THE JOURNAL OF SYMBOLIC LOGIC
Volume 47, Number 3, Sept. 1982

PROBABILITIES OVER RICH LANGUAGES, TESTING AND RANDOMNESS

HAIM GAIFMAN AND MARC SNIR[1]

§0. Introduction. The basic concept underlying probability theory and statistics is a function assigning numerical values (probabilities) to events. An "event" in this context is any conceivable state of affairs including the so-called "empty event"—an a priori impossible state. Informally, events are described in everyday language (e.g. "by playing this strategy I shall win $1000 before going broke"). But in the current mathematical framework (first proposed by Kolmogoroff [Ko 1]) they are identified with subsets of some all-inclusive set Ω. The family of all events constitutes a field, or σ-field, and the logical connectives 'and', 'or' and 'not' are translated into the set-theoretical operations of intersection, union and complementation. The points of Ω can be regarded as possible worlds and an event as the set of all worlds in which it takes place. The concept of a field of sets is wide enough to accommodate all cases and to allow for a general abstract foundation of the theory. On the other hand it does not reflect distinctions that arise out of the linguistic structure which goes into the description of our events. Since events are always described in some language they can be indentified with the sentences that describe them and the probability function can be regarded as an assignment of values to sentences. The extensive accumulated knowledge concerning formal languages makes such a project feasible. The study of probability functions defined over the sentences of a rich enough formal language yields interesting insights in more than one direction.

Our present approach is not an alternative to the accepted Kolmogoroff axiomatics. In fact, given some formal language L, we can consider a rich enough set, say Ω, of models for L (called also in this work "worlds") and we can associate with every sentence the set of all worlds in Ω in which the sentence is true. Thus our probabilities can be considered also as measures over some field of sets. But the introduction of the language adds mathematical structure and makes for distinctions expressing basic intuitions that cannot be otherwise expressed. As an example we mention here the concept of a *random sequence* or, more generally, a *random world*, or a world which is typical to a certain probability distribution.

Historically, the concept of a random sequence goes back to von-Mises who attempted to found probability theory on it. Intuitively, a random infinite binary

Received April 15, 1980.

[1]This research has been partly supported by the Israeli Academy of Sciences, Grant No. 015–6937.

sequence is one which can be considered as a typical outcome of an ideal experiment of independent coin tossings, with probability $\frac{1}{2}$ to each side. There is a long history of proposed definitions and the subject has had a revival in the last fifteen years starting with the works of Kolmogoroff [Ko 2] and Martin-Löf [ML]. As it turns out, random sequences are those which exhibit phenomena whose probability is 1. (For example, we know that with probability 1 the relative frequency of 0's should be, in the limit, $\frac{1}{2}$; hence a random sequence should have this property.) In all proposals randomness is defined as the satisfaction of a certain class of properties that have probability 1. The differences arise out of choosing different classes. Now these properties, or laws, are expressible by sentences of our language, L. A *notion of randomness* is nothing other than the satisfaction of all sentences of probability 1 which belong to a certain class Φ. By increasing Φ we get, in general, a stronger notion (this is made precise in terms of serveral theorems in §5) and thus we get a whole scale of randomness notions. Of course Φ should not be arbitrarily chosen. It should represent a certain natural level of sentence complexity—natural from the point of view of logic or from that of probability theory or statistics. Definitions proposed by Martin-Löf [ML] and Schnorr [Sc] are analysed in §5 and shown to arise out of such choices of Φ.

We get a general framework for analysing and comparing various notions of randomness and we also gain generality on the following two counts: (i) The randomness notions apply with respect to any arbitrary probability, not necessarily a Bernoulli distribution. (ii) The entities in question need not be binary sequences. We get a general notion of a random world, i.e., a world which is typical for a certain probability distribution. The exact nature of our "worlds" is determined by the predicates available in our language. For example, they can be sequences of natural numbers, or continuous functions from reals into reals, or continuous paths in a k-dimensional space, etc. We can define, for instance, what it means for a particular path in a 3-dimensional space to be a "random", i.e., typical, Brownian motion, or for a sequence of reals to be a typical outcome of a certain Markov process. Some theorems stated for random binary sequences (for the symmetric coin tossing case) turn out to have a much wider validity. Sometimes the generalization is obvious (Martin-Löf's universal test, which is true for all recursive probabilities and worlds of arbitrary nature). Sometimes it is far from obvious and necessitates a completely new proof (Schnorr's result that his notion is actually weaker than Martin-Löf's, it is true for a very wide class of recursive probabilities). The term 'random' is apt in the symmetric coin tossing case, but is somewhat misleading for arbitrary probabilities. For the probability may be such that a world which is typical with respect to it is not at all, intuitively speaking, random. In an earlier draft we used the term "possible world" (the idea being that a world is "impossible" to the extent that it satisfies sentences of probability 0, so possibility means the satisfaction of sentences of probability 1). But since "random" is already current for the Bernoulli case we prefer it in order not to multiply the terminology; (this accords well with another misleading conventional term viz., 'random variable'—which in fact is a function that may be as nonrandom as you like).

Another concept discussed in §5 is that of *randomness with respect to a family of probability distributions*; for example, a Bernoulli-random binary sequence is one which can be regarded as a typical outcome of *some* coin tossing experiment, where the probability of each side is left unspecified. Here randomness means that there exists a probability in the family with respect to which the world is random. We have in mind natural families, such as all Bernoulli distributions, or all Markov processes (with, say, a given fixed set of states). Since the notion of a random world is well defined with respect to each particular probability, the notion of a world random with respect to the family is thereby uniquely determined. The problem is to find whether this notion can be characterized by means of a set of sentences of L and, if so, to get a characterizing set which is relatively simple. Some partial results are presented in §5 (without proofs) concerning the special case of Bernoulli distributions. On the whole this area is wide open.

The language L to be used in this work is obtained by adding to a purely mathematical language L_0 an empirical part consisting of finitely many so-called *empirical predicates* and/or *empirical function symbols.* L_0 is some language of first-order arithmetic. It has variables that can be quantified, ranging over N (the set of natural numbers) names for members of N, for addition and multiplication and, possibly, for finitely many other fixed relations (or functions) over N. All worlds (or models) have N as their common domain and share the same interpretation of the symbols of L_0. The relations and functions over N that correspond to the empirical symbols are not fixed and each choice determines a particular world. We do not need in this work empirical objects for they can be represented, through the empirical predicates, by numbers; e.g., $P(v)$ can be read as asserting that the vth object (in some presupposed enumeration) is red, or that the vth toss yields heads. An empirical monadic predicate represents in each world a binary sequence (where the nth member is 0 or 1 according as $P(n)$ holds or not). Such empirical binary sequences can be also represented by an empirical function symbol, say f, provided that we consider only worlds satisfying $\forall v(f(v) = 0 \vee f(v) = 1)$. In this case the sentence is regarded as an axiom and its probability should be 1. There is a recursive encoding of worlds into binary sequences; hence, for many purposes, it suffices to treat of empirical binary sequences. But in order to represent the intended structure as clearly as possible one should treat of various empirical symbols rather than a single monadic predicate into which everything is encoded. (Just so, one considers a real-valued discrete Markov process as a distribution over sequences of reals, not over the encoding binary sequences.) Also, the encoding may affect the notion of randomness, for the translation that goes with it may take us out of the underlying class of sentences. (In this work it does not happen, for sentences are mainly classified according to quantifier depth. But it may arise when finer classifications are used.) To see how various entities can be naturally represented as worlds within our framework, note that, by standard techniques, integers and rationals can be fully treated within L_0; if necessary one can add their names and variables ranging over them. Now an empirical sequence of real numbers is representable by a 2-place predicate $R(u, v)$, where v ranges over rationals and $R(u, v)$ asserts that the uth member is $< v$. Continuous functions from reals into reals can be represented by a 2-place predicate, where both variables range over the rationals and $R(v_1, v_2)$ asserts that the function's value for v_1 is $< v_2$; hence, continuous paths in a k-dimensional space can be naturally represented by k binary empirical predicates. (Of course, certain sentences are to be added as axioms, e.g., $\forall u, v_1, v_2\,(R(u, v_1) \wedge v_1 \leq v_2 \rightarrow R(u, v_2))$.)

Besides randomness, various other aspects of the theory of probabilities over L are treated. The motivation for this mathematical treatment is due to basic well-known philosophical questions concerning subjective probability. A probability over L represents a certain a priori point of view. When evidence is gathered this point of view changes. Instead of the initial probability we have the conditional probability, given the collected knowledge. A *testing procedure*, as defined in §2, is a function which, given what is known, picks up the sentence to be tested next. Usually the testing is of relatively simple sentences—atomic empirical sentences, or sentences of a particular form, whereas the hypothesis (the sentence to be confirmed or disconfirmed) is more complex. If the initial probability of an actually true hypothesis is not zero then its conditional probability, as the evidence increases, may tend to 1, meaning that, in the long run, we become convinced of its truth. Under certain general conditions this will happen with probability 1. But if the initial value is 0 it stays so. Assigning value 0 to consistent sentences represents a certain dogmatism. For it amounts to an a priori commitment that no evidence can alter (unless we change our initial probability distribution). Two probability functions are *compatible* if they are equally dogmatic, i.e., they assign the value 0 to the same sentences. (This definition is good for probabilities that are definable in L_0; for arbitrary probabilities a somewhat stronger condition is postulated, cf. §1.) Under certain general conditions two compatible initial distributions will, with probability 1, approach each other as the evidence grows. (Theorems 2.1, 2.2 and 2.4 are precise mathematical formulations of this fact.) Thus empirical evidence will bring together any two points of view provided that they are not dogmatic with respect to each other.

Probability functions are determined by the values assigned to the sentential combinations of atomic empirical sentences (what we call elementary sentences). They are classified according to the difficulty of defining these values (e.g., a probability is called *recursive* if its restriction to elementary sentences is recursive). It turns out that in order to get less dogmatic probabilities (i.e., probabilities which assign positive values to consistent sentences in wider classes) the complexity of the function must increase. This is precisely expressed in Theorems 3.12, 3.14 and Corollary 3.13. For definable probabilities one can always decrease dogmatism at the cost of increasing the complexity.

Note that the richness of L enables one to formulate in it many hypotheses, for example, the hypothesis that a world, say a binary sequence, is random with respect to some given definable probability (where "random" is taken in any of the natural senses given in §5), in particular that a given empirical binary sequence is random with respect to a Bernoulli distribution (with, say, $p = \frac{1}{2}$). On the other hand we can formulate in L "deterministic" hypotheses, e.g. that the binary sequence is recursive, or primitive recursive, or producible by a certain type of Turing machine. Thus, our framework can serve as an explication, or as a model, of a Bayesian point of view. It shows that, at least as far as mathematics is concerned, such a point of view is tenable.

We should also like to point out that through the generalizing of randomness notions such as Martin-Löf's and Schnorr's, this work is also related to the algorithmic approach developed by Chaitin, Levine and others. (We apologize for not mentioning all to whom credit is due. For a survey and a bibliography up to 1977 cf. [Ch].)

The classification used by us is mainly according to quantifier depth. It is intended to give an all-around view. But we feel that the more interesting problems may arise with a finer classification, within the domain of recursive probabilities and restricted types of hypotheses. Phenomena such as the relation between being nondogmatic and being complex may repeat themselves on a finer scale.

By its nature the subject is on the border line of logic and probability (with possible ramifications to the theory of algorithms). It is also motivated by philosophical questions concerning subjective probability, induction and Carnap's work. Aiming at a mixed audience, we have presented the proofs with somewhat greater detail than we would have done otherwise. We also quoted, when necessary, basic facts both from probability and from logic.

§1 presents the basic framework. §2 presents the material concerning collected evidence and testing methods. It is not presupposed in any of the later sections. From §3 on we treat of definable probabilities. The results up to and including Theorem 3.3 are basic and are used throughout the rest of the work. Questions concerning absolute continuity and the assignment of nonzero values to consistent sentences are treated in the rest of §3. This forms a separate part not needed for later sections. All questions concerning randomness are treated in §5. The material of §4 is used to derive some of the theorems of §5, but it can be skipped if the reader is prepared to accept the statements without detailed proofs. By and large, different items in this work can be read independently.

The first to set up a program of defining probabilities on formalized sentences was Carnap [Ca 1], [Ca 2], though the approach is implicit in earlier works in this domain, e.g., [Ra], [Koo], [Ho], to cite only a few examples out of a long list. (For a bibliography and comments concerning the Bayesian approach cf. [Sa].) Carnap considered only a fragment of a language: a countable list of individual names, finitely many monadic predicates and no quantifiers. Also his probabilities were assumed to be symmetric and this was extremely restrictive. The idea of using a full-fledged first order language and arbitrary probabilities was introduced by Gaifman [Ga 1], [Ga 2]. A variant of this idea was later suggested by Łos [Ło] (cf. Fenstad's paper [Fe]). Gaifman's work was extended considerably by Scott and Krauss [S-K].

We thank Fenstad for calling our attention to the work of Hájek and Havránek [H-H] which was unavailable to us when we wrote the paper and which aims, among the rest, at formalizing the Neyman-Pearson theory of testing. Hájek and Havránek use a certain formal setup for describing empirical data and another one for stating statistical hypotheses (such as that X and Y are independent random variables, or that X is normally distributed). The essential difference is that we employ an a priori probability defined on one language, in which both data and hypotheses are described. This is indeed the whole point of our Bayesian approach. Statistical hypotheses are to be translated in our framework into statements concerning randomness. Thus "X is normally distributed" is to be converted into something of the form: "Our world is random with respect to the family of normal distributions", where "random" can be explicated in more than one way. To avoid further lengthening of an overlong paper, we have not gone into the aspects of such translations. Another manuscript concerning randomness is planned to take care of it.

§1. **The basic framework.** Let N = set of natural numbers = $\{0, 1, 2, \ldots, n, \ldots\}$. Let L_0 be a first order language for arithmetic. L_0 has names for the members of N, symbols for addition and multiplication and, possibly, finitely many additional relations and/or function symbols which denote certain fixed relations and functions over N. It has variables that take values in N, quantifiers ($\forall v$ — 'for all v', $\exists v$ — 'there exists v') and sentential connectives ($\neg$ — 'not', $\vee$ — 'or', etc.). We use $\bigvee$ and $\bigwedge$ for finite disjunctions and conjunctions, $\mathbf{u}$, $\mathbf{v}$, ... are used for finite strings of variables and, if $\mathbf{v} = v_0, \ldots, v_{n-1}$, then $\exists\mathbf{v}$ and $\forall\mathbf{v}$ are, respectively, $\exists v_0 \exists v_1 \cdots \exists v_{n-1}$ and $\forall v_0 \forall v_1 \cdots \forall v_{n-1}$.

Let L be obtained by adding to L_0 finitely many predicates and/or function symbols, referred to, jointly, as *empirical symbols* and, separately, as *empirical predicates* and *empirical function symbols*.

A *model for* L consists of an interpretation of the empirical symbols; it is an assignment which correlates with every empirical k-place predicate a subset of N^k and with every empirical k-ary function symbol a function from N^k into N. The symbols of L_0 are interpreted in the same standard way in all models.

Mod_L is the set of all models for L. We shall refer to models as "worlds". The intuitive idea is that pure mathematics is the same but that empirical properties can turn out differently in different worlds; e.g., if $P(v)$ asserts that the vth toss yields heads, then there will be worlds in which $P(13)$ is true and worlds in which it is false.

We use 'n', 'm', ... ambiguously for natural numbers and also for their numerals (names in L_0) and '$\mathbf{n}$', '$\mathbf{m}$', ... for finite strings of the form $n_0, \ldots, n_{k-1}$. Similarly '$+$', '$\cdot$', '$\leq$', '$=$' are used ambiguously ('$\leq$' can be introduced either as a predicate in L_0, or as a shorthand for the definition $\exists x(v_1 + x = v_2)$). When writing $\varphi(v_0, \ldots v_{k-1})$ we usually assume that all its free variables are among $v_0, \ldots, v_{k-1}$. $\varphi(n_0, \ldots, n_{k-1})$ is the sentence obtained by replacing every free occurrence of v_i by n_i.

'$w \models \varphi$' means that the sentence φ is true in the world w. A formula $\varphi(v_0, \ldots, v_{n-1})$ is said to be *valid* if $w \models \varphi(n_0, \ldots, n_{k-1})$ for all $w \in \text{Mod}_L$ and all $n_0, \ldots, n_{k-1}$; we denote this by $\models \varphi$. Two formulas φ, ψ are *equivalent* if $\models \varphi \leftrightarrow \psi$. Evidently, if φ is a sentence in L_0 then either $\models \varphi$ or $\models \neg\varphi$.

DEFINITION 1.1. A *probability on* L is a nonnegative real-valued function Pr, defined for the sentences of L, such that the following conditions hold:

(1.1) (i) If $\models \varphi \leftrightarrow \psi$ then $\Pr(\varphi) = \Pr(\psi)$.

(1.1) (ii) If $\models \varphi$ then $\Pr(\varphi) = 1$.

(1.1) (iii) If $\neg(\varphi \wedge \psi)$ then $\Pr(\varphi \vee \psi) = \Pr(\varphi) + \Pr(\psi)$.
(1.1) (iv)

$$\Pr(\exists v\varphi(v)) = \operatorname{Sup}\{\Pr(\varphi(n_1)\vee \cdots \vee\varphi(n_k)): n_1, \ldots, n_k \in N, k = 1, 2 \ldots\}.$$

The *conditional probability of* φ *given* ψ, denoted $\Pr(\varphi|\psi)$, is defined as $\Pr(\varphi \wedge \psi)/\Pr(\psi)$. It is a function of two arguments defined for all (φ, ψ) such that $\Pr(\psi) > 0$.

(1.1)(i) means that Pr depends only on the equivalence class of the sentence and (1.1)(ii) and (1.1)(iii) correspond to the requirements of a normalized finitely additive measure. We presuppose all their elementary consequences such as $\Pr(\neg\varphi) = 1 - \Pr(\varphi)$, $\models\varphi\rightarrow\psi \Rightarrow \Pr(\varphi) \leq \Pr(\psi)$, etc. On the sentences of L_0, Pr assumes the values 1 and 0, 1– for true sentences, 0– for false ones.

Given the first 3 requirements, (1.1)(iv) is equivalent to

$$\Pr(\exists v\varphi(v)) = \lim_{n\to\infty} \Pr(\bigvee_{i<n} \varphi(i))$$

(as well as to

$$\Pr(\forall v\varphi(v)) = \lim_{n\to\infty} \Pr(\bigwedge_{i<n} \varphi(i))).$$

Note, however, that (1.1)(iv) is also meaningful in the more general case where N is replaced by an uncountable domain. We shall remark on this possibility at the end of this section. (Originally a condition of the form (1.1)(iv) had appeared in [Ga 1] and was later referred to in [S-K] as Gaifman's condition.) It is easily seen that if $\mathbf{v} = v_0, \ldots, v_{m-1}$, then

$$\Pr(\exists\mathbf{v}\varphi(\mathbf{v})) = \operatorname{Sup}\{\Pr(\varphi(\mathbf{n}_0)\vee \cdots \vee \varphi(\mathbf{n}_{k-1})): \mathbf{n}_0, \ldots, \mathbf{n}_{k-1} \in N^m, k = 1, 2, \ldots\}.$$

A similar equality holds for $\forall\mathbf{v}\varphi(\mathbf{v})$.

