatives (i)-(iii), that if a set E in [0,1] is both measurable and "homogeneous," then $m(E)$ is either 0 or 1. By so doing he proved the first Zero-One Law, and indeed one which is vastly more general than Borel's in the context of binary digits. He also conjectured that $m_*(V_0) < 1$ although he could easily have established (as above) that $m(V_0) = 1$. Significantly, from the historic viewpoint, both his Zero-One Law and the conjecture $m_*(V_0) < 1$ are couched entirely in the language and framework of Lebesgue measure, and exhibit no link with probability. It is part (indeed a large part) of Borel's achievement in 1909 to approach these questions from a probabilistic viewpoint.

2.d. AN OVERSIGHT OF BOREL

It is supremely ironic that Borel himself suffered from a corresponding oversight, arising at least in part from a similar devotion to a main program: *Borel did not see that his results were equally describable in the language of measure theory.* This last assertion is based on a detailed examination, to be published elsewhere, of Borel's paper (and his earlier and later works as well). However, it is possible at least to sketch here the basis for this surprising conclusion.

First, Borel's approach to the study of denumerable sequences of "coin-tosses" with probability p_n of success on the n^{th} toss, can be mapped, by a suitable generalized Steinhaus map, to the interval [0,1], but only by introducing a measure on [0,1] which generalizes Lebesgue's measure. No such generalization of measure existed in 1909, and indeed it was not until Radon [1913], Carâthéodory [1914], and Fréchet [1915], that such generalizations were achieved [4].

Concerning the special case $p_n = 1/2$ ($n = 1,2,3,\ldots$), it is true that Borel remarked that a "geometric point of view" would lead to questions of measure, which he preferred to leave aside. However there is much internal evidence that the key notion of σ-additivity was not a cardinal feature of probability for Borel: neither as an axiom nor a theorem, nor even a reliable heuristic guide. (For example, Borel sought alternative proofs that the probability of a countable union of sets, each of probability zero, was zero, specifically to avoid appeal to this principle.)

Most significant, perhaps, is Borel's hope, revealed in his concluding lines, that the entire tangle of ideas enveloping the continuum (most especially, its non-denumerable and hence "unknowable" character) might be a side-stepped by using only denumerable trials:

> "When the theory of denumerable probabilities will have been developed in the manner just indicated, it will be interesting to compare the acquired results with those which one obtains in the theory of continuous or geometric probability.
>
> "There *exists* certainly (if it is not a misuse to employ the verb *to exist*) in the geometric continuum some elements which cannot be defined: such is the real sense of the important and celebrated proposition of Mr. Georg Cantor: the continuum is not denumerable. The day when these undefinable elements might really be put aside and when one might no longer require that they intervene more or less implicitly, there would certainly result a great simplification in the methods of Analysis; I would be happy if the preceding pages could contribute to conveying the interest which would be attached to the study of such questions." [Borel 1909, 271] (Italics in the original.)

This passage argues powerfully that Borel considered "denumerable probability" as an alternative to measure theory and/or geometric probability, and not a mere restatement of it, especially in the crucial cases of dyadic and continued fraction expansions which were discussed by him.

In summary, Van Vleck was unconcerned with the fact that measure might be interpreted as probability, while Borel was not fully aware that probability might be interpreted as measure. (Borel remained of course, a notorious skeptic concerning non-measurable sets.) These facts surely contributed to the delay in interpreting Van Vleck's work as a contribution to probability theory.

3.a. THE IMPACT OF VAN VLECK'S WORK

The first (and almost the last) mention we have found of any realization that Borel and Van Vleck were dealing with closely related phenomena occurred in 1910, in a paper by G. Faber [1910]. The first part of this paper was devoted to the discussion of those conditions on the behavior of a function at dyadic rational points in [0,1], which are sufficient or necessary conditions for the almost-everywhere continuity, differentiability, and rectifiability of the function. In particular, Faber constructed a function which he described as being "as little differentiable as possible" for a rectifiable function. His function, which was monotonic, failed to be differentiable precisely at the complement of Borel's set of normal numbers (denoted B_0 above) introduced the previous year. Lebesgue's general theorem to the effect that monotonicity implies that the set of non-differentiable points is of measure zero was of course well known, and was acknowledged by Faber was seminal to his own study. Faber readily concluded, by applying Lebesgue's theorem to his example, that the measure of the set of points B_0^c failing to satisfy (1) is zero:

$$m(B_0^c) = 0 \tag{17}$$

so that

$$m(B_0) = 1. \tag{17'}$$

Immediately after drawing this conclusion, Faber compares his result with both those of Van Vleck and Borel. The reader will observe in the quotation below that Faber gives an exposition of Van Vleck's result not in terms of a set (and its complement) but in terms of its characteristic function, denoted $\phi(x)$. The assertion that V_+ and V_- (See (6) and (10) above) are congruent by reflection about $x = 1/2$ then becomes

$$\phi(1 - x) = 1 - \phi(x),$$

if x is not in the "undecidable" set of V_0. Referring to the set we have denoted B_0^c, those x for which $A(x) = 1$, $B(x) = 1$ fails, Faber writes:

> "This theorem appears interesting to me from many points of view.
>
> "First, it gives a simple example of a set, *which is not only everywhere dense, but also has the cardinality of the continuum in every interval, however small, and nonetheless has measure zero.*
>
> "Borel recently proved, following a formulation of suitable definitions concerning denumerable probabilities, that the probability that a point belong to the above set is zero. The comparison of the above theorem with the Borel result suggests the

question:

"Is the probability--according to the Borel set-up, which possibly might need to be extended to answer this question--that a number belongs to a prescribed set of zero measure, always equal to zero? And conversely: is a set always of measure zero, if the probability that a point belongs to it is equal to zero?

"Finally the above theorem concerning the set M [Authors' note: B_0^c in our notation] is closely connected with an example given by Mr. Van Vleck of a non-analytically representable function [5] $\phi(x)$ $(0 \leq x \leq 1)$. The Van Vleck function takes only two values 0 and 1, and except for a set of measure zero [Authors' note: the dyadic rationals] takes one value at the point x and the other value at the point $1-x$; the definition of $\phi(x)$ remains incomplete, insofar as for a given x it cannot be decided whether $\phi(x) = 1$ and $\phi(1-x) = 0$ or contrariwise $\phi(x) = 0$ and $\phi(1-x) = 1$; for those x for which $\lim \nu_n/\mu_n$ exists and is different from 1, one can easily make the decision, and Mr. Van Vleck is encouraged to inquire whether this remark might point the way to overcoming the incompleteness ("Unvollständigkeit") of the definition of $\phi(x)$. That this is not the case is shown by my above-proved theorem; for one is led exactly to the x-set, for which $\lim_{n\to\infty} \nu_n/\mu_n = 1$, in completing the definition; ..." [Faber 1910, 400]. (Italics in the original).

It is to be noted that while Van Vleck's construction focuses attention on V_0 as the points difficult to assign to E or E^c (See (9)), Faber shows that the subset, defined by (1), is the real nub of the matter.

To summarize, Faber has proved (in measure-theoretic not probabilistic, terms) Borel's theorem $m(B_0) = 1$ (See (2)) and hence that Van Vleck's hopes are dashed since $V_0 \supset B_0$ implies $m(V_0) \geq m(B_0)$ [6]. Further, Faber's comments furnish a remarkably vivid picture of the "state of the art" as late as 1910, at least in the mind of an analyst as up-to-date as was Faber. The implications $m(E) = 0$ implies $P(E) = 0$ and $P(E) = 0$ implies $m(E) = 0$ are viewed as a possibility, not a tautology, and the more general identity

$$P(E) = m(E)$$

had to await Steinhaus for a suitably incisive formulation and accompanying proof. (For other evidence of the imperfect realization of the link between probability and measure theory in the period between Borel and Steinhaus, the reader is referred to the authors' forthcoming paper on this, and to the Doctoral dissertation by J. Barone [1974].)