Pr is uniquely determined by its values for quantifier-free sentences. The proof of this is by induction on the number of alternating blocks of quantifiers for sentences in prenex normal form using the fact that, up to logical equivalence, each level of the hierarchy is closed under conjunctions and disjunctions. Here the hierarchy is the natural extension of the arithmetic hierarchy to L:

A $\Sigma_0(L)$*-formula* (and also a $\Pi_0(L)$*-formula*) is one in which all quantifiers are bounded. By a *bounded quantifier*, we mean one of the forms $\exists v < t$ or $\forall v < t$, where t is either a numeral or a variable (and where '$\exists v < t(\ldots)$', '$\forall v < t(\ldots)$' stand for '$\exists v(v < t \wedge \ldots)$', '$\forall v(v < t \rightarrow \ldots)$', respectively). A formula φ is $\Sigma_n(L)$, also denoted '$\varphi \in \Sigma_n(L)$', if it has the form $\exists\mathbf{v}_1\forall\mathbf{v}_2 \cdots \psi$ where $\psi \in \Sigma_0(L)$ and the prefix consists of n alternating quantifiers. A $\Pi_n(L)$-formula is defined dually. For our purposes the difference between equivalent formulas can be disregarded. When doing so we shall speak of 'conditions'; thus a $\Sigma_n(L)$-condition is any formula which is equivalent to a $\Sigma_n(L)$-formula. A $\Delta_n(L)$-condition is one which is both $\Sigma_n(L)$ and $\Pi_n(L)$. (But 'condition' is also used to refer to what the formula expresses, e.g., $\exists vP(v)$ is a $\Sigma_1(L)$-condition on P.) A $\Sigma_n(L_0)$-formula

($\Pi_n(L_0)$-formula, $\Delta_n(L_0)$-condition) is a formula of L_0 which is $\Sigma_n(L)$ ($\Pi_n(L)$, $\Delta_n(L)$). For L_0 we get the usual arithmetic hierarchy. We presuppose all elementary facts concerning this hierarchy, such as that each level is closed, up to logical equivalence, under conjunctions, disjunctions and bounded quantification (e.g., $\varphi, \psi \in \Sigma_n(L) \Rightarrow \varphi \vee \psi$, $\varphi \wedge \psi$, $\exists v < t\ \varphi$ and $\forall v < t\ \varphi$ are equivalent to $\Sigma_n(L)$ formulas, etc.), or that by applying negation we pass from $\Sigma_n(L)$ to $\Pi_n(L)$ and vice versa. We recall also that all recursive relations are definable by $\Delta_1(L_0)$ conditions and that, the converse is also true, provided that the predicates and function symbols in L_0 name recursive entities. In addition to this we use the following classification:

An *atomic empirical formula* is one of the form $R(t_0, \ldots, t_{k-1})$ or $f(t_0, \ldots, t_{k-1}) = t_k$ where R and f are empirical symbols and each t_i is either a numeral or a variable. A *basic empirical formula* is either an atomic empirical formula or a negation of one. An *elementary empirical formula* or, for short, *elementary formula*, is a sentential combination of atomic empirical formulas.

BASIC FACT 1.1. *Every probability on L is uniquely determined by its values on the conjunctions of basic empirical sentences, a fortiori by its values on elementary sentences.*

PROOF. An elementary sentence ψ is equivalent to $\bigvee_{i<n} \psi_i$ where each ψ_i is a conjunction of basic empirical sentences and $\models \neg(\psi_i \wedge \psi_j)$ for $i \neq j$. Hence $\Pr(\psi) = \sum_{i<n}\Pr(\psi_i)$. Sentences in L_0 are either true or false, hence a sentential combination of L_0-sentences and elementary sentences is equivalent to an elementary sentence and its probability is uniquely determined. Given a $\Sigma_0(L)$-sentence, keep replacing subformulas of the forms $\exists v < n\ \varphi(v)$, $\forall v < n\ \varphi(v)$ by $\bigvee_{i<n}\varphi(i)$ and $\bigwedge_{i<n}\varphi(i)$, respectively, until all quantifiers are eliminated. The resulting sentence can now be rewritten equivalently either in the form $\exists \mathbf{v}\varphi(\mathbf{v})$, or in the form $\forall \mathbf{v}\psi(\mathbf{v})$, where φ, ψ are quantifier-free and all function symbols occur only in the form $f(t_0, \ldots, t_{n-1}) = t_n$, where each t_i is a variable or a numeral (e.g., rewrite $R(f_1(f_2(n), m), f_3(n_1, n_2), f_2(m))$ either as

$$\exists v_1, v_2, v_3, v_4[\sigma(v_1, v_2, v_3, v_4) \wedge R(v_2, v_3, v_4)]$$

or as

$$\forall v_1, v_2, v_3, v_4[\sigma(v_1, v_2, v_3, v_4) \rightarrow R(v_2, v_3, v_4)]$$

where

$$\sigma = [f_2(n) = v_1 \wedge f_1(v_1, m) = v_2 \wedge f_3(n_1, n_2) = v_3 \wedge f_2(m) = v_4]).$$

The Pr-value for the disjunctions $\varphi(\mathbf{n}_1) \vee \cdots \vee \varphi(\mathbf{n}_k)$ is uniquely determined, hence also for $\exists \mathbf{v}\varphi(\mathbf{v})$. Thus Pr is uniquely determined for $\Sigma_0(L)$ sentences and we proceed by induction on n to prove that it is uniquely determined on $\Sigma_n(L) \cup \Pi_n(L)$. Q.E.D.

DEFINITION 1.2. A probability Pr is *definable* if its restriction to elementary sentences is definable by some formula of L_0.

In this definition we treat probabilities as real-valued functions of sentences, where sentences are identified with their Gödel numbers. A real-valued function f is definable by a formula φ of L_0 if for every argument m and every n the formula defines a rational number which approximates $f(m)$ with error $\leq 1/n$. (The formal definition is given in §3.) This is the natural generalization of the concept of a recursive real-valued function (every recursive real-valued f has a $\Delta_1(L_0)$-definition). By standard techniques we can treat in L_0 rational numbers (represent, say, positive rationals as pairs of natural numbers, for negative numbers add some marker, etc.). We can add, with no loss of generality, names for all rationals, variables ranging over them and a predicate that distinguishes N from the rest of the rationals; every formula in this language is translated in the obvious way into our original L_0.

A probability Pr, as a function over *all* sentences of L, can never be defined in L_0; because for sentences φ of L_0 we have $\models\varphi \Leftrightarrow \Pr(\varphi) = 1$, implying that from Pr we get a truth definition for L_0. But Pr is uniquely determined by its restriction to elementary sentences, and we use the complexity of this restriction as a measure for the complexity of Pr. Thus in §3 we shall classify the probabilities according to the level in the arithmetic hierarchy at which the restriction is definable. For example, a Bernoulli distribution over binary sequences is obtained by taking a monadic empirical predicate P and putting

$$\Pr\left(\bigwedge_{i<k} P(n_i) \wedge \bigwedge_{i<l} \neg P(m_i)\right) = p^k(1-p)^l,$$

for all consistent conjunctions of basic empirical sentences, where $0 \leq p \leq 1$. It is easily seen that if p is a recursive real then Pr is recursive, hence $\Delta_1(L_0)$. The level at which p is definable (if at all) is also the level of Pr. Usually the probabilities encountered in statistics are recursive and, a fortiori, definable. However, one encounters certain natural families of probabilities (such as that of all Bernoulli distributions) which contain nondefinable probabilities (e.g., take in our example a nondefinable p). In §5 it is shown why nondefinable probabilities have to be taken into account.

BASIC FACT 1.2. *Let* $\Pr_0$ *be a real-valued function defined for the elementary sentences and satisfying* (1.1)(i)–(1.1)(iii). *Then* $\Pr_0$ *is extendible to a probability over L iff, for all empirical function symbols f and all* $\mathbf{n}$,

$$\operatorname{Sup}\left\{\Pr_0\left(\bigvee_{j\in J}(f(\mathbf{n}) = j)\right): J \subset N, J\text{-finite}\right\} = 1.$$

By Basic Fact 1.1 the extension is unique. If there are no empirical function symbols the condition is vacuously satisfied. That the condition is necessary is obvious from $\models\exists v(f(\mathbf{n}) = v)$. The other direction will be derived from Carathéodory's extension theorem for finitely additive measures. Since N is countable the supremum is equal to $\lim_k \Pr_0(\bigvee_{j<k}(f(\mathbf{n}) = j))$, but in the more general form the claim is true for uncountable domains. The proof given here relies, however, on the countability of N. First, we introduce some additional notation and topologize Mod_L:

If φ is a sentence in L the $\mathrm{Mod}(\varphi) =_{\mathrm{Df}} \{w \in \mathrm{Mod}_L\colon w \models \varphi\}$, if Φ is a set of sentences then $\mathrm{Mod}(\Phi) =_{\mathrm{Df}} \bigcap_{\varphi\in\Phi} \mathrm{Mod}(\varphi)$. Φ is *consistent* if $\mathrm{Mod}(\Phi) \neq \emptyset$. We topologize Mod_L by taking as a base for the open sets all sets of the form $\mathrm{Mod}(\varphi)$ where φ is an elementary sentence. Then, for every elementary φ, $\mathrm{Mod}(\varphi)$ is clopen, the open sets are of the form $\bigcup_{\varphi\in\Phi} \mathrm{Mod}(\varphi)$ and the closed ones of the form $\mathrm{Mod}(\Phi)$, where Φ is a family of elementary sentences. *If L has no empirical function symbols then* Mod_L *is compact.* [It suffices to show that for any family Φ of elementary sentences, if $\mathrm{Mod}(\Phi') \neq \emptyset$, for every finite $\Phi' \subset \Phi$ then $\mathrm{Mod}(\Phi) \neq \emptyset$. Regard each atomic empirical sentence as a different sentential variable and, by applying the compactness theorem for the sentential calculus, get an assignment of truth values which makes every sentence in Φ true. This assignment determines a world w such that $w \models R(\mathbf{n})$ iff $R(\mathbf{n})$ has value 'True'; then $w \in \mathrm{Mod}(\Phi)$.] If L has empirical function symbols Mod_L is not compact, for $\bigcap_{i\in N}(\mathrm{Mod}(f(\mathbf{n}) \neq i)) = \emptyset$ but every finite subset has a nonempty intersection. The family $\{\mathrm{Mod}(\varphi)$: φ a sentence in $L\}$ is a field, i.e. closed under complementation and finite unions. We let $\mathscr{B}$ = *σ-field generated by this field* (where a σ-field is a field closed under countable unions). By a *probability over* $\mathscr{B}$ we mean a σ-additive normalized measure defined over $\mathscr{B}$.

BASIC FACT 1.3. *Every probability* Pr *on L determines a unique probability* Pr* *on* $\mathscr{B}$ *such that* $\Pr^*(\mathrm{Mod}(\varphi)) = \Pr(\varphi)$ *for all sentences.*

This fact also remains true in the case of uncountable domains. We prove 1.2 and 1.3 together:

Let $\Pr_0$ be as in 1.2. Put $\mathscr{B}_0 = \{\mathrm{Mod}(\varphi)$: φ elementary sentence$\}$ and let $\Pr_0^*(\mathrm{Mod}(\varphi)) = \Pr_0(\varphi)$. Since $\Pr_0$ satisfies (1.1)(i)–(1.1)(iii), $\Pr_0^*$ is a well-defined, finitely additive measure on $\mathscr{B}_0$. By Carathéodory's theorem $\Pr_0^*$ is extendible to a σ-additive measure on the generated σ-field iff for all $\{A_n: \mathbf{n} \in N\} \subset \mathscr{B}_0$, $\bigcap_n A_n = \emptyset$ implies $\Pr_0^*(\bigcap_{n<k} A_n) \to_k 0$. If L has no empirical function symbols, Mod_L is compact; $\bigcap_n A_n = \emptyset$ implies that, for some k, $\bigcap_{n<k} A_n = \emptyset$. Hence the condition holds trivially. Since $\mathrm{Mod}(\exists v\varphi(v)) = \bigcup_{n\in N}\mathrm{Mod}(\varphi(n))$, the σ-field generated by $\mathscr{B}_0$ includes every $\mathrm{Mod}(\varphi)$, where φ is a sentence of L, hence it is $\mathscr{B}$. (Here the countability of N is used.) Let Pr* be the σ-additive measure on $\mathscr{B}$ extending $\Pr_0^*$; put $\Pr(\varphi) =_{\mathrm{Df}} \Pr^*(\mathrm{mod}(\varphi))$. It is easily seen that Pr is a probability on L (the σ-additivity of Pr* and the fact that N is countable imply (1.1)(iv)). This argument also proves 1.3. If there are empirical function symbols let $\tilde{L}$ be the language obtained by replacing every k-ary empirical function symbol f by a $(k + 1)$-ary empirical predicate R_f. Every model w of L can be regarded as a model for $\tilde{L}$ by interpreting each R_f as $\{(\mathbf{n}, m): w \models f(\mathbf{n}) = m\}$. In this way Mod_L can be identified with a subset X of $\mathrm{Mod}_{\tilde{L}}$. It is easily seen that X consists of all models that satisfy all the sentences $\exists v R_f(\mathbf{n}, v)$ and $R_f(\mathbf{n}, m) \wedge R_f(\mathbf{n}, m') \to m = m'$, where f is any empirical k-ary function symbol, $\mathbf{n} \in N^k$ and $m, m' \in N$. Define $\tilde{\Pr}_0$ for the elementary sentences of $\tilde{L}$ as follows: Given an elementary sentence ψ of $\tilde{L}$ let ψ' be the sentence obtained by replacing every $R_f(\mathbf{n}, m)$ by $f(\mathbf{n}) = m$ and let $\tilde{\Pr}_0(\psi) =_{\mathrm{Df}} \Pr_0(\psi')$. As is easily seen $\Pr_0$ satisfies (1.1)(i)–(1.1)(iii). By the claim for languages without empirical function symbols, we get a probability $\tilde{\Pr}$ on $\tilde{L}$ and a corresponding probability $\tilde{\Pr}^*$ over $\tilde{\mathscr{B}}$ (the σ-field generated by $\{\mathrm{Mod}(\varphi)$: φ a sentence of $\tilde{L}\}$).

$$\tilde{\Pr}(R_f(\mathbf{n}, m) \wedge R_f(\mathbf{n}, m') \to m = m')$$
$$= \Pr_0(f(\mathbf{n}) = m \wedge f(\mathbf{n}) = m' \to m = m') = 1.$$

Also

$$\widetilde{\Pr}(\exists v R_f(\mathbf{n}, v)) = \mathrm{Sup}\{\widetilde{\Pr}_0(\textstyle\bigvee_{j\in J} R_f(\mathbf{n}, j)) : J \subset N, J\text{-finite}\}$$
$$= \mathrm{Sup}_J \Pr_0(\textstyle\bigvee_{j\in J} f(\mathbf{n}) = j) = 1.$$

This implies that $\widetilde{\Pr}^*(X) = 1$. Every $A \in \mathscr{B}$ is a subset of X, and if we put $\Pr^*(A) =_{\mathrm{Df}} \widetilde{\Pr}^*(A)$ we get a probability over $\mathscr{B}$ which extends $\Pr_0^*$. Q.E.D

We shall always denote by 'Pr*' *the probability over* $\mathscr{B}$ *determined by* Pr. Actually, we can identify the two and regard Pr as a probability over $\mathscr{B}$. Note that $\mathscr{B}$ consists of the Borel sets of Mod_L and that Mod_L is what is known as a Polish space, i.e., a separable space that can be provided with a metric under which it is complete. (Enumerate in one sequence all sentences $R(\mathbf{n})$ and all terms $f(\mathbf{n})$ where R and f range over all empirical symbols. If $w_1, w_2 \in \mathrm{Mod}_L$ and $w_1 \neq w_2$, then for some i, the ith member of the sequence has, in w_1 and w_2, either different truth values or different numerical values. Let k be the smallest i satisfying this and put $d(w_1, w_2) = (k + 1)^{-1}$.)

As is well known, $\Pr^*(A)$ is equal to the infimum of $\Pr^*(B)$, where B ranges over all open sets containing A, as well as to the supremum of all $\Pr^*(C)$, where C ranges over closed subsets included in A.

Recall that a measure μ_2 is *absolutely continuous* with respect to μ_1 if for all A's in our σ-field $\mu_1(A) = 0 \Rightarrow \mu_2(A) = 0$. We therefore define:

If $\Pr_i$, $i = 1, 2$, are probabilities over L then $\Pr_2$ is *absolutely continuous with respect to* $\Pr_1$ if $\Pr_2^*$ is absolutely continuous with respect to $\Pr_1^*$. We denote it $\Pr_2 \prec \Pr_1$. We say that $\Pr_2$ *and* $\Pr_1$ *are compatible* if $\Pr_2 \prec \Pr_1$ and $\Pr_1 \prec \Pr_2$.

Evidently $\Pr_2 \prec \Pr_1$ implies that, for all sentences, $\Pr_1(\varphi) = 0 \Rightarrow \Pr_2(\varphi) = 0$. As explained in the introduction this means that $\Pr_2$ is at least as dogmatic as $\Pr_1$. Compatibility implies that they are equally dogmatic. In general, the condition that for every sentence φ, $\Pr_1(\varphi) = 0 \Rightarrow \Pr_2(\varphi) = 0$ is weaker than $\Pr_2 \prec \Pr_1$. But it will be shown in §3 that if $\Pr_i$, $i = 1, 2$, are definable, then the two conditions are equivalent. Thus, for definable probabilities, compatibility means exactly the sharing of the same a priori commitments. Without assuming the definability of the $\Pr_i$ we can characterize $\Pr_2 \prec \Pr_1$ by the condition:

For every sequence $\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots$ *of elementary sentences*, $\Pr_1(\bigwedge_{i<n}\varphi_i) \to_n 0$ *implies* $\Pr_2(\bigwedge_{i<n}\varphi_i) \to_n 0$.

(This follows from the fact that every $\Pr^*(A)$ is the supremum of $\Pr^*(B)$, where B ranges over closed subsets of A.)

Another characterization is:

For every sequence of elementary sentences $\mathrm{Inf}_n \Pr_1(\varphi_n) = 0$ *implies* $\mathrm{Inf}_n \Pr_2(\varphi_n) = 0$.

(This condition evidently implies the previous one. That it is necessary follows from standard measure theory.)

If T is a consistent set of sentences we say that Pr *is a probability with respect to the theory* T if $\Pr(\varphi) = 1$ for all $\varphi \in T$. The idea is that T is an empirical presupposed theory. For example assume a repeated experiment whose outcome, at each repetition, is known a priori to be a natural number $\leq m$ (e.g., the turning of a roulette, $m = 36$). If $f(n)$ is the nth outcome then T will contain the sentence

$\forall v(f(v) \leq m)$. Other examples in which a background theory arises in a natural way have been indicated in the introduction.

The framework can be generalized to uncountable domains as follows. Consider an arbitrary countable mathematical language L_0 with some fixed "standard" model M_0 not necessarily countable. Expand L_0 by adding to it names for all members of M_0, then add to it empirical symbols (whose interpretation is not fixed). The probability is to be defined on the language obtained in this way. In [Ga 2] L_0 was taken as pure logic and M_0 as an arbitrary set. In [S-K] Scott and Krauss have shown that this framework works if we consider L_{ω_1,ω_0}-logic instead of the first-order one. In the present paper L_0 is a first-order arithmetical language and M_0 its standard model. We can go further and consider a second-order standard structure; for example, N and all its subsets, where L_0 has variables ranging over subsets of N and a formula $v \in u$ asserting that the number v is a member of u. But then, in order for our Basic Fact 1.2 to continue to hold one should, in general, change the definition of validity by including in Mod_L also some "nonstandard" models. For instance, in the case of second-order arithmetic just mentioned, we can take as Mod_L all the submodels of standard models whose domain forms an elementary substructure of M_0 (with respect to L_0). When the domain is uncountable, $\mathscr{B}$ is not σ-generated by $\{\text{Mod}(\varphi)\colon \varphi\text{-elementary}\}$ and requirement (1.1)(iv) of the definition does not reduce to the σ-additivity of the measure; thus, not every σ-additive measure on $\mathscr{B}$ is of the form Pr*. Our Basic Fact 1.2 leads, for uncountable domains, to measures that do not arise in the standard measure-theoretic constructions. It is not clear to what extent such probabilities, in the case, say, of second-order arithmetic, correspond to natural probabilities that arise in probability theory and statistics. (A certain connection has been pointed out by Scott and Krauss in [S-K].)

§2. **Increasing evidence and testing procedures.** The increase of empirical knowledge can be represented as a process in which the truth or falsity of an increasing number of sentences becomes known. So imagine a sequence of sentences φ_0, φ_1, $\ldots$, φ_n, $\ldots$ and suppose that at stage n we know a conjunction of the form $\bigwedge_{i<n} \pm \varphi_i$ to be true, where $\pm\varphi_i$ is either φ_i or its negation. At that stage we have, instead of the initial probability, Pr, the conditional probability $\Pr(\psi | \bigwedge_{i<n} \pm \varphi_i)$. We shall investigage how the conditional probability behaves, as n increases. Instead of sequences $\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots$, we could have considered sequences $X_0, X_1, \ldots, X_n, \ldots$ of sets in $\mathscr{B}$ (i.e., Borel subsets of Mod_L) and the behaviour of $\Pr^*(Y | \bigcap_{i<n} \pm X_i)$, where $\pm X_i$ is either X_i or its complement. The claims of this section are general measure-theoretic and do not depend essentially on the linguistic structure. (Theorem 2.1(I) is an immediate consequence of the properties of Polish spaces and the martingale convergence theorem, and (II) is proved by general probabilistic techniques. To our knowledge it has not been stated before.) But the formulation in terms of a language is extremely suggestive and bears directly on the philosophical question of learning from experience. When an initial global point of view is represented by a probability over a comprehensive language, the process of learning is reflected in the conditional probability and the statements of this section constitute answers, of a sort, to two questions: Are we sure that increasing

evidence will, in the long run, reveal the truth and that it will bring together initially different points of view? These answers are not "absolute", but "inner answers", like the law of large numbers. Philosophers who are acquainted with this law and with its significance are, in general, unaware that by using more sophisticated mathematics the same type of answers can be provided on a global scale, for a whole language and for arbitrary probabilities. Later in §5 the linguistic structure becomes essential, when we consider definable probabilities, definable testing procedures and tie the results of this section with the concept of randomness.

For every world w and every sentence φ let $\varphi^{(w)}$ be either φ or $\neg\varphi$ according as $w \models \varphi$ or not. If the accumulation of data is represented by the sequence φ_0, $\varphi_1, \ldots, \varphi_n, \ldots$, then, at the nth stage, the evidence in the world w is $\bigwedge_{i<n}\varphi_i^{(w)}$. If ψ is some hypothesis then for its truth or falsity to be revealed in w in the long run means that $\Pr(\psi | \bigwedge_{i<n}\varphi_i^{(w)})$ tends, as $n \to \infty$, to 1 if $w \models \psi$, to 0 otherwise. One should think of the φ_i's as relatively simple sentences whose truth values can be found by available testing methods, whereas ψ cannot be tested directly in this way. In Carnap's framework these were taken as the atomic empirical sentences. Here we do not fix any particular sequence of φ_i's and the results are valid in general. The basic property of the φ_i's that we need here is that of *separation*. It is defined as follows:

Say that a sentence φ separates the two worlds w_1, w_2 if it is true in one of them and false in the other. *A class Φ of sentences separates a set X of worlds* if every two different worlds in X are separated by some $\varphi \in \Phi$. *A class Φ is separating* if it separates Mod_L. This is the obvious correlate of the well-known notion of separation by a family $\{X_i : i \in I\}$ of sets, or by a family $\{F_i : i \in I\}$ of functions; namely, the *family separates a set X* if for every two different $x_1, x_2 \in X$ there exists $i \in I$ such that X_i contains one and not the other, or, in the case of functions, such that $F_i(x_1) \neq F_i(x_2)$. Evidently the class of all atomic empirical sentences is separating.

THEOREM 2.1. *Let $\{\varphi_i : i = 0, 1, \ldots\}$ be separating. Let $[\psi](w)$ be defined as* 1 *if $w \models \psi$, as* 0 *otherwise. Then*:

(I) *For every ψ, $\Pr(\psi | \bigwedge_{i<n}\varphi_i^{(w)}) \to_n [\psi](w)$ almost everywhere.*

(II) *If $\Pr_1$ is compatible with* Pr *then*

$$[\mathrm{Sup}_\psi |\Pr_1(\psi | \bigwedge_{i<n}\varphi_i^{(w)}) - \Pr(\psi | \bigwedge_{i<n}\varphi_i^{(w)})|] \to_n 0$$

almost everywhere.

Here "*almost everywhere*" means that the statement in question is true for all worlds belonging to some set of probability 1. We shall abbreviate it as "*a.e.*". Note that in (II) "almost everywhere" is the same with respect to Pr as with respect to $\Pr_1$, since the sets of probability 1 are the same for the two probabilities. The supremum in (II) is taken over all ψ in the language. For each particular ψ, (I) implies that

$$|\Pr_1(\psi | \bigwedge_{i<n} \varphi_i^{(w)}) - \Pr(\psi | \bigwedge_{i<n} \varphi_i^{(w)})| \to_n 0,$$

but this is weaker than (II), because it does not imply a convergence uniform in ψ.

(II) means that for every sequence $\psi_0, \ldots, \psi_n, \ldots,$

$$[\mathrm{Pr}_1(\psi_n | \bigwedge_{i<n} \varphi_i^{(w)}) - \mathrm{Pr}(\psi_n | \bigwedge_{i<n} \varphi_i^{(w)})] \to 0;$$

but the limit of $\mathrm{Pr}(\psi_n | \bigwedge_{i<n} \varphi_i^{(w)})$ need not exist. For example, let P be an empirical monadic predicate and let $\psi_n = \varphi_n = P(n)$. Then

$$[\mathrm{Pr}_1(P(n) | \bigwedge_{i<n} P(i)^{(w)}) - \mathrm{Pr}(P(n) | \bigwedge_{i<n} P(i)^{(w)})] \to 0$$

but $\mathrm{Pr}(P(n) | \bigwedge_{i<n} P(i)^{(w)})$ need not tend to a limit; thus, if the evidence is $P(i)$, for all even i's $< n$ and $\neg P(i)$, for all odd i's $< n$, then, for reasonable Pr's, it is to be expected that the conditional probability will approach 1 or 0 according as n is even or odd. (II) implies that no matter how the conditional probabilities behave separately they will, with probability 1, come arbitrarily close as $n \to \infty$. In other words, with probability 1, two persons holding mutually nondogmatic initial views will, in the long run, judge similarly. Still, one cannot without further assumptions give any estimate how long is "in the long run" (cf. [Ga 3] for a discussion of this point in connection with Goodman's puzzle). Also the convergence is guaranteed with probability 1, where "probability" refers to the presupposed prior. (I) and (II) form an "inner justification" but they do not constitute a justification of the particular prior.

PROOF OF THEOREM 2.1. (I) Recall that if F is an integrable function defined over a probability space $(\Omega, \mathscr{B}, \mu)$ and if $\mathscr{D}$ is a σ-field, $\mathscr{D} \subset \mathscr{B}$, then $E(F|\mathscr{D})$ is the conditional expectation of F, given $\mathscr{D}$; this is a $\mathscr{D}$-measurable function such that $\int_A F d\mu = \int_A E(F|\mathscr{D}) d\mu$ for all $A \in \mathscr{D}$.

Such a function exists by the Radon-Nikodým theorem and is determined uniquely almost everywhere (i.e., every two functions satisfying the condition coincide a.e.). Recall also that if $\{B_i : i \in N\}$ is a partition of Ω into mutually disjoint sets and $\mathscr{D}$ is the σ-field generated by it, then the $\mathscr{D}$-measurable functions are those that are constant over each B_i. In this case $\int_{B_i} F d\mu = E(F|\mathscr{D})(w) \cdot \mu(B_i)$, for all $w \in B_i$. For the case $F = \chi_A$, where $\chi_A(w)$ is 1 or 0, according as $w \in A$ or not, one gets

$$\mu(A \cap B_i) = \int_{B_i} \chi_A d\mu = E(\chi_A|\mathscr{D})(w) \cdot \mu(B_i), \quad \text{for } w \in B_i.$$

Hence, if $\mu(B_i) > 0$, we have

$$E(\chi_A|\mathscr{D})(w) = \mu(A \cap B_i)/\mu(B_i) \quad \text{for all } w \in B_i.$$

If $\mathscr{D}_0 \subset \mathscr{D}_1 \subset \cdots \subset \mathscr{D}_n \cdots$, where each $\mathscr{D}_i$ is a σ-field and $\bigcup_n \mathscr{D}_n$ generates our $\mathscr{B}$ then $E(F|\mathscr{D}_n) \to F$ a.e.

This is a version of Doob's martingale convergence theorem. (It is the only form in which we shall use this theorem. For definition and details cf. [Br, chapter 5].) Another basic fact which we shall use is the following:

In a Polish space, a family of Borel sets which separates the set of all points generates the σ-field of all Borel sets (*cf.* [Chr]).

Now let $\mathscr{D}_n$ = field generated by $\{\mathrm{Mod}(\varphi_i): i < n\}$. Then $\mathscr{D}_n$ is a finite field (hence σ-field) generated by the partition whose members are the sets of the form

$\mathrm{Mod}(\bigwedge_{i<n} \pm\varphi_i)$. For every $w \in \Omega$, $\mathrm{Mod}(\bigwedge_{i<n}\varphi_i^{(w)})$ is the member of the partition to which w belongs. Given a sentence ψ the characteristic function of $\mathrm{Mod}(\psi)$ is $[\psi]$. It follows that if $\Pr(\bigwedge_{i<n}\varphi_i^{(w)}) > 0$, then

$$E([\psi]\|\mathscr{D}_n)(w) = \Pr(\psi|_{i<n}\varphi_i^{(w)}).$$

Since the union of all sets $\mathrm{Mod}(\bigwedge_{i<n} \pm\varphi_i)$ of probability 0, $n = 0, 1, 2, \ldots$, is of probability 0, the last equality holds a.e. Since $\{\varphi_i: i \in N\}$ separates Ω, $\bigcup_n\mathscr{D}_n$ generates $\mathscr{B}$; hence $E([\psi]\|\mathscr{D}_n)(w) \to [\psi](\mathrm{w})$ a.e. This implies (I).

(II) We shall use the following standard fact concerning conditional expectations: If f_1 is $\mathscr{D}$-measurable then $E(f_1 \cdot f_2|\mathscr{D}) = f_1 \cdot E(f_2|\mathscr{D})$ a.e. (cf. [Br, §25.2]). $\Pr_1 \prec \Pr$ we have, by the Radon-Nikodým theorem, a nonnegative function g such that $\Pr_1^*(A) = \int_A g\, d\Pr^*$ for all $A \in \mathscr{B}$ and, as is well known, $\int h\, d\Pr_1^* = \int hg\, d\Pr^*$ for every integrable function h. Let $E(\ |\ )$ and $E_1(\ |\ )$ be the conditional expectations for the probabilities $\Pr^*$ and $\Pr_1^*$, respectively. Then, with $\mathscr{D}_n$ as in (I), we have

$$\Pr_1(\psi|\bigwedge_{i<n} \varphi_i^{(w)}) - \Pr(\psi|\bigwedge_{i<n} \varphi_i^{(w)}) = E_1([\psi]\|\mathscr{D}_n) - E([\psi]\|\mathscr{D}_n).$$

Consequently it suffices to find an upper bound of $E_1([\psi]\|\mathscr{D}_n) - E([\psi]\|\mathscr{D}_n)$ which is independent of ψ and which converges to 0 as $n \to \infty$. Another basic fact is

$$E_1([\psi]\|\mathscr{D}_n) \cdot E(g|\mathscr{D}_n) = E([\psi] \cdot g|\mathscr{D}_n) \quad \text{a.e.}$$

(For, since $E_1([\psi]\|\mathscr{D}_n)$ is $\mathscr{D}_n$-measurable, we have

$$E_1([\psi]\|\mathscr{D}_n) \cdot E(g|\mathscr{D}_n) = E((E_1([\psi]\|\mathscr{D}_n) \cdot g)|\mathscr{D}_n) \text{ a.e.}$$

For any $A \in \mathscr{D}_n$ the integral of this over A, with respect to $\Pr^*$ turns out to be $\int_A[\psi]d\Pr_1^*$, which is also the integral of the right-hand side.) Consequently $E_1([\psi]\|\mathscr{D}_n) - E([\psi]\|\mathscr{D}_n)$ is equal to

$$(E(g|\mathscr{D}_n))^{-1}(E([\psi] \cdot g|\mathscr{D}_n) - E([\psi]\|\mathscr{D}_n) \cdot E(g|\mathscr{D}_n)).$$

Since $\Pr^* \prec \Pr_1^*$, we have $g(w) > 0$ a.e. Also $E(g|\mathscr{D}_n) \to g$ a.e. Therefore it suffices to show that

$$|(E([\psi] \cdot g|\mathscr{D}_n) - E([\psi]\|\mathscr{D}_n) \cdot E(g|\mathscr{D}_n))| < \xi_n \to 0 \quad \text{a.e.}$$

where ξ_n does not depend on ψ. Now put $[\psi] \cdot g = [\psi] \cdot E(g|\mathscr{D}_n) + [\psi](g - E(g|\mathscr{D}_n))$. Applying $E(\ |\mathscr{D}_n)$ to both sides and using the fact that $E(g|\mathscr{D}_n)$ is $\mathscr{D}_n$-measurable, we get

$$\begin{aligned} &|E([\psi] \cdot g|\mathscr{D}_n) - E([\psi]\|\mathscr{D}_n) \cdot E(g|\mathscr{D}_n)| \\ &\qquad \le E(|[\psi] \cdot (g - E(g|\mathscr{D}_n))|\ |\mathscr{D}_n) \le E(|g - E(g|\mathscr{D}_n)|\ |\mathscr{D}_n). \end{aligned}$$

(Here we also used the linearity of $E(\ |\ )$ and the basic properties $E(h|\mathscr{D}) \le E(|h|\ |\mathscr{D})$ and $h \le h' \Rightarrow E(h|\mathscr{D}) \le E(h'|\mathscr{D})$ a.e.) It suffices to show that

$$E(g - E(g|\mathscr{D}_n)\ |\mathscr{D}_n) \xrightarrow[n]{} 0 \quad \text{a.e.}$$

If $\int g^2 d\mathrm{Pr}^* < \infty$, use the Cauchy-Schwartz inequality for conditional expectations $E^2(f|\mathscr{D}) \leq E(f^2|\mathscr{D})$ (obtainable from $0 \leq E((f - E(f|\mathscr{D}))^2|\mathscr{D}) = E(f^2|\mathscr{D}) - E^2(f|\mathscr{D})$) and get

$$E^2(|g - E(g|\mathscr{D}_n)|\,|\mathscr{D}_n) \leq E((g - E(g|\mathscr{D}_n))^2|\mathscr{D}_n) = E(g^2|\mathscr{D}_n) - E^2(g|\mathscr{D}_n) \to 0 \quad \text{a.e.}$$

(because $E(g^2|\mathscr{D}_n) \to g^2$, $E(g|\mathscr{D}_n) \to g$ a.e.). If $\int g^2 d\mathrm{Pr}^* = \infty$ let g_k be the k-truncation of g, i.e., $g_k(w) = g(w)$ if $g(w) < k$ and $g_k(w) = 0$ otherwise. Then

$$\begin{aligned}|g - E(g|\mathscr{D}_n)| &\leq |g_k - E(g_k|\mathscr{D}_n)| + |g - g_k| + |E(g_k|\mathscr{D}_n) - E(g|\mathscr{D}_n)| \\ &\leq |g_k - E(g_k|\mathscr{D}_n)| + |g - g_k| + E(|g - g_k|\,|\mathscr{D}_n).\end{aligned}$$

By the previous argument $E(|g_k - E(g_k|\mathscr{D}_n)|\,|\mathscr{D}_n) \to_n 0$ a.e. Also

$$E(E(|g - g_k|\,|\mathscr{D}_n)|\mathscr{D}_n) = E(|g - g_k|\,|\mathscr{D}_n) \quad \text{and}$$
$$E(|g - g_k|\,|\mathscr{D}_n) \to |g - g_k| \quad \text{a.e.}$$

Put $\eta_{k,n} = E(|g_k - E(g_k|\mathscr{D}_n)|\,|\mathscr{D}_n)$. Then

$$E(|g - E(g|\mathscr{D}_n)|\,|\mathscr{D}_n) \leq \eta_{k,n} + 2E(|g - g_k|\,|\mathscr{D}_n) \to_n 2\cdot|g - g_k| \quad \text{a.e.}$$

Since the left-hand side does not depend on k and for every w there exists k such that $g(w) - g_k(w) = 0$, the left-hand side tends to 0 a.e. Q.E.D.

As remarked before, the proof is valid in the general case where $\{\mathrm{Mod}(\varphi_i): i \in N\}$ is replaced by an arbitrary separating family $\{X_i: i \in N\}$ of members of $\mathscr{B}$, and where $\mathrm{Mod}(\psi)$ is replaced by Y where Y ranges over $\mathscr{B}$. Actually, the general case is derivable from the theorem as stated, using the fact that given any countable family $\{Z_i: i \in N\}$ of members of $\mathscr{B}$, we can enlarge L_0 by a mathematical predicate so that in the enriched language we have a formula $\lambda(v)$ such that $Z_n = \mathrm{Mod}(\lambda(n))$ for all n. (This is done using well-known encoding techniques, by which sequences of Borel sets are encoded into binary sequences.)

If $X \in \mathscr{B}$ is any set such that $\mathrm{Pr}^*(X) = 1$ and if $\{\varphi_i: i \in N\}$ separates X, then we can carry out a proof of Theorem 2.1 in which Mod_L is replaced by X. (With every sentence σ we associate $\mathrm{Mod}'(\sigma)$, defined as $\mathrm{Mod}(\sigma) \cap X$. Then $\{\mathrm{Mod}'(\varphi_i): i \in N\}$ separates X, hence it σ-generates $\mathscr{B}'$, where $\mathscr{B}'$ is the σ-field generated by all $\mathrm{Mod}'(\sigma)$ where σ ranges over all atomic empirical sentences and where complementation is with respect to X. Since $\mathrm{Pr}^*(X) = 1$, the measure, say Pr^{**}, induced by Pr^* on $\mathscr{B}'$ satisfies $\mathrm{Pr}^{**}(\mathrm{Mod}'(\sigma)) = \mathrm{Pr}(\sigma)$.)

Say that Φ is *a.e. separating* (with respect to a given probability) if it separates a set of probability 1.

The observation just made proves

THEOREM 2.1′. *In Theorem* 2.1 *we can replace the condition that* $\{\varphi_i: i \in N\}$ *is separating by the condition that it is a.e. separating.*

Testing procedures. Evidence need not be collected in a predetermined order. The knowledge of the truth values of $\varphi_0, \ldots, \varphi_{n-1}$ may affect the choice of the next sentence φ_n to be examined. The procecure by which the choice is performed is representable by a function t as follows. If φ is the sentence describing evidence so far collected, then $t(\varphi)$ is the sentence to be tested next, after which testing, the evidence will be either $\varphi \wedge t(\varphi)$ or $\varphi \wedge \neg t(\varphi)$. This motivates the following definition and notation:

A *testing procedure* is a function t from sentences to sentences. t yields, for every world w, a sequence of sentences $\varphi(w, n)$ where $\varphi(w, 0)$ is the sentence $0 = 0$. $\varphi(w, n) = t(\bigwedge_{i<n}\varphi(w, i))$ if w satisfies this sentence, $\varphi(w, n) = \neg t(\bigwedge_{i<n}\varphi(w, i))$ otherwise.

We associate with t the *protocol function* $\bar{t}$ defined by

$$\bar{t}(w, n) = \bigwedge_{i<n}\varphi(w, i).$$

Our definition of testing procedure is sufficiently wide to accommodate all statistical testing practices. For example, consider a case of a test based on random samplings. Say individuals are picked by lottery and are examined with respect to having a certain property—named by the predicate P; e.g., '$P(x)$' stands for 'x approves of the president's policy'. For the sake of simplicity assume all individuals to be arranged in a sequence $0, 1, \ldots, n, \ldots$ and that for each i a random device, such as the outcome of a toss of a coin, determines whether i is included in the sample. To reflect this situation we include in the language an empirical predicate R such that $R(i)$ is to read 'the ith toss yields a 'yes'.' Now consider the following procedure:

(i) $t(0 = 0) = R(0)$.

(ii) Let φ be a conjunction of basic empirical sentences and let n be the largest i occurring in it. Then $t(\varphi) = P(n)$, provided that $R(n)$ occurs in φ as a conjunct and neither $P(n)$ nor $\neg P(n)$ occur in φ. In all other cases, $t(\varphi) = R(n + 1)$.

It is easily seen that this t is, in fact, the random sampling procedure. The assumptions concerning the randomness of the choosing device are to be expressed through the prior probability Pr.

A testing procedure t *separates the worlds* w_1, w_2 if there exists n such that $t(w_1, n) \neq t(w_2, n)$. *It separates a set* X if it separates every two different worlds in X. It is *separating* if it separates Mod_L and it is *a.e. separating* (with respect to some given probability) if it separates a set of probability 1.

The information yielded in the world w by the first n testings of the prodedure t is represented by $\bar{t}(w, n)$. The testing of a fixed sequence $\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots$ is a special procedure such that $t(0 = 0) = \varphi_0$ and $t(\bigwedge_{i<n}\pm\varphi_i) = \varphi_n$. Theorem 2.1′ generalizes to all testing procedures:

THEOREM 2.2. *Let t be an a.e. separating testing procedure. Then*:

(I) *For every* ψ, $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$ *a.e.*

(II) *If* $\Pr_1$ *is compatible with* Pr *then*

$$\mathrm{Sup}_\psi|\Pr_1(\psi|\bar{t}(w, n)) - \Pr(\psi|\bar{t}(w, n))| \to 0 \quad a.e.$$

PROOF. We reduce the case of an arbitrary procedure to that of a fixed sequence of sentences. With every finite binary sequence s associate a sentence σ_s as follows:

$\sigma_\varnothing = (0 = 0)$, $\sigma_{s^\frown\langle x\rangle} = \sigma_s \wedge t(\sigma_s)$ if $x = 1$,

$\sigma_{s^\frown\langle x\rangle} = \sigma_s \wedge \neg t(\sigma_s)$ if $x = 0$.