3.b. THE DEFINITION OF "HOMOGENEOUS": THE WORK OF KNOPP AND JACOBSTHAL

The Van Vleck Zero-One Law is the alternative (i)-(iii) cited above for sets in [0,1] which have "homogeneous" character in the terminology of Van Vleck. The precise definition of "homogeneous" given by Van Vleck is best understood by defining it in two stages, as indeed Van Vleck did. The first, and simpler notion, is that a set E in [0,1] is homogeneous if given any sub-interval $[a,b]$, $0 \leq a \leq b \leq 1$, the set $E \cap [a,b]$ is "geometrically similar" to the set E itself. More precisely, the mapping $x \to a + (b-a)x$ maps the set E onto the set $E \cap [a,b]$. This condition is however a little too stringent to permit the construction of abundant examples. The second, and final definition, is that a set E in [0,1] is homogeneous if, given an arbitrary sub-interval $[a,b]$ of [0,1], there exists a countable union of sub-intervals i_ν of $[a,b]$ having at most end-points in common, and having total length arbitrarily close to $b-a$ such that E is "geometrically similar", in the above sense, to each set $E \cap i_\nu$. If a set E has the property that its rule for membership depends only on the "ultimate form" of the dyadic expansion $x = .x_1x_2x_3...x_n...$, then E is easily seen to be homogeneous since the sub-intervals i_ν of an arbitrarily given $[a,b]$ may be taken to be its maximal sub-intervals of the form $[p/2^n, (p+1)/2^n]$ for various choices of n and p. The mapping $x \to p/2^n + x/2^n$ then maps E onto $E \cap [p/2^n, (p+1)/2^n]$, since x and $p/2^n + x/2^n$ have the same "ultimate form" for their dyadic expansions in the sense of (8) and (8') above. Of course similar constructions with bases other than 2 produce homogeneous sets as well.

In 1915 Knopp and Jacobsthal [Jacobsthal and Knopp 1915] proposed a slightly wider definition of homogeneous, again with emphasis on the fact that measurability was not assumed. The Knopp-Jacobsthal definition is that a set E of [0,1] is "homogeneous with density d" if for a suitable choice of the constant d one has for every sub-interval $[a,b]$ of [0,1]:

$$\frac{m^*(E \cap [a,b])}{b-a} = d. \tag{19}$$

The left-hand side of (19) is called the density of E in the interval $[a,b]$ and is also denoted $D_E(a,b)$. To avoid confusion with Van Vleck's definition, we will call sets with property (19) sets of "constant density." The main Knopp-Jacobsthal results were again theorems of zero-one type. In particular they proved the following theorems (where the numeration is our own):

Theorem 1. A set of constant density d necessarily has $d = 0$ or $d = 1$ (The Knopp-Jacobsthal Zero-One Law).

Theorem 2. If $D_E(a,b) \leq r < 1$ for all $[a,b] \subset [0,1]$, and some fixed r, then E is of constant density 0.

Theorem 3. If $D_E(a,b) \geq r > 0$ for all $[a,b] \subset [0,1]$ and some fixed r, then E is of constant density 1. (Knopp and Jacobsthal acknowledge Caratheodory's remark that the last two theorems, which imply the first, are contained in Lebesgue's theorem that almost every point of a set E is "a point of density 1" in the terminology of Lebesgue; however they contend that their methods are more elementary and that they do not assume measurability.)

The definition of constant density, given above, is broader than the "homogeneity" of Van Vleck, whom Knopp and Jacobsthal cite for his earlier work which they described as "for the purpose of the construction of a non-measurable set." A set which is homogeneous in the Van Vleck sense is necessarily of constant density (so that with the aid of their results the alternatives (i)-(iii) of Van Vleck naturally follow). The converse is not true: if E is a set of constant density, the same need not be true of E^c, while if E is homogeneous in Van Vleck's sense, E^c is again so.