Obviously, for every world w, $\bar{t}(w, k) = \sigma_{x_1\ldots x_k}$, where x_i is 1 or 0 according as $w \models t(\bar{t}(w, i - 1))$ or not, $i = 1, \ldots, k$. (For $k = 0$, $\bar{t}(w, 0) = \sigma_\varnothing$.) Put

$$\varphi_k = \bigvee_{s\in\{0,1\}^k} (\sigma_s \wedge t(\sigma_s)) \qquad (\varphi_0 = (0 = 0 \wedge t(0 = 0))).$$

For each k, every w determines unique $s \in \{0, 1\}^k$ such that $w \models \sigma_s$. Consequently, $w \models \varphi_k \Leftrightarrow w \models t(\bar{t}(w, k))$. Letting $k = 0, \ldots, n - 1$, we find that $\bigwedge_{k<n} \varphi_k^{(w)}$ is equivalent to $\bar{t}(w, n)$ (where $\varphi_k^{(w)}$ is defined as in Theorem 2.1). Hence, in every world, the information yielded by t is the same as that obtained by testing the fixed sequence $\varphi_0, \ldots, \varphi_n, \ldots$. Also, t is a.e. separating iff $\{\varphi_0, \ldots, \varphi_n, \ldots\}$ is. Now apply Theorem 2.1′. Q.E.D.

Though, in principle, testing procecures are reducible to testing fixed sequences of sentences, the notion of a testing procedure is much more convenient. For, usually, the sentences of the fixed sequence will be complex and considerably less transparent than those obtained by applying the procedure. Just so, the notion of a game of many moves is very useful, though, in principle, it reduces to the one-move game in which each player chooses a strategy.

Testing for partial knowledge. Testing procedures have been so far considered as providing, in the limit, complete knowledge of the world. Consequently the values $\Pr(\psi|\bar{t}(w, n))$ could be used to test any sentence ψ. If, however, we are interested only in a certain restricted type of information then we can use tests which are not exhaustive. This in fact is the case in statistical testing. For example, one wishes to estimate the limit frequency of a certain empirical predicate P, i.e., the limit as $n \to \infty$ of k_n/n, where k_n is the cardinality of $\{i < n: w \models P(i)\}$ (say, the a priori probability Pr is such that the existence of the limit is accorded probability 1). In this case we need not use a procedure which separates a.e., it suffices that we separate a.e. between every two worlds with different limiting frequencies; thus, a procedure based on random sampling, as described before, will do. These considerations lead to the following terminology:

Let Pr be a fixed probability on L and let Φ_1, Φ_2 be two families of sentences, Φ_1 *separates not less than* Φ_2 if there is a set $X \subset \mathrm{Mod}_L$ of probability 1 such that every $\{w_1, w_2\} \subset X$ which is separated by Φ_2 is separated by Φ_1. This relation is evidently reflexive transitive and determines an equivalence relation: Φ_1 and Φ_2 are *equivalent* if each separates not less than the other.

LEMMA 2.3. *Let Φ be a family of sentences and let $\sigma\Phi$ be the family consisting of all sentences ψ such that* $\mathrm{Mod}(\psi)$ *is equal, modulo a set of measure* 0, *to a set σ-generated by* $\{\mathrm{Mod}(\varphi): \varphi \in \Phi\}$. *Then* (i) *$\sigma\Phi$ is equivalent to Φ*; (ii) *$\Phi' \subset \sigma\Phi$ for every Φ' which is equivalent to Φ*; (iii) *Φ_1 separates not less than Φ_2 iff $\sigma\Phi_2 \subset \sigma\Phi_1$*.

PROOF. (i) Evidently $\Phi \subset \sigma\Phi$; hence $\sigma\Phi$ separates not less then Φ. Since altogether there are countably many sentences, we get a set $X \subset \mathrm{Mod}_L$ of probability 1 such that, for each $\psi \in \sigma\Phi$, $\mathrm{Mod}(\psi) \cap X$ is in the algebra of subsets of X σ-generated by $\{\mathrm{Mod}(\varphi) \cap X: \varphi \in \Phi\}$. This implies that every $\{w_1, w_2\} \subset X$ separated by $\sigma\Phi$ is separated by Φ. (ii) follows from (iii), since we get $\Phi' \subset \sigma\Phi' \subset \sigma\Phi$. (iii) follows from the basic fact that if $\mathscr{F}_1$, $\mathscr{F}_2$ are countable families of Borel sets in a Polish space and $\mathscr{F}_1$ separates all pairs separated by $\mathscr{F}_2$, then the σ-field generated by $\mathscr{F}_1$ includes the one generated by $\mathscr{F}_2$. Q.E.D.

We are now in position to generalize Theorem 2.2 to the case of testing for partial knowledge:

THEOREM 2.4. *Let t be a testing procedure and* Ψ *a class of sentences such that, in some set of worlds of probability* 1, *t separates every pair which is separated by* Ψ. *Then*:

(I) *For all* $\psi \in \Psi$, $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$ *a.e.*

(II) *If* $\Pr_1$ *is compatible with our presupposed* Pr *then*

$$\operatorname{Sup}_{\psi\in\Psi} |\Pr_1(\psi|\bar{t}(w, n)) - \Pr(\psi|\bar{t}(w, n))| \underset{n}{\to} 0 \quad a.e.$$

PROOF. By the proof of Theorem 2.2 we can reduce the case of an arbitrary t to that of a fixed sequence $\{\varphi_0, \ldots, \varphi_n, \ldots\}$, where $\bar{t}(w, n)$ is replaced by $\bigwedge_{i<n}\varphi_i^{(w)}$. Let $\Phi = \{\varphi_0, \ldots, \varphi_n, \ldots\}$, then Φ separates no less than Ψ. Let $\sigma\Phi$ be defined as in Lemma 2.3. The arguments used in the proofs of Theorems 2.1 and 2.1′ can be applied to show (I) and (II) of 2.1′, provided that in (I) $\psi \in \sigma\Phi$, and in (II) the supremum is taken over all $\psi \in \sigma\Phi$. (One replaces the Borel field $\mathscr{B}$ by the σ-field $\mathscr{B}'$ generated by all the sets $\operatorname{Mod}(\varphi)$, $\varphi \in \Phi$.) Since $\Psi \subset \sigma\Phi$ the claim follows. Q.E.D.

Here is another generalization which we state without giving the proof:

THEOREM 2.5. *Let t be a testing procedure*, Ψ *a family of sentences and* σ *a sentence such that the following holds. For some X of probability* 1 *we have*: (i) *t separates any two* $w_1, w_2 \in X$ *which are separated by* $\{\sigma\}$; (ii) *t separates any two* $w_1, w_2 \in X \cap \operatorname{Mod}(\sigma)$ *which are separated by* Ψ. *Then* (I) *and* (II) *of Theorem* 2.4 *hold for a.e.* $w \in \operatorname{Mod}(\sigma)$.

(The set $\operatorname{Mod}(\sigma)$ can be replaced by an arbitrary $X \in \mathscr{B}$.)

Testing procedures can be investigated further in order to throw light on existing statistical practices, e.g., randomization in sample choosing. The idea is to arrive at some general results of the kind given here, which will show how the properties of our prior probability determine the choice of the preferable testing procedure. Other factors such as the speed of convergence (in (I) of Theorem 2.4) or the cost of testing can be introduced in the framework as well.

§3. **Definable probabilities.** There is a standard technique for defining real numbers and real-valued functions in a first-order arithmetical language; reals are identified with infinite binary sequences and one considers definitions that determine for each n the nth digit of the sequence. The following variant comes to the same thing, but is more useful for our purposes. (In clause (iii) of the definition we treat formulas via their Gödel numbers.)

DEFINITION 3.1. (i) Let f be a real-valued function such that $\operatorname{Domain}(f) \subset N^k$. A rational-valued function g is said to be an *approximating function for f* (or simply, an *approximation for f*) if;

$\operatorname{Domain}(g) = \operatorname{Domain}(f) \times N$;

$|f(\mathbf{x}) - g(\mathbf{x}, y)| \leq 1/y$ for all $\mathbf{x} \in \operatorname{Domain}(f)$, $y \in N$.

(It would make no difference had we replaced $1/y$ by $1/2y$ or 2^{-y}, or any positive recursive function which tends to 0 as $y \to \infty$.)

(ii) A real-valued f is definable in L_0 if it has an approximating function definable by a formula of L_0. It is $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, $\Delta_n(L_0)$, *recursive*) if it has a $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, $\Delta_n(L_0)$, recursive) approximating function. (Recall that a function g is definable by the formula λ if for all $\mathbf{n}, m$ we have $g(\mathbf{n}) = m \Leftrightarrow\ \models \lambda(\mathbf{n}, m)$, and that g is $\Sigma_n(L_0)$ if it is definable by a $\Sigma_n(\lambda_0)$ formula; similarly for $\Pi_n(L_0)$ and $\Delta_n(L_0)$.)

(iii) A probability Pr over L is *definable* if its restriction to elementary sentences is definable in L_0. If this restriction is $\Sigma_n(L_0)$ (recursive, etc.) we say that Pr is *$\Sigma_n(L_0)$-definable, or, for short, $\Sigma_n(L_0)$ (recursive, etc.).*

Recall that a relation R (or a function f) is $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, etc.) in a relation S if it is $\Sigma_n(L_0')$ ($\Pi_n(L_0')$ etc.), where L_0' is obtained from L_0 by adding a name for S. Thus, we can speak also of probabilities that are $\Sigma_n(L_0)$ ($\Pi_n(L_0)$ etc.) in some relation.

Note that given any probability Pr, we can extend L_0 to L_0' by adding to L_0 a predicate that denotes an approximating function for Pr over the elementary sentence. L_0' and L_0 have the same empirical part, hence the same classes of elementary sentences. Consequently Pr has a unique extension to L' (L' = the extended L) and this probability is $\Sigma_0(L_0')$-definable.

Unless the context indicates otherwise, we shall use 'Σ_n', 'Π_n' and 'Δ_n' for $\Sigma_n(L_0)$, $\Pi_n(L_0)$ and $\Delta_n(L_0)$.

As is well known, for $n \geq 1$, a Σ_n-function whose domain is a Π_n-set is also a Π_n-function, hence it is a Δ_n-function. (For if $f(\mathbf{x}) = y \Leftrightarrow \exists z(R(\mathbf{x}, y, z))$, where R is Π_{n-1}, then also $f(\mathbf{x}) = y \Leftrightarrow \mathbf{x} \in \text{Domain}(f) \wedge \forall y', z(R(\mathbf{x}, y'z) \rightarrow y = y')$.) Since the class of elementary sentences is recursive it follows that Σ_n-probabilities are the same as Δ_n-probabilities. We shall therefore use the Δ_n scale in order to classify probabilities.

From now on we put $\ulcorner\varphi\urcorner =_{\text{Df}}$ *Gödel number of* φ.

We shall assume throughout well-known, or easily verifiable, assertions such as that, for $n \geq 1$, the sum and product of Δ_n-functions are Δ_n and that a function which is partially recursive in a Δ_n-function is also Δ_n. The following lemma will be used in order to pass from probabilities to conditional probabilities.

LEMMA 3.1. (i) *Let g be an approximating function for f. Then there is a rational-valued function $\tilde{g}$ which is partially recursive in g satisfying*:

$$\text{Domain}(\tilde{g}) = \{(\mathbf{x}, y): \mathbf{x} \in \text{Domain}(f) \ \& \ f(\mathbf{x}) \neq 0\}$$

and for every $(\mathbf{x}, y) \in \text{Domain}(\tilde{g})$, $|f(\mathbf{x}) - \tilde{g}(\mathbf{x}, y)| \leq f(\mathbf{x})/y$ (*i.e. the relative error is* $\leq 1/y$).

(ii) *Let Φ be a class of sentences closed under conjunctions. Given any approximating function for the restriction of* Pr *to Φ, there is a function, partially recursive in it, which approximates the conditional probability* $\Pr(\psi|\varphi)$ *where* (ψ, φ) *ranges over* $\{(\psi, \varphi): \psi, \varphi \in \Phi \ \& \ \Pr(\varphi) > 0\}$.

PROOF. We shall prove (i) for the case of nonnegative f, which is the case we need. For arbitrary f the proof is essentially the same. Given $\mathbf{x}$ and y let $z_0(\mathbf{x}, y)$ = smallest z such that $g(\mathbf{x}, (y+2)\cdot z) \geq 1/z$. Such a z exists iff $\mathbf{x} \in \text{Domain}(f)$ and $f(\mathbf{x}) > 0$; otherwise z_0 is undefined. If $z_0(\mathbf{x}, y)$ is defined then

$$f(\mathbf{x}) \geq g(\mathbf{x}, (y+2)\cdot z_0) - 1/(y+2)\cdot z_0 \geq 1/z_0 - 1/2z_0.$$

Put $\tilde{g}(\mathbf{x}, y) =_{\text{Df}} g(\mathbf{x}, 2y\cdot z_0(\mathbf{x}, y))$. Then $\tilde{g}(\mathbf{x}, y)$ approximates $f(\mathbf{x})$ with error $\leq 1/2yz_0$, hence with relative error $\leq 1/y$. Evidently, z_0, hence $\tilde{g}$, is partial recursive

(ii) Use (i) and get a function $\tilde{g}$ for the case where g approximates Pr's restriction to Φ. For $\psi, \varphi \in \Phi$ define

$$h(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner, y) =_{\mathrm{Df}} g(\ulcorner\psi \wedge \varphi\urcorner, 4yy')/\tilde{g}(\ulcorner\varphi\urcorner, 4y),$$

where y' = smallest z such that $\tilde{g}(\ulcorner\varphi\urcorner, 4y) \geq 1/z$. Then

$$\begin{aligned} h(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner, y) &= (\Pr(\psi \wedge \varphi) + \delta_1)/\Pr(\varphi)(1 + \delta_2) \\ &= \Pr(\psi|\varphi) + \Pr(\psi|\varphi)[(1 + \delta_2)^{-1} - 1] + \delta_1/(1 + \delta_2)\Pr(\varphi), \end{aligned}$$

where

$$|\delta_1| \leq 1/4yy', \quad \Pr(\varphi) \geq (1 - 1/4y)1/y', \quad |\delta_2| \leq 1/4y.$$

An easy calculation shows that the error is $\leq (1/4y)(4/3 + 16/9) < 1/y$. Q.E.D.

LEMMA 3.2. *Let f be a partial real-valued function over N^k whose domain is a Δ_n-set, $n \geq 1$. The following are equivalent*:

(i) *f is Δ_n.*

(ii) *The relations $f(\mathbf{x}) \leq a$ and $f(\mathbf{x}) \geq a$ as relations over $N^k \times Q$ (Q = set of rationals) are Π_n.*

(iii) *The relations $f(\mathbf{x}) < a$ and $f(\mathbf{x}) > a$ are Σ_n.*

PROOF. (i) $\Rightarrow$ (ii). Let g be a Δ_n-approximating function. Then $f(\mathbf{x}) \leq a \Leftrightarrow \forall y(g(\mathbf{x}, y) \leq a + 1/y)$ and the condition $g(\mathbf{x}, y) \leq a + 1/y$ can be rewritten as $x \in \mathrm{Domain}(f) \wedge \forall z(g(\mathbf{x}, y) = z \rightarrow z \leq a + 1/y)$. Since $g(\mathbf{x}, y) = z$ is a Δ_n-condition the second conjunct is Π_n. The first conjunct is Δ_n; the conjunction is therefore Π_n and so is $f(\mathbf{x}) \leq a$. The other inequality is similarly treated.

(ii) $\Rightarrow$ (iii). $f(\mathbf{x}) < a \Leftrightarrow \mathbf{x} \in \mathrm{Domain}(f) \wedge f(\mathbf{x}) \ngeq a$ and $f(\mathbf{x}) > a \Leftrightarrow \mathbf{x} \in \mathrm{Domain}(f) \wedge f(\mathbf{x}) \nleq a$.

(iii) $\Rightarrow$ (i). If $f(\mathbf{x}) > a$ and $f(\mathbf{x}) < a$ are Σ_n so is the relation $a < f(\mathbf{x}) < b$, as a relation in a, $\mathbf{x}$, b. There is a Π_{n-1}-relation R so that

$$a < f(\mathbf{x}) < b \Leftrightarrow \exists t R(a, b, \mathbf{x}, t).$$

Given $\mathbf{x}$, y define $g(\mathbf{x}, y)$ as the first coordinate, a, of the first pair (a, z) (in some fixed recursive enumeration of $Q \times N$) such that $R(a, a + (y + 1)^{-1}, \mathbf{x}, z)$. Q.E.D.

(Note that the proof implies that if a real-valued function whose domain is Δ_n ($n \geq 1$) has a Δ_n-approximating function, then it also has Δ_n-approximating functions f_1, f_2 satisfying $f_1(\mathbf{x}, y) \leq f(\mathbf{x}) \leq f_2(x, y)$, for all $\mathbf{x} \in \mathrm{Domain}(f)$ and all y.)

The following theorem is often used in the rest of this work.

THEOREM 3.3. (i) *If the restriction of* Pr *to $\Sigma_m(L)$ is Δ_n ($n \geq 1$), then so is its restriction to all sentential combinations of the sentences in $\Sigma_m(L) \cup \Pi_m(L)$. (The same is true if we replace $\Sigma_m(L)$ by $\Pi_m(L)$.)*

(ii) *If the restriction of* Pr *to $\Sigma_m(L)$ is Δ_n ($n \geq 1$), then its restriction to Σ_{m+k} is Δ_{n+k}.*

(iii) *If* Pr *is Δ_n-definable then its restriction to the sentential combinations of $\Sigma_m(L) \cup \Pi_m(L)$-sentences is Δ_{m+n}.*

These assertions are by and large what one would expect. It is easy to see that the definability of Pr's restriction to Σ_m implies the definability of the restriction to Σ_{m+1}. But the straightforward construction yields estimates higher (hence get our estimates we have to use inductively the two characterizations given in Lemma 3.2.

PROOF. (i) We have

$$\Pr(\neg\varphi) > a \Leftrightarrow \Pr(\varphi) < 1 - a, \qquad \Pr(\neg\varphi) < a \Leftrightarrow \Pr(\varphi) > 1 - a.$$

Since for every $\varphi \in \Pi_m(L)$ ($\varphi \in \Sigma_m(L)$), $\neg\varphi$ is equivalent to some effectively constructed $\varphi' \in \Sigma_m(L)$ ($\varphi' \in \Pi_m(L)$), Lemma 3.2 implies that if Pr is Δ_n over $\Sigma_m(L)$ it is also Δ_n over $\Pi_m(L)$ and vice versa. Consequently, if it is Δ_n over $\Sigma_m(L)$ (or over $\Pi_m(L)$) it is also Δ_n over $\Sigma_m(L) \cup \Pi_m(L)$. For $m = 0$, $\Sigma_m(L) = \Pi_m(L)$ and it is already closed under sentential combinations. In general, every sentential combination, say σ, of sentences from $\Sigma_m(L) \cup \Pi_m(L)$ can be effectively rewritten in the equivalent form $\bigvee_{i<k}\sigma_i$ where the σ_i are mutually exclusive ($i \neq j \Rightarrow \models(\sigma_i \to \neg\sigma_j)$) and $\sigma_i = \sigma_i' \wedge \sigma_i''$, $\sigma_i' \in \Sigma_m(L)$, $\sigma_i'' \in \Pi_m(L)$. In order to get an approximate value for $\Pr(\sigma)$ with error $< \varepsilon$ it suffices to get approximate values with error $< \varepsilon/k$ for $\Pr(\sigma_i)$, $i = 0, \ldots, k - 1$, and to sum them up. Hence we have to show that Pr's restriction to all sentences $\sigma' \wedge \sigma''$, $\sigma' \in \Sigma_m(L)$, $\sigma'' \in \Pi_m(L)$ is Δ_n. Let $\varphi = \sigma' \wedge \sigma''$. Say $\sigma' = \exists\mathbf{u}\varphi'$, $\sigma'' = \forall\mathbf{v}\varphi''$, $\varphi' \in \Pi_{m-1}(L)$, $\varphi'' \in \Sigma_{m-1}(L)$. By '$\forall\mathbf{u} < l$' we mean $\forall u_0 < l\ \forall u_1 < l \cdots \forall u_{j-1} < l$, where $\mathbf{u} = u_0, \ldots, u_{j-1}$, similarly for $\exists\mathbf{v} < l$. We get:

$$\Pr(\varphi) \leq a \Leftrightarrow \forall l\,[\Pr((\exists\mathbf{u} < l\,\varphi') \wedge \sigma'') \leq a],$$

$$\Pr(\varphi) \geq a \Leftrightarrow \forall l\,[\Pr(\sigma' \wedge \forall\mathbf{v} < l\,\varphi'') \geq a].$$

Now $(\exists\mathbf{u} < l\ \varphi') \wedge \sigma''$ can be rewritten as a $\Pi_m(L)$ sentence which depends effectively on σ', σ'' and l. Similarly, $\sigma' \wedge \forall\mathbf{v} < l\,\varphi''$ can be effectively converted into a $\Sigma_m(L)$ sentence. Hence the claim follows from Lemma 3.2.

(ii) It suffices to prove (ii) for $k = 1$. The rest will follow by induction. Let $\sigma = \exists\mathbf{u}\varphi(\mathbf{u})$, $\varphi \in \Pi_m(L)$; then

$$\Pr(\sigma) > a \Leftrightarrow \exists l\,[\Pr(\exists\mathbf{u} < l\,\varphi(\mathbf{u})) > a],$$

$$\Pr(\sigma) < a \Leftrightarrow \exists j \forall l\,[\Pr(\exists\mathbf{u} < l\,\varphi(\mathbf{u})) \leq a - 1/j].$$

The condition within the first square brackets is, by Lemma 3.2, Σ_m (because $\exists\mathbf{u} < l\,\varphi$ is effectively convertible into $\Pi_m(L)$ form). The condition within the second square brackets is, again by Lemma 3.2, $\Pi_m(L)$. Hence both right-hand sides are Σ_{m+1} and the claim follows by 3.2.

(iii) This follows from (ii) and (i) once we observe that if Pr is Δ_n over the elementary sentences it is also Δ_n over $\Sigma_0(L)$ sentences. Given a $\Sigma_0(L)$ sentence, say φ, eliminate all bounded quantifiers—replacing each by a suitable disjunction or conjunction. Next, if f is a function symbol of L_0, the term $f(n_0, \ldots, n_{j-1})$ can be replaced by the numeral $\underset{\sim}{n}$ where n is the value of the corresponding function for $n_0, \ldots, n_{j-1}$. Iterate this until all mathematical function symbols are eliminated. This transformation is describable by means of a $\Delta_1(L_0)$-formula. Finally, if R is a predicate of L_0, then $R(n_0, \ldots, n_{j-1})$ is either true or false and, accordingly, replaceable by $\sigma \vee \neg\sigma$, or by $\sigma \wedge \neg\sigma$, where σ is some fixed atomic empirical sentence. Eliminate in this way all mathematical predicates. This again is a $\Delta_1(L_0)$-transformation. If L has no empirical function symbols the sentence φ' thus obtained is elementary. The passage from φ to φ' is a $\Delta_1(L_0)$-function and our claim follows from the standard fact that if $g(x, y)$ is a $\Delta_n(L_0)$-function, $n \geq 1$, and $h(x)$ is a $\Delta_1(L_0)$-function, then $g(h(x), y)$ is $\Delta_n(L_0)$. If L has empirical function symbols then by the observation made in §1 (proof of Basic Fact 1.1), φ' can be effectively transformed into two equivalent forms $\exists \mathbf{v}\varphi_1'(\mathbf{v})$ and $\forall \mathbf{v}\varphi_2'(\mathbf{v})$ where φ_1', φ_2' are elementary formulas. Then we have

$$\Pr(\exists \mathbf{v} < n\ \varphi_1') \leq \Pr(\varphi') \leq \Pr(\forall \mathbf{v} < n\ \varphi_2')$$

and the two extreme terms converge, as $n \to \infty$, to $\Pr(\varphi')$. Since $\exists \mathbf{v} < n\ \varphi_1'$ and $\forall \mathbf{v} < n\ \varphi_2'$ are effectively equivalent to elementary sentences, their probabilities can be calculated to any desired approximation using a Δ_n-function. Given $\varepsilon > 0$, calculate the two extreme terms for $n = 1, 2, \ldots,$ with error $< \varepsilon/3$. Evertually we shall find n such that the absolute difference between the two is $< \varepsilon/3$. For this n each of the two values approximates $\Pr(\varphi')$ with error $< \varepsilon$. This recipe yields a Δ_n-approximating function. Q.E.D.

REMARKS. (I) The restriction of any probability Pr to $\Sigma_m(L)$, where $m \geq 1$, cannot be Δ_k if $k < m + 1$. For if φ is a sentence in L_0 then φ is true $\Leftrightarrow \Pr(\varphi) > 1/2$. If the relation $\Pr(\varphi) > a$, where φ ranges over $\Sigma_m(L)$ sentences (and a over the rationals), is $\Delta_k(L_0)$, then we have a $\Delta_k(L_0)$-truth definition for $\Sigma_m(L_0)$. This, as is well known, implies $k \geq m + 1$. Hence, for $n = 1$ Theorem 3.3(ii) cannot be improved in any case.

(II) For every $n \geq 1$ there exists a Δ_n-definable probability whose restriction to $\Sigma_m(L)$, $m = 0, 1, 2, \ldots,$ is not Δ_{m+n-1}. This shows that the levels obtained in Theorem 3.3 cannot, in general, be improved. However, we also have:

(III) For every $n > 1$, there exists a probability which is Δ_n, but not Δ_{n-1}, whose restriction to $\Sigma_m(L)$ is Δ_n for all $m < n$, Δ_{m+1} for all $m \geq n$. Hence, for $n > 1$, there are nontrivial cases where the levels obtained in 3.3 can be improved. In general, one can consider the function $d(m)$ = smallest k such that Pr is Δ_k over $\Sigma_m(L)$. Then, by 3.3 and Remark (I), we have two inequalities:

$$d(m_1 + m_2) \leq d(m_1) + m_2, \qquad d(m) > m.$$

By Remark (II) there are, for each n, examples such that $d(m) = m + n$ and, by the present remark, we can also have $d(m) = n$, for $m < n$, $d(m) = m + 1$ for $m \geq n$. What other d's are possible?

An outline of the constructions that prove (II) and (III) is given in an appendix.

The last remark shows that, whereas Δ_n-definability implies that on the mth level the probability is Δ_{m+n}, the converse implication may not hold.

The following essentially well-known fact will be used in the sequel.

ENUMERATION THEOREM. *For each $m \geq 1$ and each k there exists a formula $\gamma(v_0, \ldots, v_k)$ such that for every $\Sigma_m(L)$ formula with k free variables $\varphi(v_1, \ldots, v_k)$ we have*

$$\models \forall v_1 \cdots v_k[\gamma(\ulcorner\varphi\urcorner, v_1 \cdots v_k) \leftrightarrow \varphi(v_1 \cdots v_k)].$$

If $m > 0$ then γ is $\Sigma_m(L)$ and for $m = 0$, γ can be chosen as a $\Delta_1(L)$-formula (*i.e., we have $\gamma' \in \Sigma_1(L)$, $\in \Pi_1(L\ \gamma'')$ such that $\models \gamma \leftrightarrow \gamma' \leftrightarrow \gamma''$*). *The same holds with '$\Pi_m(L)$' replacing '$\Sigma_m(L)$'.*

The corresponding fact for L_0 is well known. A variant of it is the existence of a universal Σ_m-relation of $k + 1$ arguments which enumerates all Σ_m-relations of k arguments. (Another variant for L_0 is that we have for $m \geq 1$ a Σ_m-truth predicate for Σ_m-formulas and a Δ_1-truth predicate for Σ_0-formulas.) The proof is by standard technique. We only indicate how to handle the empirical part. One starts by constructing a formula $\delta(v_0, \ldots, v_m)$ asserting that v_0 is the Gödel number of an atomic empirical formula which is satisfied by $v_1, \ldots, v_m$. (Here m is the maximal number of free variables in atomic empirical formulas; it is finite because L has finitely many empirical symbols.) δ is a disjunction of all the following formulas:

$$(v_0 = \ulcorner P(v_1, \ldots, v_j)\urcorner) \wedge P(v_1, \ldots v_j),$$
$$(v_0 = \ulcorner f(v_1, \ldots, v_j) = v_{j+1}\urcorner) \wedge f(v_1, \ldots, v_j) = v_{j+1},$$

where P ranges over empirical predicates, f over empirical function symbols, and where $v_0 = \ulcorner \ldots \urcorner$ asserts that v_0 is the Gödel number of .... Then proceed as in the proof for L_0.

Continuity and compatibility for definable probabilities. For definable probabilities continuity and compatibility can be expressed by natural conditions that refer to sentences of L only:

THEOREM 3.4. *If* $\Pr_1$, $\Pr_2$ *are definable then* $\Pr_2 \prec \Pr_1$ *iff* $\Pr_1(\varphi) = 0 \Rightarrow \Pr_2(\varphi) = 0$ *for all sentences* φ. *If the probabilities are* Δ_n*-definable*, $n \geq 1$, *then it suffices to require that the implication hold for all* $\varphi \in \Pi_{n+1}(L)$.

PROOF. Suppose $\Pr_2 \not\prec \Pr_1$. Then there exists a $\delta > 0$ such that for every $\varepsilon > 0$ there is an elementary sentence ψ with $\Pr_1(\psi) \leq \varepsilon$ and $\Pr_2(\psi) \geq \delta$. In particular, we have a sequence φ_i, $i = 1, 2, \ldots$, of elementary sentences such that $\Pr_1(\varphi_i) \leq 1/2^i$ and $\Pr_2(\varphi) \geq \delta$. Suppose this sequence is defined in L_0 by a formula $\sigma(x, y)$, i.e. $\models \sigma(n, i) \Leftrightarrow n = \ulcorner \varphi_i \urcorner$. By the enumeration theorem there is a $\Delta_1(L)$ formula γ which satisfies $\models \gamma(\ulcorner \varphi \urcorner) \leftrightarrow \varphi$ for all elementary sentences φ.

Using σ and γ we construct a sentence φ which is equivalent to the infinitary sentence $\bigwedge_{n=0}^{\infty}(\bigvee_{i=n}^{\infty} \varphi_i)$, that is to say, $\mathrm{Mod}(\varphi) = \bigcap_{n=0}^{\infty} \bigcup_{i=n}^{\infty} \mathrm{Mod}(\varphi_i)$:

$$\varphi = \forall y\, \exists z > y\, \exists x\, (\sigma(x, z) \wedge \gamma(x)).$$

Clearly $\Pr_2(\varphi) \geq \delta$, but $\Pr_1(\varphi) \leq \sum_{i=n}^{\infty} 1/2^i$ for all n, implying $\Pr_1(\varphi) = 0$. If σ is a Σ_n-formula then φ is a $\Pi_{n+1}(L)$-sentence. It remains to build the formula σ. Let $f_i^*(x, y)$ be a Δ_n-approximating function for $\Pr_i$'s restriction to elementary sentences, $i = 1, 2$. Let $k \in N$ be such that $4/k \leq \delta$. Then

$$\Pr_2(\psi) \geq 4/k \Rightarrow f_2^*(\ulcorner \psi \urcorner, k) \geq 3/k \Rightarrow \Pr_2(\psi) \geq 2/k.$$

Similarly,

$$\Pr_1(\psi) \leq 1/2^{i+1} \Rightarrow f_1^*(\ulcorner \psi \urcorner, 2^{i+2}) \leq 3/2^{i+2} \Rightarrow \Pr_1(\psi) \leq 1/2^i.$$

Thus we have a first (i.e., with the smallest Gödel number) elementary sentence φ_i such that $f_2^*(\ulcorner \varphi_i \urcorner, k) \geq 3/k$ and $f_1^*(\varphi_i, 2^{i+2}) \leq 3/2^{i+2}$. This φ_i satisfies $\Pr_2(\varphi_i) \geq 2/k$ and $\Pr_1(\varphi_i) \leq 1/2^i$. The sequence φ_i is defined by the following $\sigma(x, y)$:

$$(f_2^*(x, k) \geq 3/k) \wedge (f_1^*(x, 2^{y+2}) \leq 3/2^{y+2}) \wedge$$
$$\forall x' < x\, \neg[(f_2^*(x', k) \geq 3/k) \wedge (f_1^*(x, 2^{y+2}) \leq 3/2^{y+2})]$$

where x and x' range on Gödel numbers of elementary sentences. Now if g is a rational-valued Δ_n-function, $n \geq 1$, and Domain(g) is Δ_n, then $g(\mathbf{x}) \geq a$ and $g(\mathbf{x}) \leq a$ are Δ_n. (For $g(\mathbf{x}) \geq a$ is equivalent to $\exists y(g(\mathbf{x}) = y \wedge y \geq a)$ as well as to $\mathbf{x} \in D(g) \wedge \forall y(g(\mathbf{x}) = y \rightarrow y \geq a)$. Similarly for $g(\mathbf{x}) \leq a$.) Hence σ is Σ_n. Q.E.D.

The theorem presupposes a finite number of empirical symbols; with infinitely many empirical symbols one can construct Δ_1-probabilities Pr_i, $i = 1, 2$, such that for any finite sublanguage, L', the restriction of Pr_2 to the sentences of L' is absolutely continuous with respect to the restriction of Pr_1, but $\mathrm{Pr}_2 \not\prec \mathrm{Pr}_1$. The characterization of absolute continuity is not valid if the probabilities are not definable as one can see from easily constructed counterexamples. But we do not know the answer to the following:

Problem. Is Π_{n+1} in the second claim the best possible? In particular, is Π_2 the lowest level for Δ_1-definable probabilities? A positive answer implies the existence of two recursive probabilities Pr_i, $i = 1, 2$, such that $\mathrm{Pr}_2 \not\prec \mathrm{Pr}_1$ but $\mathrm{Pr}_1(\varphi) = 0 \Rightarrow \mathrm{Pr}_2(\varphi) = 0$ for all $\varphi \in \Pi_1(L)$.

COROLLARY 3.5. *Two Δ_n-definable probabilities are compatible iff they have the value* 0 *for the same Π_{n+1} sentences.*

COROLLARY 3.6. *If* Pr *is Δ_n-definable, there is a $\Sigma_{2n+2}(L)$ sentence such that, for every Δ_n-definable* Pr′, Pr′ $\prec$ Pr *iff* $\mathrm{Pr}'(\psi) = 0$.

PROOF. A sentence ψ which is equivalent to the infinitary disjunction of all $\Pi_{n+1}(L)$-sentences for which Pr has value 0 will obviously do the job. By Theorem 3.3, $\{\varphi \in \Pi_{n+1}(L): \mathrm{Pr}(\varphi) = 0\}$ is a Π_{2n+1}-set. Let $\rho(x)$ be a $\Pi_{2n+1}(L_0)$-formula defining it. By the enumeration theorem we have a $\Pi_{n+1}(L)$-formula γ satisfying $\models \gamma(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$, for all $\Pi_{n+1}(L)$ φ's. Put $\psi = \exists x[\rho(x) \wedge \gamma(x)]$. Q.E.D.

Again, is Σ_{2n+2} the best possible?

Assigning nonzero values to consistent sentences. We now take up the question of consistent sentences having probability 0. Recall that, in this context, consistency is the *semantic* condition of having a model in Mod_L. As remarked before, in order to be nondogmatic Pr should assign positive values to consistent sentences. If, however, Pr is definable this cannot be achieved for all sentences:

THEOREM 3.7. *There is a consistent $\Pi_2(L)$-sentence, φ, such that* $\mathrm{Pr}(\varphi) = 0$ *for every definable* Pr. *If, moreover, L has an empirical function symbol then there exists such a $\Pi_1(L)$-sentence.*

PROOF. Let P be a monadic empirical predicate. (If L does not have monadic empirical predicates, take instead any atomic empirical formula with one free variable.) Let φ be the sentence asserting that P is a truth predicate for L_0. That is to say, φ asserts that the class defined by P satisfies the inductive conditions of a truth definition. φ is a conjunction of the sentences which formalize the following conditions (where sentences are treated via their Gödel numbers): $\forall v[P(v) \rightarrow v$ is a sentence of $L_0]$, $\forall v[v$ is atomic $\rightarrow (P(v) \leftrightarrow \mathrm{True}_0(v))]$, where True_0 is a Δ_1 truth definition for atomic formulas of L_0, $\forall v_1, v_2[v_1 =$ negation of $v_2 \rightarrow (P(v_1) \leftrightarrow \neg P(v_2))]$. $\forall v_1, v_2, v_3[v_1 =$ conjunction of v_2 and $v_3 \rightarrow (P(v_1) \leftrightarrow P(v_2) \wedge P(v_3))]$ and finally $\forall v_1, v_2[v_1$ is a sentence obtained by one universal quantification of $v_2 \rightarrow (P(v_1) \leftrightarrow \forall u P(\mathrm{Sub}(v_2, u)))]$, where $\mathrm{Sub}(v_2, u)$ is the formula describing the result of substituting the name of the number u for the free variable of v_2. All, except for the last condition, are Π_1. The last is Π_2. Hence φ is Π_2. Now in every $w \in \mathrm{Mod}(\varphi)$, P must be the truth predicate for L_0. Hence, for every sentence ψ of L_0, $\mathrm{Mod}(\varphi) \cap \mathrm{Mod}(P(\ulcorner\psi\urcorner))$ is either $\mathrm{Mod}(\varphi)$ or empty according as $\models\psi$ or $\models\neg\psi$. If $\Pr(\varphi) > 0$, then for each sentence ψ of L_0, $\models\psi \Leftrightarrow \Pr(\varphi \wedge P(\ulcorner\psi\urcorner)) > 0$, which, were Pr definable, would yield a truth definition for L_0 inside L_0. Hence, if Pr is definable, $\Pr(\varphi) = 0$.

To establish the second claim, take an empirical unary function symbol F (or any empirical term with one free variable) and construct φ as before except that '$P(v)$' is to be replaced by '$F(v) = 0$' and in the last condition the clause $P(v_1) \leftrightarrow \forall u P(\mathrm{Sub}(v_2, u))$ is to be replaced by

$$[F(v_1) = 0 \rightarrow \forall u(F(\mathrm{Sub}(v_2, u)) = 0)] \\ \wedge [F(v_1) > 0 \rightarrow F(\mathrm{Sub}(v_2, F(v_1) - 1)) > 0].$$

This means that in the case where $\forall u\psi(u)$ is false $F(\ulcorner\forall u\psi(u)\urcorner) - 1$ is a counterexample to the generalization. Arguing as before, $\mathrm{Mod}(\varphi \wedge (F(\ulcorner\psi\urcorner) = 0))$ is either $\mathrm{Mod}(\varphi)$ or empty according as $\models\psi$ or $\models\neg\psi$. It is easy to check that φ is $\Pi_1(L)$. Q.E.D.

The following is another theorem that shows the extent to which definable probabilities are dogmatic.

THEOREM 3.8. *Let* Pr *be a definable probability with a definable approximating function f, P an empirical monadic predicate, k a positive integer. There exists in L_0 a formula $\lambda(v)$, expressing a property recursive in f and k, such that if $\psi(v)$ is the formula $P(v) \leftrightarrow \lambda(v)$ then* $\Pr(\psi(m) \mid \bigwedge_{i<m}\psi(i)) \leq 1/2 + 1/k$ *whenever* $\Pr(\bigwedge_{i<m}\psi(i)) > 0$. *(For $m = 0$, $\bigwedge_{i<m}\psi(i)$ is, by definition, $0 = 0$.) In particular, if* Pr *is $\Delta_n(L_0)$ then ψ is $\Delta_n(L)$, and if* Pr *is recursive then ψ is recursive in $P(v)$.*

(Instead of $P(v)$ we can use any empirical formula of $\Sigma_0(L)$ with one free variable.)

Note that the claim implies that, for all m, $\Pr(\bigwedge_{i<m}\psi(i)) \leq (1/2 + 1/k)^m$; hence, for $k > 2$, $\Pr(\forall v\psi(v)) = 0$. Obviously, $\forall v\psi(v)$ is consistent (being equivalent to an infinite conjunction $\bigwedge_i P'(i)$ where each $P'(i)$ is either $P(i)$ or $\neg P(i)$). Thus if Pr is Δ_n, we get a consistent $\Pi_n(L)$-sentence of probability 0. The sentence depends on Pr; hence if $n > 1$ this implication, by itself, is weaker than Theorem 3.7. The interest of Theorem 3.8 lies in the special form of our sentence (which serves later to derive Theorem 3.12) as well as in the case where Pr is recursive. Another point of interest is that Pr is dogmatic with respect to ψ in a strong sense: no matter how large m, evidence that $\psi(i)$ holds for all $i < m$ yields (when compatible with Pr) a conditional probability of $\psi(m)$ which is $\leq$ the fixed number $1/2 + 1/k$.

The proof is based on a diagonal construction. (The idea of using a diagonal argument in a probabilistic context is due to Putnam [Pu]; his construction proves a weaker claim.)