Not until 1926 did Knopp [1926], this time alone, employ his his wider notion of sets of constant density in connection with the Borel Law of Normal Numbers and allied questions. He then observed that a considerable number of recently proved theorems concerning asymptotic density of zeros and ones in the dyadic expansions $x = .x_1x_2...x_n...$ assert that the set of x for which one or another asymptotic assertion is made is a set of measure either zero or one. Specifically, if $\nu_n(x)$ is the number of ones among $x_1,x_2,...,x_n$, then $\nu_n(x) = n/2 + o(n)$ holds for a set of measure 1 (Borel); $\nu_n(x) = n/2 + O(n^{1/2+\varepsilon})$ holds for a set of measure 1 for any given $\varepsilon > 0$ (Hausdorff); $\nu_n(x) = n/2 + O(\sqrt{n \log n})$ holds for a set of measure 1 (Hardy-Littlewood); $\nu_n(x) = n/2 + O(\sqrt{n})$ holds for a set of measure 0 (Hardy-Littlewood); $\nu_n(x) = n/2 + O(\sqrt{n \log \log n})$ holds for a set of measure 1 (Khintchine).

Knopp shows that all such theorems define sets of constant density, and hence of density 0 or 1, thus, in a sense, unifying these earlier results. In fact, each of these sets has a rule for membership which depends on the "ultimate form" of the dyadic expansion, and is therefore "homogeneous" in Van Vleck's sense. Since they are evidently measurable sets, they must satisfy the Zero-One alternative of Van Vleck.

As to probabilistic interpretations, the situation in 1926 had progressed considerably in the direction dimly viewed by Faber: "It is almost a purely terminological matter to use the concept of probability for these theorems: if a point-set M lies in the interval $[0,1]$, one can without further ado describe its measure μ as the probability that an arbitrary real number belongs to the set *M*. In this manner of expression, which many of the above mentioned authors use, the most strikingly prominent phenomenon is that for all of these questions the (transfinite) probability (sic) that a real number satisfies the conditions of the above theorems always has either one or the other of the two values 0 or 1 and none of the values between 0 and 1 come into play." [Knopp 1926, 411].

In effect, Knopp testifies to the fact that by 1926 (if not earlier) at least a certain class of problems were regarded as interchangeably expressible as problems of measure or of probability. The revealing adjective "transfinite" presumably indicates that "ordinary" probability by contrast, is *not* associated with measure. The class of problems in question all deal with complicated sets in [0,1], direct descendents of problems of "geometric probability" of the 19th century. In fact, Steinhaus had been much clearer on this point three years earlier. The full modern scope of probability-as-measure is not foreshadowed by Knopp's relatively modest remarks. He had, however, helped focus the concept of a Zero-One Law.

Nonetheless, it seems to the authors that Van Vleck was closer than Knopp to the viewpoint ultimately adopted by Kolmogorov, inasmuch as Van Vleck's construction, unlike Knopp's, depends palpably on the dyadic representations and their "ultimate form," which leads to the cylinder set concept successfully exploited by Kolmogorov. In any even, Van Vleck's priority is unquestioned.

It is perhaps surprising that in the literature of probability theory, Van Vleck is never mentioned [7]. His Zero-One Law, and his proximity to cylinder sets and tail events was a remarkable achievement in 1908. The last reference to Van Vleck we have located is a lone one by Knopp [1926] to the definition of "homogeneous" given by Van Vleck as being narrower than his own.

4.a. KNOPP, VAN VLECK AND ERGODIC THEORY

While the folk-history of probability theory has up to now bypassed Van Vleck, the contribution of Knopp [1926] has met a better fate. In what follows, we explain how this came about. Therefore it is appropriate to sketch the contents of Knopp [1926] as a preliminary. We will conclude with our final reformulation of Van Vleck's Zero-One Law in contemporary terms.

Having introduced his concept of "constant density" described above, and his Zero-One Law (cited as Theorem 1 above, from Jacobsthal and Knopp [1915]), Knopp was able in 1926 to introduce several examples of (measurable) sets having constant density and hence measure zero or one. The examples cited in the previous section were indeed not the only ones presented in Knopp [1926]. Another class of examples was also given, constucted by imposing conditions on the "ultimate form" of the elements of the continued fraction expansion of x (here x is an irrational number in the unit interval). We describe the meaning of "ultimate form" and indicate Knopp's second class of examples below.