PROOF. The basic idea is to define inductively a sequence $P'(i)$, where $P'(i)$ is either $P(i)$ or $\neg P(i)$, choosing $P'(n)$ as $P(n)$ if $\Pr(P(n) \mid \bigwedge_{i<n} P'(i)) \leq \frac{1}{2}$, as $\neg P(n)$ otherwise. However this inductive definition is not recursive in f since it involves the verifying of inequalities such as $\Pr(P(0)) \leq \frac{1}{2}$. Hence we proceed as follows. By Lemma 3.1 there exists an approximating function $\tilde{f}(x, y, z)$ for the

conditional probabilities of elementary sentences which is partially recursive in f. If $\tilde{f}(\ulcorner\varphi_1\urcorner, \ulcorner\varphi_2\urcorner, 2k) \leq 1/2 + 1/2k$ then $\Pr(\varphi_1|\varphi_2) \leq 1/2 + 1/2k + 1/2k = 1/2 + 1/k$. Conversely, if $\tilde{f}(\ulcorner\varphi_1\urcorner, \ulcorner\varphi_2\urcorner, 2k) > 1/2 + 1/2k$ then $\Pr(\varphi_1|\varphi_2) > 1/2 + 1/2k - 1/2k = 1/2$, implying $\Pr(\neg\varphi_1|\varphi_2) < 1/2 + 1/k$. Let C be the class defined by induction:

$$m \in C \Leftrightarrow \tilde{f}\left(\ulcorner P(m)\urcorner, \ulcorner\bigwedge_{i<m} P^C(i)\urcorner, 2k\right) \leq \frac{1}{2} + \frac{1}{2k}$$

where $P^C(i)$ is defined as $P(i)$ if $i \in C$, as $\neg P(i)$ if $i \notin C$. The right-hand side means that f is defined and its value satisfies the inequality. If $\Pr(\bigwedge_{i<m}P^C(i)) = 0$, then $m \notin C$; since this will also hold for all $m' > m$, C will be finite. If, for all m, $\Pr(\bigwedge_{i<m}P^C(i)) > 0$, then $\tilde{f}$ on the right side is always defined, implying that C is recursive in $\tilde{f}$ and k and, consequently, recursive in f and k. If $\Pr(\bigwedge_{i<m}P^C(i)) > 0$ then $\Pr(P^C(m)|\bigwedge_{i<m}P^C(i)) \leq 1/2 + 1/k$. Let $\lambda(v)$ be a formula in L_0 defining the set C. Then $P(n) \leftrightarrow \lambda(n)$ is equivalent to $P^C(n)$. Q.E.D.

We now turn to the construction of nondogmatic probabilities.

THEOREM 3.9. *Let* Ψ *be a set of sentences such that* $\mathrm{Mod}(\psi)$ *is closed for all* $\psi \in \Psi$. *Let* Ξ *be the set of all consistent conjunctions* $\psi \wedge \varphi$, *where* $\psi \in \Psi$ *and* φ *is a conjunction of basic empirical sentences. Then there is a probability* Pr *which is recursive in* Ξ *such that* $\Pr(\psi) > 0$ *for all consistent* $\psi \in \Psi$.

PROOF. Let $\Psi_0 = \{\psi \in \Psi: \psi \text{ is consistent}\}$. We shall construct a rational-valued function, $H(x, y)$, recursive in Ξ such that for $\psi \in \Psi_0$, $H(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$ as a function of φ, where φ ranges over elementary sentences, satisfies the conditions of a finitely-additive probability (Definition 1.1) and such that $H(\ulcorner\psi\urcorner, \ulcorner\bigvee_{i<n}(f(\mathbf{m}) = i)\urcorner) \to 1$ as $n \to \infty$ for every empirical term $f(\mathbf{m})$. By Theorem 1.1 this function extends to a unique probability $\Pr_\psi$ on L. Furthermore, H is constructed so that $\Pr_\psi(\psi) = 1$ for all $\psi \in \Psi_0$. Then we use an enumeration $\psi_0, \psi_1, \ldots, \psi_n, \ldots$ of Ψ_0 (possibly with repetitions) which is recursive in Ξ (it exists for, fixing any atomic empirical φ, we have $\psi \in \Psi_0$ iff $\psi \wedge \varphi \in \Xi$ or $\psi \wedge \neg\varphi \in \Xi$). Define $\Pr =_{\mathrm{Df}} \sum(1/2^{i+1})\cdot\Pr_{\psi_i}$. Then $\psi \in \Psi \Rightarrow \Pr(\psi) > 0$. In order to approximate $\Pr(\varphi)$ up to $1/m$, take k such that $1/2^k \leq 1/m$ and compute $\sum_{i<k}\Pr_{\psi_i}(\varphi)$. Hence Pr is recursive in Ξ. The main point is to get H.

Let $s = (\sigma_0, \sigma_1, \ldots, \sigma_1, \ldots)$ be a sequence of atomic empirical formulas, either finite or infinite. We associate with s a function $H_s(x, y)$, using a procedure recursive in s and Ξ. For every sentence ψ and every conjunction $\bigwedge_{i<n}\pm\sigma_i$, where where $n <$ length of s and $\pm\sigma_i$ is either σ_i or $\neg\sigma_i$, let $H_s(\ulcorner\psi\urcorner, \ulcorner\bigwedge_{i<n}\pm\sigma_i\urcorner)$ be determined by the following induction on n. For $n = 0$ we let $\bigwedge_{i<n} \pm\sigma_i$ be $0 = 0$ and put $H_s(\ulcorner\psi\urcorner, \ulcorner 0=0\urcorner) = 1$. Supposing $H_s(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$ to be defined where $\varphi = \bigwedge_{i<n}\pm\sigma_i$, let $\varphi_0 = \varphi \wedge \sigma_n$, $\varphi_1 = \varphi \wedge \neg\sigma_n$. If of the two sentences $\psi \wedge \varphi_0$, $\psi \wedge \varphi_1$ exactly one, say $\psi \wedge \varphi_i$, is in Ξ, define $H_s(\ulcorner\psi\urcorner, \ulcorner\varphi_i\urcorner) = H_s(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$ and $H_s(\ulcorner\psi\urcorner, \ulcorner\varphi_{1-i}\urcorner) = 0$; if not, put $H_s(\ulcorner\psi\urcorner, \ulcorner\varphi_0\urcorner) = H_s(\ulcorner\psi\urcorner, \ulcorner\varphi_1\urcorner) = \frac{1}{2}\cdot H_s(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$. If φ is any sentential combination of members of s it can be effectively transformed into an equivalent sentence $\bigvee_{j\in J}\varphi_j$ where the φ_j's are mutually exclusive conjunctions of the form $\bigwedge_{i<n}\pm\sigma_i$ and we put $H_s(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner) =_{\mathrm{Df}} \sum_{j\in J}H_s(\ulcorner\psi\urcorner, \ulcorner\varphi_j\urcorner)$. It is easily seen that this definition is legitimate and yields a

finitely additive probability, say $\widetilde{\Pr}_{\psi,s}$, over all sentential combinations of the members of s. Note that if $s_1{}^\frown s_2$ is the concatenation of s_1 with s_2 then $H_{s_1{}^\frown s_2}$ extends H_{s_1}. Next we show that given $\psi \in \Psi_0$ we can enumerate all atomic empirical sentences in a sequence $s(\psi)$ such that if $s = s(\psi)$ then $\widetilde{\Pr}_{\psi,s}(\bigvee_{i<n} f(\mathbf{m}) = i) \to 1$ as $n \to \infty$ for every empirical function symbol f and every $\mathbf{m} \in N^k$. The enumeration will depend on ψ and will be recursive in Ξ; thus, if H is defined over $\{(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$: $\psi \in \Psi_0$, φ-elementary$\}$ by: $H(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner) =_{\mathrm{Df}} H_{s(\psi)}(\ulcorner\psi\urcorner, \ulcorner\varphi\urcorner)$, then H is recursive in Ξ.

In order to facilitate the presentation we use sequences s, in which sentences may be repeated. The definition of H_s applies equally well and it is easy to see that $H_s = H_{s'}$ where s' is the subsequence of all first occurrences of sentences in s. Let $s = (\sigma_0, \ldots, \sigma_{r-1})$, $\varphi = \bigwedge_{i<r} \pm\sigma_i$, $\psi \in \Psi_0$ and suppose that $f(\mathbf{m})$ is some term, where f is an empirical function symbol. Let t_n be the sequence $(f(\mathbf{m}) = 0, \ldots, f(\mathbf{m}) = n-1)$.

Claim. For every $k > 0$ there exists n such that

$$H_{s^\frown t_n}(\ulcorner\psi\urcorner, \ulcorner\varphi \wedge \textstyle\bigwedge_{i<n} f(\mathbf{m}) \neq i\urcorner) < 1/k.$$

PROOF. Let t be the infinite sequence $f(\mathbf{m}) \neq 0, \ldots, f(\mathbf{m}) \neq i, \ldots$. If $\psi \wedge \varphi \wedge \bigwedge_{i<j} f(\mathbf{m}) \neq i$ is consistent let $w \in \mathrm{Mod}_L$ be a model where it is true. For some $j' > j$, $w \models f(\mathbf{m}) = j'$; then $\psi \wedge \varphi \wedge (\bigwedge_{i<j'} f(\mathbf{m}) \neq i) \wedge f(\mathbf{m}) = j'$ is consistent. By our construction this implies that $H_{s^\frown t}(\ulcorner\psi\urcorner, \ulcorner\varphi \wedge \bigwedge_{i<j'+1} f(\mathbf{m}) \neq i\urcorner)$ is either $\frac{1}{2} \cdot H_{s^\frown t}(\ulcorner\psi\urcorner, \ulcorner\varphi \wedge \bigwedge_{i<j'} f(\mathbf{m}) \neq i\urcorner)$ or 0. Hence in the sequence $H_{s^\frown t}(\ulcorner\psi\urcorner, \ulcorner\varphi \wedge \bigwedge_{i<n} f(\mathbf{m}) \neq i\urcorner)$, $n = 0, 1, 2, \ldots$, either from a certain point on we get 0 or there are infinitely many places where the value is halved. Hence for a large enough n the value is $< 1/k$.

Now take a fixed effective enumeration of atomic empirical sentences and a fixed effective enumeration of all terms of the form $f(\mathbf{m})$ where each term occurs infinitely many times. For $\psi \in \Psi_0$ construct $s(\psi)$ using a step-by-step prolongation. At the $2j$th step we add the jth atomic empirical sentence. Suppose that in $2j$ steps we got the sequence $(\sigma_0, \ldots, \sigma_l)$. Let $f(\mathbf{m})$ be the jth term in our enumeration; then at the $2j+1$ step extend the sequence to $\sigma_0, \ldots, \sigma_l, f(\mathbf{m}) = 0, \ldots, f(m) = n$ where n is large enough so that $H_s(\ulcorner\psi\urcorner, \ulcorner\bigwedge_{i<l} \pm\sigma_i \wedge \bigwedge_{i<n} f(\mathbf{m}) \neq i\urcorner) < 1/j$ for each of the 2^l conjunctions $\bigwedge_{j<l} \pm\sigma_j$, where s is the extended sequence. Obviously the resulting H is recursive in Ξ. It is easy to see that

$$\widetilde{\Pr}_{\psi, s(\psi)}(\textstyle\bigwedge_{i<n} f(\mathbf{m}) \neq i) \to 0 \quad \text{as } n \to \infty,$$

for every term $f(\mathbf{m})$; hence $\widetilde{\Pr}_{\psi, s(\psi)}(\bigvee_{i<n} f(\mathbf{m}) = i) \to 1$.

It remains to show that $\Pr_\psi(\psi) = 1$. If $\sigma_0, \ldots, \sigma_n, \ldots$ is the sequence $s(\psi)$, then by the definition of H, if $\psi \wedge \bigwedge_{i<k} \pm\sigma_i$ is inconsistent, then $H(\ulcorner\psi\urcorner, \ulcorner\bigwedge_{i<k} \pm\sigma_i\urcorner) = 0$. This implies that $\Pr_\psi(\varphi) = 0$ whenever φ is elementary and inconsistent with ψ. Since $\mathrm{Mod}(\psi)$ is closed, $\mathrm{Mod}(\neg\psi)$ is the union of all $\mathrm{Mod}(\varphi)$'s where φ is elementary and inconsistent with ψ. Hence $\Pr_\psi(\neg\psi) = 0$. Q.E.D

The same technique proves the following generalization of Theorem 3.9.

THEOREM 3.9′. *If*

(i) $\mathrm{Mod}(\psi)$ *is closed for all* $\psi \in \Psi$,

(ii) Φ *is a recursive set of consistent* $\Sigma_1(L)$ *sentences*,

(iii) Ξ *consists of all* $\psi \wedge \varphi$, *which are consistent with* Φ *where* $\psi \in \Psi$ *and* φ *is a conjunction of basic empirical sentences,*

then there is a Pr *recursively definable in* Ξ *such that* $\Pr(\psi) > 0$ *for all* $\psi \in \Psi$ *consistent with* Φ and $\Pr(\varphi) = 1$ *for all* $\varphi \in \Phi$.

Letting Ψ be the set of all $\Pi_1(L)$ sentences we get

COROLLARY 3.10. *If L has no empirical function symbols then there exists a* Δ_2*-definable probability that assigns a positive value to each consistent* Π_1*-sentence.*

PROOF. Let Ξ be the set of all consistent $\psi \wedge \varphi$, where ψ is $\Pi_1(L)$ and φ is a conjunction of basic empirical sentences. If ψ is $\forall \mathbf{v}\alpha(\mathbf{v})$ (where α is $\Sigma_0(L)$) then $\psi \wedge \varphi$ is consistent iff every finite conjunction $\varphi \wedge \alpha(\mathbf{n}_0) \wedge \cdots \wedge \alpha(\mathbf{n}_{k-1})$ is consistent. For, having no empirical function symbols, Mod_L is a compact space and if each intersection $\mathrm{Mod}(\varphi) \cap \bigcap_{i<k} \mathrm{Mod}(\alpha(\mathbf{n}_i))$ is nonempty then $\mathrm{Mod}(\varphi) \cap \bigcap_{\mathbf{n}\in N^k}\mathrm{Mod}(\alpha(\mathbf{n})) \neq \emptyset$. The consistency of $\Sigma_0(L)$-sentences is expressible in L_0 by a Δ_1-formula (where the sentences are represented by their Gödel numbers). It follows that Ξ is a $\Pi_1(L_0)$-set. Since $\mathrm{Mod}(\psi)$ is closed for all $\psi \in \Pi_1(L)$, we get the desired probability as a function recursive in Ξ, which function is Δ_2. Q.E.D

By Theorem 3.7, the condition that L has no empirical function symbols cannot be omitted. Neither can we, by Theorem 3.8, obtain the probability on a level lower than Δ_2. One can, however, get, for more restricted sets Ψ of $\Pi_1(L)$-sentences, Δ_1 (or recursively definable) probabilities that assign positive values to the consistent members of Ψ, for example:

COROLLARY 3.11. *Let* $G(x, y)$ *be a total recursive function of two arguments. Let* $\tau(u, v)$ *be a term in* L_0 *that defines G. If F is a one-place empirical function symbol then there is a recursively definable probability* Pr *such that* $\Pr(\forall v(F(v) = \tau(m, v))) > 0$ *for all* $m \in N$. *If P is a monadic empirical predicate, there is a recursively definable probability* Pr *such that* $\Pr(\forall v[P(v) \leftrightarrow \tau(m, v) = 0]) > 0$ *for all* $m \in N$.

PROOF. We can effectively find, given any conjunction φ of basic empirical sentences and any $m \in N$, whether $\forall v(F(v) = \tau(m, v)) \wedge \varphi$ is consistent. Namely, for each n, if φ contains a conjunct of the form $F(n) = k$, or of the form $F(n) \neq k$, calculate $G(m, n)$ and check whether $G(m, n) = k$ or $G(m, n) \neq k$. Then apply Theorem 3.9. The second claim is similarly proved. Q.E.D.

By Theorem 3.8 we cannot have a recursively definable Pr such that

$$\Pr(\forall v(F(v) = \sigma(v))) > 0$$

for all terms σ that define total recursive functions.

THEOREM 3.12. *For every* Δ_n*-definable probability* Pr (*where* $n \geq 1$) *there exist a* Δ_n*-definable* $\widetilde{\Pr}$, *and a* Δ_n*-sentence* σ *such that* $\Pr(\sigma) = 0$ *but* $\widetilde{\Pr}(\sigma) = 1$.

PROOF. Let $P(v)$ be an atomic empirical formula either of the form $R(v, n_1, \ldots, n_{k-1})$, or of the form $f(v, n_1, \ldots, n_{k-1}) = n_k$. Let σ be $\forall v(P(v) \leftrightarrow \lambda(v))$, where $\lambda(v)$ is the $\Delta_n(L_0)$ formula of Theorem 3.8. Then $\Pr(\sigma) = 0$. Using Theorem 3.9, we construct a Δ_n-definable probability $\widetilde{\Pr}$ such that $\widetilde{\Pr}(\sigma) > 0$ (in fact, we will have $\widetilde{\Pr}(\sigma) = 1$). Let Ψ of Theorem 3.9 be $\{\sigma\}$ and let Ξ be the set of all consistent $\sigma \wedge \varphi$, where φ is a conjunction of basic empirical sentences. Since $\widetilde{\Pr}$ is recursive in Ξ it suffices to show that Ξ is $\Delta_n(L_0)$. Now σ asserts the equivalence of $P(v)$

with a formula $\lambda(v)$ which does not contain any empirical symbols; hence a conjunction φ of basic empirical sentences is consistent with σ iff φ' is consistent, where φ' is obtained by replacing every $P(n)$ occurring in φ by $\lambda(n)$. We have $\varphi' = \varphi_1' \wedge \varphi_2'$ where φ_1' is the conjunction of all conjuncts in φ' of the form $\lambda(n)$ or $\neg\lambda(n)$ and φ_2' is the rest of the conjunction; it is easy to see that φ' is consistent iff φ_1' is consistent and φ_2' is consistent. φ_2' is consistent iff it contains no conjunction of a sentence and its negation, φ_1' is consistent iff it is true. Hence all reduces to verifying the truth of conjunctions of the form $\bigwedge_{i<k}\lambda(n_i) \wedge \bigwedge_{j<l}\neg\lambda(m_j)$. Using the fact that $\lambda(v)$ is equivalent to a $\Sigma_n(L_0)$-formula as well as to a $\Pi_n(L_0)$-formula and that truth for $\Sigma_n(L_0)$ ($\Pi_n(L_0)$) sentences is definable by a $\Sigma_n(L_0)$ ($\Pi_n(L_0)$) formula, it follows that the truth of φ_2' can be stated using a $\Delta_n(L_0)$-formula. All in all, Ξ is $\Delta_n(L_0)$. Q.E.D.

COROLLARY 3.13. *For every Δ_n-probability* Pr, $n \geq 1$, *there exists a Δ_n-probability* Pr′ *such that* Pr *is absolutely continuous with respect to* Pr′, *but not vice versa.*

PROOF. Build Pr as in Theorem 3.12. Now define $\mathrm{Pr}' = \frac{1}{2}\mathrm{Pr} + \frac{1}{2}\widetilde{\mathrm{Pr}}$. Q.E.D.

THEOREM 3.14. *For each $n \geq 1$ there exists a Δ_{n+1}-definable probability such that every Δ_n-definable probability is absolutely continuous with respect to it.*

The proof relies on the following lemma which asserts that all Δ_n-definable probabilities can be enumerated on the Δ_{n+1} level.