4.b. BOREL AND KNOPP

It should be remarked that Knopp was following the order of Borel's 1909 paper in general outline. That paper was divided into three major sections, the first concluding a zero-one law, the second containing an application (and a refinement) of this zero-one law to the binary (and more generally n-ary) expansions of points from the unit interval, and the third section dealt similarly with terms arising in the continued fraction expansion of (irrational) numbers lying in the unit interval. The Borel Law of Normal Numbers is the chief result of the second section; the chief result of the third section, i.e. the continued fraction application, attracted like notice, and had been pursued by authors such as F. Bernstein [1911; 1912], F. Hausdorff [1914], A. Khintchine [1923; 1924], and shortly after, P. Lévy [1929]. These sets of constant density presented by Knopp thus belong to a tradition well known to probabilists.

Let us introduce the notation

$$x = \frac{1}{a_1 +}\,\frac{1}{a_2 +}\,\ldots\frac{1}{a_n +}\,\ldots$$

for the continued fraction expansion of an arbitrary irrational number x lying in the unit interval. The elements $a_1, a_2, \ldots, a_n, \ldots$ are positive integers determined uniquely by x, and conversely any infinite sequence of positive integers determine a unique corresponding irrational number.

A set E will be said to be determined by the "ultimate form" of the continued fraction expansions of its members if

$$x \in E \text{ if and only if every } x_n \in E \text{ where}$$

$$x_n = \frac{1}{a_{n+1} +}\,\frac{1}{a_{n+2} +}\,\ldots\,. \tag{20}$$

We have introduced the phrase "ultimate form" (not used by Knopp) to stress the analogy with the earlier examples of both Van Vleck and Knopp using binary expansions.

The chief result of Knopp is that every measurable set constructed to satisfy (20) is of constant density, and hence of measure zero or one.

The reader will observe in particular the close analogy of (20) with (8'), our earlier reformulation of Van Vleck's Zero-One Law. Examples of sets E constructed in accordance with condition (20) are: all (irrational) x such that $a_n = 7$ for infinitely many values of n, or such that the sequence $\{a_n\}$ is bounded, or such that the sequence $\{a_n\}$ has a given upper bound, or such that $\Sigma 1/a_n$ diverges, or converges, etc.. (It is immediate that the sets E formed in the above manner are measurable, since each element, $a_n = a_n(x)$, is a measurable function of x.) These sets are *not* homogeneous in Van Vleck's sense.

4.c. MARCZEWSKI AND RYLL-NARDZEWSKI: THE T-TRANSFORMATION AND ERGODIC THEORY

A key formulation of condition (20), due apparently to E. Marczewski (cf. Ryll-Nardzewski [1951, 74]) is obtained by introducing the function $[y]$ = greatest integer $\leq y$, and the transformation

$$T(x) = 1/x - [1/x]$$

which maps the unit interval onto itself [8].
If

$$x = \frac{1}{a_1 +} \frac{1}{a_2 +} \cdots \frac{1}{a_n +} \cdots .$$

then

$$T(x) = \frac{1}{a_2 +} \frac{1}{a_3 +} \cdots \frac{1}{a_n +} \cdots .$$

In terms of T, (20) can be reformulated as either

(20') $x \in E$ if and only if $T^n(x) \in E$ for every $n = 1,2,3,\ldots$

or, more succinctly,

$$T^{-1}(E) = E. \qquad (21)$$

The idea of exploiting these reformulations, especially (21), is due to Ryll-Nardzewski [1951], who describes (21) in contemporary terms by saying E is *invariant* with respect to the transformation T.

4.d. THE S-TRANSFORMATION

An analogous transformation

$$S(x) = 2x - [2x]$$

maps the unit interval onto itself.
If $x = .b_1b_2\ldots$, $b_i = 0$ or 1, then $S(x) = .b_2b_3\ldots$, and condition (8'), can be reformulated in terms of S as either $x \in E$ if and only if $S^n(x) \in E$ for every $n = 1,2,3,\ldots$ or, more succinctly,

$$S^{-1}(E) = E. \qquad (8'')$$

Thus Van Vleck's examples of homogeneous sets defined by the "ultimate form" property are precisely the sets E which are invariant with respect to the transformation S.