LEMMA 3.15. *Let $\mathscr{E}$ be the set of Gödel numbers of elementary sentences. For each $n \geq 1$ there exists a Δ_{n+1} rational-valued function $H(x_0, x_1, y)$ such that, for each $m \in N$, $H(m, x_1, y)$, as (x_1, y) ranges over $\mathscr{E} \times N$, approximates the restriction of a probability Pr_m to elementary sentences and such that $\{\mathrm{Pr}_m : m \in N\}$ is the set of all Δ_n-definable probabilities.*

PROOF OF 3.15. Every Δ_n-probability Pr determines a conditional probability $\mathrm{Pr}(\varphi|\psi)$ whose restriction to $\{(\varphi, \psi): \varphi, \psi\text{-elementary } \mathrm{Pr}(\psi) > 0\}$ is, by Lemma 3.1, Δ_n. We can add to this domain all pairs $(\varphi, 0 = 0)$, φ-elementary, because $\mathrm{Pr}(\varphi|0 = 0) = \mathrm{Pr}(\varphi)$. Now let $\varphi_0, \ldots, \varphi_k, \ldots$ be some fixed recursive enumeration of all atomic empirical sentences and let C be the set of all conjunctions $\bigwedge_{i<k} \pm \varphi_i$, $k = 0, 1, \ldots$ (where $\bigwedge_{i<0} \pm\varphi_i =_{\mathrm{Df}} (0=0)$). Let f be a function mapping C into $[0, 1]$ such that

$$(*) \qquad f(\bigwedge_{i<k} \pm\varphi_i) = \mathrm{Pr}(\varphi_k | \bigwedge_{i<k} \pm\varphi_i),$$

whenever $\mathrm{Pr}(\bigwedge_{i<k}\pm\varphi_i) > 0$. Then $\mathrm{Pr}(\bigwedge_{i<k} \pm \varphi_i) = \prod_{j<k} f^{(j)}(\bigwedge_{i<j}\pm\varphi_i)$, where $f^{(j)}(\ldots)$ is defined as $f(\ldots)$ or as $1 - f(\ldots)$ according as $\pm\varphi_j = \varphi_j$ or $\pm\varphi_j = \neg\varphi_j$. Hence f determines the probability over C. Since every elementary sentence is equivalent to a disjunction of mutually exclusive members of C, it determines Pr. Also, from an approximation to f, we can get an approximation to Pr by applying a fixed recursive scheme. (In order to calculate $\mathrm{Pr}(\bigwedge_{i<k}\pm\varphi_i)$ with error $< \varepsilon$ calculate each $f(\bigwedge_{i<j}\pm\varphi_i)$ with error $< \varepsilon/k^2$, $j = 0, \ldots, k-1$, and form the above-mentioned product of the approximated $f^{(j)}$'s.) Note that *any* mapping f of C into [0, 1] determines a finitely-additive probability over the elementary sentences which satisfies (*). If L has no empirical function symbols then this determines a probability over L. With empirical function symbols, the holding of $\mathrm{Pr}(\bigvee_{i<k}(t = i)) \to_k 1$, for every atomic empirical term t, is not guaranteed. We

shall construct a Δ_{n+1}-function $F(x_0, x_1, y)$, with rational values in [0, 1], satisfying the following:

(i) For each m, $F(m, x_1, y)$, as (x_1, y) ranges over $\{\ulcorner\phi\urcorner : \phi \in C\} \times N$, is an approximating function for some f_m: $C \to [0, 1]$.

(ii) If L has empirical function symbols then the $\Pr_m$ determined by f_m satisfies $\Pr_m(\bigvee_{i<k}(t = i)) \to_k 1$ for all atomic empirical terms t.

(iii) If g is any Δ_n-approximation of a partial function f that determines (through (*)) a probability on L, then, for some $m, f(\phi) = f_m(\phi)$ for all $\phi \in C \cap \text{Domain}(f)$.

Evidently, this will imply 3.15. Let $G(x_0, x_1, y)$ be a rational-valued Δ_n-function such that every rational-valued partial Δ_n-function of (x_1, y) is equal to some $G(m, x_1, y)$. (To obtain this let $U(x_0, x_1, y, z)$ be a Σ_n-relation, where z ranges over Q, such that every Σ_n-relation over $N \times N \times Q$ is equal to some $U(m, x_1, y, z)$. Let $U = \exists z' U_0(x_0, x_1, y, z, z')$ where U_0 is Π_{n-1}. Define $G(x_0, x_1, y) = z$ by the condition: z is the first coordinate in the first pair (z, z') satisfying $U_0(x_0, x_1, y, z, z')$; here we use some fixed recursive enumeration of $Q \times N$.) We shall construct F so that, for each m, if $g(x_1, y) = G(m, x_1, y)$ then $F(m, x_1, y)$ satisfies (iii) with respect to it. First we handle the case where L has no empirical function symbols. Then we shall indicate the modifications necessary to guarantee (ii). To simplify the notation we use 'ϕ' both for ϕ and for its Gödel number $\ulcorner\phi\urcorner$. Let $\phi_0, \phi_1, \ldots, \phi_k, \ldots$ be a fixed recursive enumeration of C. From now on fix m and put $G_m(x_1, y) = G(m, x_1, y)$. We construct the function F_m (i.e., $F(m, \ldots)$) by stages, according to the following program. At the kth stage we define $F_m(\phi_i, k)$ for all $i \leq k$ and $F_m(\phi_k, i)$ for all $i < k$. We also construct an increasing sequence $C_0 \subset C_1 \subset \cdots \subset C_k \subset \cdots$ of finite subsets of C. (Roughly speaking, $\bigcup_k C_k$ consists of those ϕ's for which the function that is presumably approximated by G_m is undefined.) The program may enter a certain state described as "raised flag". Once the flag is raised it stays so throughout the construction. (Roughly speaking, the flag is raised once it is discovered that G_m does not approximate a probability-determining function.)

Put $G'_m(x, y) = G_m(x, y)$ if $G_m(x, y) \in [0, 1]$, $G'_m(x, y) = 0$ if $G_m(x, y) < 0$, $G'_m(x, y) = 1$ if $G_m(x, y) > 1$ and G'_m is undefined if G_m is undefined.

Stage 0. Put $C_0 = \varnothing$. If $G_m(0 = 0, 0)$ is defined put

$$F_m(0 = 0, 0) = G'_m(0 = 0, 0),$$

otherwise put $F_m(0 = 0, 0) = 1$ and raise the flag.

Stage $k + 1$ *for an unraised flag*. If $G_m(\phi_{k+1}, k + 1)$ is defined, put $C_{k+1} = C_k$, otherwise put $C_{k+1} = C_k \cup \{\phi_{k+1}\}$. For $j < k + 1$ put $F_m(\phi_{k+1}, j) = G'_m(\phi_{k+1}, k + 1)$ if $\phi_{k+1} \notin C_{k+1}$, $F_m(\phi_{k+1}, j) = 1$ otherwise. It remains to determine $F_m(\phi_i, k + 1)$ for $i < k + 1$. This is done in k substages. At the ith substage if $\phi_i \notin C_{k+1}$ and if the flag is not raised (i.e., has not been raised at a preceding substage), check whether $G_m(\phi_i, k + 1)$ is defined and if so whether

$$\left[G'_m(\phi_i, k + 1) - \frac{1}{k + 1}, G'_m(\phi_i, k + 1) + \frac{1}{k + 1}\right]$$

$$\cap \bigcap_{j=1}^{k} \left[F_m(\phi_i, j) - \frac{1}{j}, F_m(\phi_i, j) + \frac{1}{j}\right]$$

is not empty. If both hold put $F_m(\psi_i, k+1) = G'_m(\psi_i, k+1)$. Otherwise, put $F_m(\psi_i, k+1) = F_m(\psi_i, k)$ and raise the flag. If the flag is raised already put $F_m(\psi_i, k+1) = F_m(\psi_i, k)$. For $\psi_i \in C_{k+1}$ put $F_m(\psi_i, k+1) = 1$.

Stage $k+1$ *for a raised flag*. Put $F_m(\psi_{k+1}, j) = 1$ for all $j < k+1$ and $F_m(\psi_i, k+1) = F_m(\psi_i, k)$ for all $i < k+1$. (The construction of C_{k+1} is no longer necessary.)

We leave it as an exercise to check that (i) and (iii) hold. The condition that $G_m(x, y)$ is defined is a Σ_n-condition on (m, x, y). Consequently, all clauses in the program are Δ_{n+1}. This implies that $F(x_0, x_1, y)$ is recursive in a Δ_{n+1} relation, hence it is Δ_{n+1}.

If we have empirical function symbols then atomic empirical sentences of the form $t = i$, $t = i'$, where $i \neq i'$, are mutually exclusive. Hence introduce a clause in the program that $F_m(\psi_k, l) = 0$ whenever $\psi_k = \bigwedge_{i<j} \pm \varphi_i$, $\varphi_j = (t = i')$ and $t = i$ is a conjunct in ψ_k. In order to guarantee (ii) we interpose additional stages such that for every term t and every $l \in N$ there will be an interposed stage guaranteeing that for some k, $\Pr(\bigvee_{i<k}(t = i)) > 1 - 1/l$, where Pr is the probability resulting from F_m. These additional stages are used only as long as the flag is not raised. For it is easy to see that if during the execution the flag is raised the resulting Pr satisfies (ii). Hence, in such an interposed stage one assumes that G_m approximates a probability-determining function. One checks the existence of a finite set $C' \subset C$ of an integer k and of a sufficiently large j such that $G_m(\psi, j)$ is defined for all $\psi \in C'$ and such that from these values alone it follows that the probability determined by G_m satisfies $\Pr(\bigvee_{i<k}(t = i)) > 1 - 1/l$. If the checking fails, raise the flag. Again, the passing of such a checking is a Σ_n-condition, hence all clauses are Δ_{n+1}.

Q.E.D.

The technique of the proof can be used to prove a more general enumeration claim where the probabilities enumerated are required to have value 1 for certain consistent $\Sigma_1(L)$-sentences.

PROOF OF THEOREM 3.14. If $\Pr_m$ is the probability of Lemma 3.15, put $\Pr = \sum_m (1/2)^{m+1} \Pr_m$. It is obvious that every Δ_n-definable probability is absolutely continuous with respect to Pr. Using the function H of Lemma 3.15 it is easy to construct a Δ_{n+1} approximating function for Pr. Q.E.D.

Corollary 3.13 and Theorem 3.14 imply that the set of Δ_{n+1} describable probabilities is essentially richer than that of the Δ_n-describable one. On a higher level we can get less dogmatic probabilities. There is a general phenomenon according to which nondogmatism is bought at the price of higher complexity. This seems to be more interesting for measures of complexity more refined than the arithmetical hierarchy. Consider, for example, hypotheses of the form $\forall v(F(v) = g(v))$ where F is an empirical function and g a recursive function. Suppose that Pr assigns nonzero values to all such hypotheses where g varies over some Ω. The Pr that has been constructed in Corollary 3.11 is more complex than all members of Ω. Is it true in general that such a Pr (that is to say, any approximating function for Pr's restriction to elementary sentences) cannot be in Ω? For certain pathological Ω's this need not be the case. For example, using the recursion theorem we can construct a recursive function g having only the values 0 and 1 such that over (Gödel numbers of) elementary empirical sentences g is additive and such that $g(\ulcorner F(n) = g(n)\urcorner) = 1$ for each $n \in N$ (where '$g(n)$' stands for the term in L_0 defining g); taking $\Omega = \{g\}$ we get our counterexample. The question is interesting for natural Ω's that correspond to a subhierarchy of the recursive functions. For example:

Problem. Is there a primitive recursive probability Pr such that

$$\Pr(\forall v(F(v) = g(v))) > 0$$

for every primitive recursive function g?

The generalizations treated in Corollary 3.11 and in Theorem 3.8 express "deterministic" hypotheses, i.e., that a certain empirical function (or set) coincides with some given recursive function (or set). Another sort of hypotheses is that our empirical function, or set, is random, in this or that specified sense of randomness. Hypotheses of this sort are expressible in L as well. This will be seen in §5.

§4. Arithmetical worlds and sets of positive probability.

DEFINITION. A world w is *arithmetical* (with respect to L_0) if all the empirical predicates and function symbols are interpreted as relations and functions definable by formulas of L_0. It is Σ_m (Π_m, Δ_m) if all the interpretations are $\Sigma_m(L_0)$ ($\Pi_m(L_0)$, $\Delta_m(L_0)$). It is ($\Sigma_m + \Pi_m$) if each interpretation is either $\Sigma_m(L_0)$ or $\Pi_m(L_0)$.

There are nonempty sets of the form $\mathrm{Mod}(\varphi)$, even sets containing exactly one world, which do not contain an arithmetical world. Consider, for example, a $\Pi_2(L)$-sentence asserting that a certain empirical predicate is the truth predicate for L_0; it is satisfiable, but not by an arithmetical world. (This sentence was used in Theorem 3.7.) But if $\Pr(\varphi) > 0$, where Pr is definable, then $\mathrm{Mod}(\varphi)$ contains an arithmetical world. The level of this world can be correlated with the level of φ and the level at which Pr is definable. (The other direction is trivial; if w is arithmetical then it is easy to get a definable Pr which accords to $\{w\}$ probability 1.) The results of this section are used when we compare various notions of randomness. In the proofs we shall approximate definable sets of worlds by definable compact subsets which differ from them by arbitrary small probabilities. The following notion, though not essential in the proofs, is employed here in view of its use in the next section.

DEFINITION. Recall that the open subsets $B \subset \mathrm{Mod}_L$ are the unions $\bigcup_{\varphi\in\Xi}\mathrm{Mod}(\varphi)$, where Ξ is a set of conjunctions of basic empirical sentences. If Ξ is definable by a formula $\psi(x)$ of L_0 (i.e. $\Xi = \{\varphi: \models\psi(\ulcorner\varphi\urcorner)\}$), then we say that *$B$ is describable in L_0* and the formula *ψ describes B*. Let $\mathbf{u} = u_0 \ldots, u_{p-1}$, $\mathbf{i} = i_0 \ldots, i_{p-1}$. If $\psi = \psi(\mathbf{u}, x)$ and, for every $\mathbf{i}$, $\psi(\mathbf{i}, x)$ describes $B(\mathbf{i})$ then *ψ describes the indexed family* $(B(\mathbf{i}))_{\mathbf{i}}$. An open set B is $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, $\Delta_n(L_0)$, *recursively*)*-describable* if it can be described by a formula ψ which is $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, $\Delta_n(L_0)$, defines a recursive set); in other words, if it has a representation where Ξ is $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, etc). The same notion is applied to indexed families (in which case Ξ is replaced by $\{(\varphi, \mathbf{i}): \varphi \in B(\mathbf{i})\}$).

Here are some elementary properties:

(4.1) (i) For $n \geq 1$, $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, $\Delta_n(L_0)$)-describable open sets are closed under intersections and unions.

(ii) For $n \geq 1$, $\Sigma_n(L_0)$ ($\Pi_n(L_0)$, etc.)-describable indexed families are closed under bounded intersections and unions. By this we mean that if $(B(\mathbf{i}, j))_{\mathbf{i},j}$ is $\Sigma_n(L_0)$

$(\Pi_n(L_0), \Delta_n(L_0))$-describable so is $(B'(\mathbf{i}, j))_{\mathbf{i},j}$ defined by $B'(\mathbf{i}, j) =_{\mathrm{Df}} \bigcap_{j' \le j} B(\mathbf{i}, j')$. Similarly for unions.

(iii) If f is a $\Delta_n(L_0)$-function and $(B(\mathbf{i}, j))_{\mathbf{i},j}$ is $\Sigma_n(L_0)$ $(\Pi_n(L_0), \Delta_n(L_0))$-describable, then so is $(B'(\mathbf{i}, \mathbf{j}))_{\mathbf{i},\mathbf{j}}$ where $B'(\mathbf{i}, \mathbf{j}) =_{\mathrm{Df}} B(\mathbf{i}, f(\mathbf{j}))$.

The proofs are evident. For example, suppose that $B(\mathbf{i}, j)) = \bigcup\{\mathrm{Mod}(\varphi) : \varphi \in \Xi(\mathbf{i}, j)\}$; then $\bigcap_{j' \le j} B(\mathbf{i}, j') = \bigcup\{\mathrm{Mod}(\varphi) : \varphi \in \Xi'(\mathbf{i}, j)\}$, where $\Xi'(\mathbf{i}, j)$ consists of all conjunctions $\varphi_0 \wedge \cdots \wedge \varphi_j$ where $\varphi_k \in \Xi(\mathbf{i}, k)$, $k = 0, \ldots, j$. It is easy to verify that if $\varphi \in \Xi(\mathbf{i}, j)$ is a $\Sigma_n(L_0)$ $(\Pi_n(L_0), \Delta_n(L_0))$ relation in $(\ulcorner \varphi \urcorner, \mathbf{i}, j)$, so is $\varphi \in \Xi'(\mathbf{i}, j)$.

If ψ describes the open set B (or indexed family $(B(\mathbf{i}))_{\mathbf{i}}$), let $\tilde{\psi}$ be $\exists x[\psi(x) \wedge \mathrm{True}(x)]$, where 'True( )' is a $\Delta_1(L)$ truth definition for $\Sigma_0(L)$-sentences.

Then $B = \mathrm{Mod}(\tilde{\psi})$ (or $B(\mathbf{i}) = \mathrm{Mod}(\tilde{\psi}(\mathbf{i}))$). If $\psi \in \Sigma_n(L_0)$ then $\tilde{\psi} \in \Sigma_n(L)$, and if $\psi \in \Pi_n(L_0)$ then $\tilde{\psi} \in \Sigma_{n+1}(L)$. Hence we get:

(4.2) (i) *Any $\Sigma_n(L_0)$-describable open set, or indexed family, is also $\Sigma_n(L)$ (here $n \ge 1$).*

(ii) *Any $\Pi_n(L_0)$-describable open set, or indexed family, is $\Sigma_{n+1}(L)$.*

Note that the converse is false. There are $\Sigma_2(L)$ open sets which are not describable in L_0. Take a $\Pi_2(L)$ sentence α such that $\mathrm{Mod}(\alpha) = \{w\}$ and w is not arithmetical (say α asserts that the empirical predicate is a truth definition for L_0). Then $\mathrm{Mod}(\neg\alpha)$ is not describable; for if it were we would have $\{w\} = \bigcap_{\varphi \in \Phi} \mathrm{Mod}(\varphi)$, where Φ is a definable subset of $\Sigma_0(L)$, and this would imply that w is arithmetical.

In the sequel we identify Pr with the unique probability Pr* over the Borel field which is determined by it.

THEOREM 4.3. *Let* Pr *be Δ_n-definable. With every formula $\varphi(\mathbf{u})$ of L we can associate, effectively, formulas, $\varphi_0(\mathbf{u}, v, x)$ of L_0 and $\varphi_1(\mathbf{u}, v)$ of L, such that φ_0 describes an indexed family $(B(\mathbf{i}, j))_{\mathbf{i},j}$ of open sets and φ_1 defines this family* (*i.e.*, $B(\mathbf{i}, j) = \mathrm{Mod}(\varphi_1(\mathbf{i}, j))$ *and the following hold*:

(0) *For each* $\mathbf{i}, j$,

$$B(\mathbf{i}, j) \supset \mathrm{Mod}(\varphi(\mathbf{i})) \quad and \quad \Pr[B(\mathbf{i}, j) - \mathrm{Mod}(\varphi(\mathbf{i}))] < 2^{-j};$$

(i) $\varphi \in \Sigma_0(L) \Rightarrow \varphi_0 \in \Delta_1(L_0)$; $\varphi \in \Sigma_1(L) \Rightarrow \varphi_0 \in \Sigma_1(L_0)$, $\varphi_1 \in \Sigma_1(L)$;

(ii) $\varphi \in \Sigma_m(L)$, $m \ge 2 \Rightarrow \varphi_0 \in \Sigma_{m+n-1}(L_0)$, $\varphi_1 \in \Sigma_{m+n-1}(L)$;

(iii) $\varphi \in \Pi_m(L)$, $m \ge 1 \Rightarrow \varphi_0 \in \Delta_{m+n}(L_0)$, $\varphi_1 \in \Delta_{m+n}(L)$;

(iv) $\varphi \in \Delta_m(L)$, $m \ge 1 \Rightarrow \varphi_0 \in \Delta_{m+n-1}(L_0)$, $\varphi_1 \in \Delta_{m+n-1}(L)$.

By '$\varphi_0 \in \Delta_{m+n}(L_0)$' we mean that we can construct $\varphi_0' \in \Sigma_{m+n}(L_0)$, $\varphi_0'' \in \Pi_{m+n}(L_0)$ such that $\models \varphi_0 \leftrightarrow \varphi_0' \leftrightarrow \varphi_0''$; '$\varphi_1 \in \Delta_{m+n}(L)$' is similarly interpreted. In (iv) '$\varphi \in \Delta_m(L)$' means that given $\varphi' \in \Sigma_m(L)$, $\varphi'' \in \Pi_m(L)$ such that $\models \varphi \leftrightarrow \varphi' \leftrightarrow \varphi''$, we can construct etc. Note that in (ii) if we have $\varphi_0 \in \Sigma_{m+n-1}(L_0)$ we can get by (4.2) (i) the desired $\varphi_1 \in \Sigma_{m+n-1}(L)$, but there is no analogous argument for (iii) and (iv). Note also that we can have an indexed family such that for each $\mathbf{i}$, $B(\mathbf{i}, 0) \supset B(\mathbf{i}, 1) \supset \cdots$. This is obtained by passing to $\bigcap_{k \le j} B(\mathbf{i}, k)$; the new formulas φ_0, φ_1 will have the same levels. (Apply (4.1)(ii) for the case of φ_0.) We assume throughout that $n \ge 1$.