4.e. FROM KNOPP TO METRIC TRANSITIVITY

Ryll-Nardzewski made use of (21), by observing further that the measure

$$\nu(J) = \frac{1}{\ln 2} \int_J \frac{1}{1+x}\,dx$$

("Gauss measure") has the property

$$\nu(J) = \nu(T^{-1}(J)) \tag{22}$$

for any sub-interval J of the unit interval and more generally for any (Lebesgue) measurable set J lying in the unit interval. The proof of (22) is quite elementary [Kac 1959, 89-92; Khintchine 1964, section 15; Kuzmin 1928] and seems to go back to Gauss [Urban 1923; Gnedenko 1957]. We note in passing that C. F. Gauss is thus the father of the study of the asymptotic or "ultimate form" of the elements of continued fraction expansions.

Further, since

$$\frac{1}{2 \ln 2} \le \frac{1}{\ln 2}\,\frac{1}{1+x} \le \frac{1}{\ln 2} \quad \text{if} \quad 0 \le x \le 1,$$

it is immediate that Gauss measure $\nu(\cdot)$ and Lebesgue measure $m(\cdot)$ are absolutely continuous with respect to each other, so that $\nu(J)$ equals zero or one accordingly as $m(J)$ equals zero or one respectively.

Thus Ryll-Nardzewski could recast Knopp's result, incorporating these remarks, as follows: if E is a measurable T-invariant set, then $\nu(E)$ is either zero or one. Ryll-Nardzewski definitely cites Knopp as the author of this result, since the "hard part" is to show that if E is a measurable, T-invariant set, then $m(E)$ is either 0 or 1. Thus Knopp has survived in the literature of probability theory.

Once Ryll-Nardzewski had reformulated Knopp's result in these terms, he was able to apply a body of theorems well known collectively by 1951 (but dating back as early as 1932 to Von Neumann and G. D. Birkhoff) as Ergodic Theory. The essential elements needed for application of this theory are a measure space of total measure 1 (the unit interval in this context), together with a measure $\nu(\cdot)$ on the σ-algebra of measurable sets, a transformation (the above T) which is measure-preserving (meaning, specifically (22)) and with the property that the only T-invariant measurable sets have measure 0 or 1. The transformation T is then called metrically transitive (or metrically indecomposable) with respect to the invariant measure $\nu(\cdot)$ [8]. Knopp's contribution is to have established the metric transitivity of T with respect to the non-invariant measure $m(\cdot)$, while Ryll-Nardzewski's contribution is to have applied to T the general theorems of Ergodic Theory dealing with metrically transitive transformations and their invariant measures.

4.e. VAN VLECK RECONSIDERED

In order to apply the same viewpoint adopted by Ryll-Nardzewski to the transformation S, we need only point out that

$$m(J) = m(S^{-1}(J)) \tag{23}$$

for any sub-interval J of the unit interval, or, more generally any measurable subset. In other words, S is measure-preserving for Lebesgue measure $m(\cdot)$. This calculation is orders of magnitude more elementary than the corresponding one needed to verify its counterpart (22) above. The real significance of Van Vleck's Zero-One Law now becomes apparent: every S-invariant measurable set must have measure zero or one, so that *the measure-preserving transformation S is metrically transitive with respect to Lebesgue measure on the unit interval.*

Ryll-Nardzewski characterized as Knopp's the theorem that the transformation T is metrically transitive (or metrically indecomposable) with respect to Lebesgue measure, since this is its content when put in the language of Ergodic Theory. There is an element of generosity in this, since the transformation T was not introduced explicitly by Knopp. Knopp's result in any case needs to be reformulated using the invariant Gauss measure $\nu(\cdot)$ to permit application of Ergodic Theory.

It would seem at least as appropriate to characterize as Van Vleck's the theorem that the transformation S is metrically transitive for Lebesgue measure, $m(\cdot)$ (and thus to assign 1908 as its date of discovery). Since S is already measure-preserving for this measure, the result permits application of Ergodic Theory as stated. It is the *first known instance* of metric transitivity, (preceding by some 20 years the general formulation of the concept by G. D. Birkhoff and P. A. Smith). This interpretation alone should argue for acknowledgement of Van Vleck's contribution.

The significance of metric transitivity can be grasped from reading an historical note by G. D. Birkhoff and B. O. Koopman [1932; especially the concluding lines] in which it is remarked that Carathéodory [1919, 580] first introduced measure theory into the realm of "dynamics". We may say in view of the preceding that without fully realizing it, Van Vleck did something of the same sort in 1908; Van Vleck's example, modified neatly to the unit square instead of the unit interval, and frequently called "the baker's transformation", was shown to be metrically transitive in 1933 by Seidel, after this had been conjectured by G. D. Birkhoff. Birkhoff was evidently unaware in 1933 of the remarkable and neglected contribution of his old friend and fellow-analyst Van Vleck, some thirty years earlier.

ACKNOWLEDGEMENT

The authors would like to acknowledge helpful correspondence from both J. H. Van Vleck and Garret Birkhoff concerning the work of, and friendship between, their fathers. The authors are especially indebted to Professor J. H. Van Vleck for examining his father's memorabilia and making available to us the dated entries concerning the elder Van Vleck's sabbatical visits to France.

NOTES

1. The well-known Borel-Cantelli lemmas observe, with slightly more precision, that if $\Sigma\ p_n$ converges, then $P(E) = 0$ even if the trials are allowed to be dependent.

2. Roughly speaking, the set B_0 consists of those sequences in which 0 and 1 appear with equal asymptotic density 1/2, while V_0 consists of those sequences for which any persistent tendency of 1's to overmatch 0's is accompanied by an equally persistent tendency of 0's to overmatch 1's.

3. The reader is referred to Rosenthal [1975] for a contemporary indication of how recently this construction of Van Vleck has been ignored.

4. Steinhaus accomplished this in 1923 for the case $p_n = 1/2$ $(n = 1,2,3,\ldots)$, and for the more general case $p_n = p$ $(n = 1,2,3,\ldots)$, $0 < p < 1$, but not for arbitrary p_n (as embodied, for example, in Borel's Zero-One Law).

5. Authors' Note: Evidently "analytische nicht darstellbare Funktion" means nonmeasurable function in our terminology.

6. Of course Borel's assertion $P(B_0) = 1$, in 1909, already dashes the hopes of Van Vleck if the reinterpretation of probability $P(\cdot)$ as Lebesgue measure $m(\cdot)$ is made explicit.

7. Van Vleck's near-miss at proving $m(V_0) = 1$ is not to be compared to the much deeper and probabilistically oriented resut result $P(B_0) = 1$ of Borel one year later.

8. In personal correspondence Professor Marczewski asserts that the role of the "shift transformation" for continued fractions was first studied by W. Doeblin [1939] in what Marczewski describes as "an almost forgotten paper."

9. The term metrically transitive was introduced by G. D. Birkhoff and P. A. Smith [1928, 365] though limited by them to one-to-one (and even analytic) transformations of a surface to itself.

BIBLIOGRAPHY

Barone, J 1975 *An Historical Analysis of the Development of Axiomatic Probability Theory* Doctoral Dissertation (New York University)

Bernstein, F 1911 Über eine Anwendung der Mengenlehre aud ein aus der Theorie der säkularen Störengen herrührendes Problem *Mathematische Annalen* 71, 417-439

_____ 1912 Über geometrische Wahrscheinlichkeit und über das Axiom der beschrankten Arithmetisierbarkeit der Beobachtungen *Mathematische Annalen* 72, 585-587

Birkhoff, G D and Koopman, B O 1932 Recent contributions to the Ergodic Theory *Proceedings of the National Academy of Sciences* 18, 279-282

Birkhoff, G D and Smith, P A 1928 Structure analysis of sur surface transformations *Journal de Mathématiques pures et appliquées* 7, 345-379

Borel, E 1909 Les probabilités dénombrables et leurs applications arithmétiques *Rendiconti del Circolo Matematico di Palermo* 27, 247-271

Carátheódory, C 1914 Über das lineare Mass von Punktmengen eine Verallgemeinerung des Langenbegriffs *Nachrichten der Academie der Wissenschaften in Gottingen II. Mathematisch-Physikalische Klass* 4, 404-426

_____ 1919 Über den Wiederkehrsatz von Poincaré *Sitzungsberichte der Preussichen Akademie der Wissenschaften zu Berlin* 32, 580-584

Doeblin, W 1939 Remarques sur la théorie métrique des fractions continues *Compositio Mathematica* 7, 353-371

Faber, G 1910 Über stetigen Funktionen *Mathematische Annalen* 69, 372-443

Fréchet, M 1915 Sur l'intégrale d'une fonctionnelle étendue à un ensemble abstrait *Bulletin de la Société Mathématique de France* 43, 248-265

Gnedenko, B 1957 Über die Arbeiten von C. F. Gauss zur Wahrscheinlichkeitsrechnung *C. F. Gauss Gedenkband* Leipzig (Teubner) 194-204

Hausdorff, F 1914 *Grundzüge der Mengenlehre* Leipzig Reprinted 1965 New York (Chelsea)

Jacobsthal, E and Knopp, K 1915 Bermerkungen über die Struktur linearer Punktmengen *Sitzungsberichte der Berliner Mathematischen Gesellschaft* 14, 121-129

Kac, M 1959 *Statistical Independence in Probability, Analysis and Number Theory* The Carus Mathematical Monographs, no. 12 (Mathematical Association of America)

Khintchine, A 1923 Ein Satz über Kettenbrüche mit arithmetischen Anwendungen *Mathematische Zeitschrift* 18, 289-306

_____ 1924 Einige Satze über Kettenbrüche mit Anwendungen auf die Theorie der Diophantischen Approximationen *Mathematische Annalen* 92, 115-125

______ 1964 *Continued Fractions* Translated by Scripta Technica, Inc. Chicago (University of Chicago Press)

Knopp, K 1926 Mengentheoretische Behandlung einiger Probleme der diophantischen Approximation und der transfiniten Wahrscheinlichkeiten *Mathematische Annalen* 95, 409-426

Kolmogorov, A 1933 *Grundbegriffe der Wahrscheinlichkeitsrechnung Ergbnisse Der Mathematik, Vol. 2, No. 3* English Translation by Nathan Morrison New York (Chelsea Publishing Company) 2nd. edition 1956

Kuzmin, R 1928 Sur un problème de Gauss *International Congress Bologna Vol. VI., 83-89*

Lévy, P 1929 Sur les lois de probabilité dont dépendent les quotients complets et incomplets d'une fraction continué *Bulletin de la Société Mathématique de France* 57, 178-194

Radon, J 1913 Theorie U. Anwendungen der absolut Additiven Mengen funktiones *Sitzungsberichte der Kaiserlichen Academie der Wissenschaften Mathematisch-Naturwissenschaftliche Klass* 122(2), 1295-1438

Rosenthal, J 1975 Nonmeasurable invariant sets *The American Mathematical Monthly* 82, 488-491

Ryll-Nardzewski, C 1951 On the Ergodic Theorems (II) *Studia Mathematica* 12, 74-79

Seidel, W 1933 Note on a metrically transitive system *Proceedings of the National Academy of Sciences* 19, 453-456

Steinhaus, H 1923 Les probabilités dénombrables et leurs rapport à la théorie de la mesure *Fundamenta Mathematicae IV*, 286-310

Urban, F 1923 *Grundlagen der Wahrscheinlichkeitsrechnung und der Theorie der Beobachtungsfehler* Leipzig (Teubner)

Van Vleck, E 1908 On non-measurable sets of points, with an example *Transactions of the American Mathematical Society* 9, 237-244