PROOF. By induction on φ. For $\varphi \in \Sigma_1(L)$, $\mathrm{Mod}(\varphi(\mathbf{i}))$ is open, hence let $\varphi_1(\mathbf{u}, v)$ be $\varphi(\mathbf{u})$, or say $\varphi(\mathbf{u}) \wedge v = v$. This yields $B(\mathbf{i}, j) = \mathrm{Mod}(\varphi(\mathbf{i}))$ for all j. This family of open sets is describable as follows. If $\varphi \in \Sigma_0(L)$ then each $\varphi(\mathbf{i})$ can be effectively transformed into $\bigvee\!\!\!\bigvee_{k<l}(\alpha_k \wedge \beta_k)$, where α_k is a conjunction of basic empirical sentences and $\beta_k \in \Sigma_0(L_0)$. If $I = \{k < l: \models \beta_k\}$ then $\mathrm{Mod}(\varphi(\mathbf{i})) = \bigcup_{k \in I} \mathrm{Mod}(\alpha_k)$. Using a $\Delta_1(L_0)$ truth definition for $\Sigma_0(L_0)$ we get a $\Delta_1(L_0)$ definition of $\{\alpha_k: k \in I\}$ and we take $\varphi_0(\mathbf{u}, x)$ (with v a dummy variable) as the $\Delta_1(L_0)$-formula which defines this set. If $\varphi = \exists \mathbf{w}\psi(\mathbf{u}, \mathbf{w})$, where $\psi \in \Sigma_0$, let $\psi_0(\mathbf{u}, \mathbf{w}, v, x)$ be the $\Delta_1(L_0)$-formula corresponding to ψ and put $\varphi_0 = \exists \mathbf{w}\psi_0$.

(ii) Let $\varphi = \exists w\psi(\mathbf{u}, w)$, where $\psi \in \Pi_{m-1}(L)$, and suppose that $\psi_0(\mathbf{u}, w, v, x)$ and $\psi_1(\mathbf{u}, w, v)$ are constructed so that $(B(\mathbf{i}, k, j))_{\mathbf{i},k,j}$ = family described by ψ_0 = $(\mathrm{Mod}(\psi_1(\mathbf{i}, k, j)))_{\mathbf{i},k,j}$ has the desired properties. Put

$$B'(\mathbf{i}, j) = \bigcup_k B(\mathbf{i}, k, j + k + 1).$$

Then

$$B'(\mathbf{i}, j) \supset \bigcup_k \mathrm{Mod}(\psi(\mathbf{i}, k)) = \mathrm{Mod}(\varphi(\mathbf{i}))$$

and

$$\Pr[B'(\mathbf{i}, j) - \mathrm{Mod}(\varphi(\mathbf{i}))] < \sum_k \Pr[B(\mathbf{i}, j, j + k + 1) - \mathrm{Mod}(\psi(\mathbf{i}, k))]$$
$$< \sum_k 2^{-(j+k+1)} = 2^{-j}.$$

Put

$$\varphi_0(\mathbf{u}, v, x) = \exists w\psi_0(u, w, v + w + 1, x) \quad \text{and}$$
$$\varphi_1(\mathbf{u}, v) = \exists w\psi_1(\mathbf{u}, w, v + w + 1).$$

Then φ_0 describes $(B'(\mathbf{i}, j))_{\mathbf{i},j}$ and φ_1 defines it and we have $\varphi_0 \in \Sigma_{m+n-1}(L_0)$, $\varphi_1 \in \Sigma_{m+n-1}(L_1)$.

(iii) Let $\varphi = \forall w\psi(\mathbf{u}, w)$, $\psi \in \Sigma_{m-1}(L)$. Let $\psi_0(\mathbf{u}, w, v, x)$ and $\psi_1(\mathbf{u}, w, v)$ be the formulas constructed for ψ and $(B(\mathbf{i}, k, j))_{\mathbf{i},jk}$ the family determined by them, $\psi_0 \in \Sigma_{(m-1)+(n-1)}(L_0)$, $\psi_1 \in \Sigma_{(m-1)+(n-1)}(L)$. Consider the sentence

$$\forall w \leq k\, \psi(\mathbf{i}, w) \wedge \neg \forall w\psi(\mathbf{i}, w);$$

its probability is nonincreasing and tends to 0 as $k \to \infty$. Hence, given j there exists k, which depends on $\mathbf{i}$ and j, for which this probability is $< 2^{-(j+1)}$. We can get such k as a $\Delta_{m+n}(L_0)$ function, say $k^*(\mathbf{i}, j)$, of $(\mathbf{i}, j)$: Rewrite the sentence in $\Pi_m(L)$ form, apply the $\Delta_{m+n}(L_0)$-approximating function for $\Pi_{m+n}(L)$-sentences (Theorem 3.3) where the error is $< 2^{-(j+3)}$ and calculate the approximate values for $k = 0, 1, \ldots$. There exists k for which the approximate value is $< 2^{-(j+2)}$ (for, otherwise, all real values are $\geq 2^{-(j+2)} - 2^{-(j+3)}$). Let $k^*(\mathbf{i}, j)$ be the first k satisfying the inequality. Then for $k = k^*$ the real value is $< 2^{-(j+2)} + 2^{-(j+3)}$. Put

$$B'(\mathbf{i}, j) = \bigcap_{k \leq k^*(\mathbf{i}, j)} B(\mathbf{i}, k, j + k + 2).$$

Then

$$B'(\mathbf{i}, j) \supset \bigcap B(\mathbf{i}, k, j + k + 2) \supset \mathrm{Mod}(\varphi(\mathbf{i}))$$

and

$$\Pr[B'(\mathbf{i}, j) - \mathrm{Mod}(\varphi(\mathbf{i}))] < 2^{-(j+1)} + \sum_{k \le k^*(\mathbf{i}, j)} 2^{-(j+k+2)} < 2^{-j}.$$

Now $B(\mathbf{i}, k, j)$ is describable by ψ_0 which is $\Sigma_{m+n-2}(L_0)$, hence also $\Delta_{m+n}(L_0)$. Since k^* is Δ_{m+n} it follows by (4.1) that $B'(\mathbf{i}, j)$ is $\Delta_{m+n}(L_0)$-describable. It is easily seen that from ψ_0 we can construct the equivalent $\Pi_{m+n}(L_0)$ and $\Sigma_{m+n}(L_0)$ forms of the describing formula φ_0. Finally φ_1 is the formula $\forall w \le k^*(\mathbf{u}, v)\psi_1(\mathbf{u}, w, v + w + 2)$ where '$\forall w < k^*(\mathbf{u}, v)$' is to be formalized via the $\Delta_{m+n}(L_0)$-definition of k^*.

This proves (i)–(iii). To show (iv) suppose that φ is equivalent to $\exists w\psi'(u, w)$ as well as to $\forall w\psi''(\mathbf{u}, w)$ where $\psi' \in \Pi_{m-1}(L)$, $\psi'' \in \Sigma_{m-1}(L)$. Then

$$\Pr(\exists w \le k\, \psi'(\mathbf{i}, w)) \le \Pr(\varphi) \le \Pr(\forall w \le k\, \psi''(\mathbf{i}, w)).$$

The left side is nondecreasing, the right side nonincreasing, and both tend to $\Pr(\varphi)$. Given j there is k such that the absolute difference between the two sides is $< 2^{-(j+1)}$. We can bet such a k as a $\Delta_{m+n-1}(L_0)$ function of $(\mathbf{i}, j)$; for we can approximate the two sides by a Δ_{m-1+n}-function. Set the error as $< 2^{-(j+4)}$ and let k^* be the first k for which the approximate right side is bigger by $< 2^{-(j+2)}$ than the approximate left side. Now repeat the construction used in the proof of (iii), with ψ'' replacing ψ. The argument is the same, but on one level lower, because k^* is Δ_{m+n-1}. Q.E.D.

For every sentence φ of L and every $\varepsilon > 0$ we get a sentence φ' such that $\mathrm{Mod}(\varphi')$ is open, includes $\mathrm{Mod}(\varphi)$ and differs from it by probability $< \varepsilon$. If $\varphi \in \Sigma_m(L)$ or $\varphi \in \Delta_m(L)$, $m \ge 2$, then the corresponding level of φ' is obtained by adding $n - 1$, where n is the level of our Pr. In particular, if Pr is Δ_1-definable, φ' has the same Σ or Δ-level as φ. If $\varphi \in \Pi_m(L)$ then $\varphi' \in \Delta_{m+n}$. It can be shown that for Δ_1-definable probabilities these levels of φ' cannot, in general, be improved. We do not know whether these are the best for Δ_n-definable probabilities where $n > 1$; we conjecture that they are. Apply this construction to $\psi = \neg\varphi$, get the corresponding ψ' and let $\varphi^* = \neg\psi'$. Then $\mathrm{Mod}(\varphi^*) \subset \mathrm{Mod}(\varphi)$ is closed and differs from $\mathrm{Mod}(\varphi)$ by probability $< \varepsilon$. This argument yields:

COROLLARY 4.4. *Let* Pr *be Δ_n-definable, $n \ge 1$. With every formula $\varphi(\mathbf{u})$ one can associate effectively a formula $\varphi^*(\mathbf{u}, v)$ such that for all* $\mathbf{i}, j$, $\mathrm{Mod}(\varphi^*(\mathbf{i}, j))$ *is a closed subset of* $\mathrm{Mod}(\varphi(\mathbf{i}))$ *differing from it by probability $< 2^{-j}$. For $m \ge 1$, $\varphi \in \Pi_m(L) \Rightarrow \varphi^* \in \Pi_{m+n-1}(L)$ (if $\varphi \in \Pi_1(L)$,* $\mathrm{Mod}(\varphi(\mathbf{i}))$ *is already closed and we can put $\varphi^* = \varphi$). Also $\varphi \in \Sigma_m(L) \Rightarrow \varphi^* \in \Delta_{m+n}(L)$, $\varphi \in \Delta_m(L) \Rightarrow \varphi^* \in \Delta_{m+n-1}(L)$.*

We want to get approximations which are compact. If L has no empirical function symbols the closed sets are also compact. In general, we need:

LEMMA 4.5. (i) *Let $\{C_i: i \in I\}$ be a family of closed subsets of* Mod_L *such that for every empirical term t (i.e., of the form $f(\mathbf{m})$ where f is an empirical function symbol) there exists $k_t \in N$ such that* $\mathrm{Mod}(\bigvee_{l<k_t}(t = l))$ *is in the family. If $\bigcap_{i\in I'} C_i \ne \varnothing$ for all finite $I' \subset I$ then $\bigcap_{i\in I} C_i \ne \varnothing$.*

(ii) *A closed $C \subset \mathrm{Mod}_L$ is compact iff for every empirical term t there is $k_t \in N$*

PROOF. (i) Every C_i is $\bigcap_{\varphi\in\Phi_i}$ Mod(φ) where Φ_i consists of elementary sentences. Hence it suffices to prove the claim assuming that each C_i is Mod(φ_i) where φ_i is elementary. Regard the atomic empirical sentences as atomic sentences of the sentential calculus. Get by the compactness theorem an assignment of truth values which makes every φ_i true. This assignment determines a model $w \in \mathrm{Mod}_L$ provided that for every empirical term t there is l such that $t = l$ gets the value 'true'. This holds because $\bigvee\!\!\!\bigvee_{l<k_t}(t = l)$ gets the value 'true'. Evidently $w \in \bigcap_{i\in I}\mathrm{Mod}(\varphi_i)$.

(ii) Any set $A = \bigcap_t \mathrm{Mod}(\bigvee\!\!\!\bigvee_{l<k_t}(t = l))$, where t ranges over all empirical terms, is closed. If $\{C_i\colon i \in I\}$ is a family of closed sets and $A \cap \bigcap_{i\in I'} C_i \neq \varnothing$ for all finite $I' \subset I$, then $A \cap \bigcap_{i\in I} C_i \neq \varnothing$. This follows from (i) by adding all sets $\mathrm{Mod}(\bigvee\!\!\!\bigvee_{l<k_t}(t = l))$ to the family. Consequently, A is compact. If C is closed and $C \subset A$ then C is compact as well. The condition is necessary for compactness, because, for each t, $C \subset \mathrm{Mod}_L = \bigcup_{l\in N} \mathrm{Mod}(t = l)$ and $\mathrm{Mod}(t = l)$ is open. Taking a finite cover we get $C \subset \mathrm{Mod}(\bigvee\!\!\!\bigvee_{l<k_t}(t = l))$. Q.E.D.

If Pr is Δ_n-definable then we have a $\Pi_n(L)$-formula $\tau(v)$ such that for every j, $\mathrm{Mod}(\tau(j))$ is compact and $\Pr(\tau(j)) > 1 - 2^{-j}$. If there are no empirical function symbols let τ be $v = v$. Otherwise use an effective enumeration of all empirical terms: $t_0, \ldots, t_i, \ldots$. Given i and j there exists k such that $\Pr(\bigvee\!\!\!\bigvee_{l<k}(t_i = l)) > 1 - 2^{-(j+i+1)}$. Using the Δ_n-approximating function for Pr over the elementary sentences, we can get such a k as a $\Delta_n(L_0)$-function of (i, j). Let k^* be this function. Let $\tau(v)$ be $\forall u \exists w < k^*(u, v)[\mathrm{True}(\ulcorner t_u = w\urcorner)]$ where 'True( )' is a Δ_1-truth definition for $\Sigma_0(L)$, $\ulcorner t_u = w\urcorner$ describes in L_0 upon substituting i, l for u, w, the Gödel number of $t_i = l$ and '$\exists w < k^*(u, v)$' is to be formalized via the $\Delta_n(L_0)$ definition of k^*. Then

$$\mathrm{Mod}(\tau(j)) = \bigcap_i \mathrm{Mod}\Big(\bigvee\!\!\!\bigvee_{l<k^*(i,j)} (t_i = l)\Big) \quad \text{and}$$
$$\Pr(\tau(j)) > 1 - \sum_i 2^{-(j+i+1)} = 1 - 2^{-j}.$$

Now, if $\varphi^*(\mathbf{u}, v)$ is the formula of Corollary 4.4 and if

$$\hat{\varphi}(\mathbf{u}, v) = \varphi^*(\mathbf{u}, v + 1) \wedge \tau(v + 1),$$

then $\mathrm{Mod}(\hat{\varphi}(\mathbf{i}, j))$ is compact and differs from $\mathrm{Mod}(\varphi(\mathbf{i}))$ by probability $< 2^{-j}$. We get:

THEOREM 4.6. *If* Pr *is Δ_n-definable* ($n \geq 1$), *then with every formula* $\varphi(\mathbf{u})$ *a formula* $\hat{\varphi}(\mathbf{u}, v)$ *can be effectively associated so that for all* $\mathbf{i}$, j, $\mathrm{Mod}(\hat{\varphi}(\mathbf{i}, j))$ *is a compact* (*possibly empty*) *subset of* $\mathrm{Mod}(\varphi(\mathbf{i}))$ *and* $\Pr(\varphi(\mathbf{i}) \wedge \neg\hat{\varphi}(\mathbf{i}, j)) < 2^{-j}$. *Furthermore, for* $m \geq 1$, $\varphi \in \Pi_m(L) \Rightarrow \hat{\varphi} \in \Pi_{m+n-1}(L)$, $\varphi \in \Sigma_m(L) \Rightarrow \hat{\varphi} \in \Delta_{m+n}(L)$ *and, for* $m \geq 2$, $\varphi \in \Delta_m(L) \Rightarrow \hat{\varphi} \in \Delta_{m+n-1}(L)$.

THEOREM 4.7. *If* Pr *is Δ_n-definable* ($n \geq 1$) *then every $\Pi_m(L)$ set of positive probability contains as a member a Δ_{m+2n-1}-world.* (*If* $m = 0$, 1 *and* $n \geq 2$ *we can get the following improvements*: *for* $m = 0$ *a Δ_1-world, for* $m = 1$ *a Δ_{n+1}-world.*)

PROOF. For $m = 0$ the claim is trivial, for any consistent $\Pi_0(L)$ sentence can be satisfied by a world in which the empirical predicates and function symbols are interpreted as finite relations and as functions that are 0 everywhere except

Let $\Pr(\varphi) > 0$ where $\varphi \in \Pi_m(L)$, $m \geq 2$. By 4.6, there exists $\hat{\varphi} \in \Pi_{m+n-1}(L)$ such that $\mathrm{Mod}(\hat{\varphi})$ is a compact subset of $\mathrm{Mod}(\varphi)$ and $\Pr(\hat{\varphi}) > 0$. Enumerate effectively all atomic empirical sentences $\sigma_0, \sigma_1, \ldots, \sigma_i, \ldots$. We shall define a sequence $\sigma_0^*, \ldots, \sigma_i^*, \ldots$ such that (i) every σ_i^* is either σ_i or $\neg\sigma_i$, (ii) $\Pr(\hat{\varphi} \wedge \bigwedge_{i<l} \sigma_i^*) > 0$ for all l, and (iii) σ_i^* (or to be precise its Gödel number) is a Δ_{m+2n-1}-function of i. From (i), (ii) and the compactness of $\hat{\varphi}$ it follows that there is a unique w satisfying all σ_i^*; this w satisfies $\hat{\varphi}$. Also w is Δ_{m+2n-1}. For if $\alpha(\mathbf{u})$ is an empirical atomic formula then for each $\mathbf{k}$ there exists $j = j(\mathbf{k})$ such that $\alpha(\mathbf{k}) = \sigma_j$. The function $j(\mathbf{k})$ is recursive and we have

$$w \models \alpha(\mathbf{k}) \Leftrightarrow \sigma_{j(\mathbf{k})} = \sigma^*_{j(\mathbf{k})}.$$

Since σ_i is a recursive function of i and σ_i^* is a Δ_{m+2n-1}-function, the right-hand side is a $\Delta_{m+2n-1}(L_0)$ condition on $\mathbf{k}$.

Define the σ_i^* by induction on i. Use a Δ_{m+2n-1}-approximating function to get approximate probabilities for $\Pi_{m+n-1}(L)$-sentences (Theorem 3.3). Say $\Pr(\hat{\varphi}) > \varepsilon > 0$ where ε is rational. At least one of $\hat{\varphi} \wedge \sigma_0$, $\hat{\varphi} \wedge \neg\sigma_0$ has probability $> \varepsilon/2$. Set the error as $< \varepsilon/4$ and get the approximate probability of $\hat{\varphi} \wedge \sigma_0$. If this value is $> \varepsilon/2$ put $\sigma_0^* = \sigma_0$, otherwise put $\sigma_0^* = \neg\sigma_0$. Evidently $\Pr(\hat{\varphi} \wedge \sigma_0^*) > \varepsilon/4$. In general, if σ_i^* are defined for $i < l$, set the error as $< \varepsilon(1/4)^l$ and get the approximate probability of $\hat{\varphi} \wedge \bigwedge_{i<l} \sigma_i^* \wedge \sigma_l$. If the value is $> (\varepsilon/2)\cdot(1/4)^{l-1}$ put $\sigma_l^* = \sigma_l$, otherwise put $\sigma_l^* = \neg\sigma_l$. Then $\Pr(\hat{\varphi} \wedge \bigwedge_{i<l} \sigma_i^*) > \varepsilon\cdot(1/4)^l$. The definition of σ_i^* is recursive in the approximating function, hence it is Δ_{m+2n-1}.

Finally, let $\varphi \in \Pi_1(L)$; then $\mathrm{Mod}(\varphi)$ is closed, but not necessarily compact. We construct the sequence $\sigma_0^*, \ldots, \sigma_i^*, \ldots$ using φ instead of $\hat{\varphi}$. If there are empirical function symbols we have, in addition, to ensure that $\mathrm{Mod}(\varphi) \cap \bigcap_i \mathrm{Mod}(\sigma_i^*) \neq \emptyset$. Enumerate effectively the empirical terms $t_0, \ldots, t_i, \ldots$. Using the argument that proves 4.6, we get a Δ_n-function k^* such that $\bigvee_{l<k^*(i)}(t_i = l)$ has probability $> 1 - (1/8)^{i+1}\cdot\varepsilon$ (where $\Pr(\varphi) > \varepsilon > 0$, ε rational). Let θ_i be this sentence. Define σ_j^* to be σ_j if the approximate probability of $\varphi \wedge \bigwedge_{i<j}(\sigma_i^* \wedge \theta_i) \wedge \sigma_j$, with error $< (\varepsilon/4)\cdot(1/8)^{j-1}$ is $> (\varepsilon/2)\cdot(1/8)^{j-1}$; define it as $\neg\sigma_j$ otherwise. An easy calculation shows that $\Pr(\varphi \wedge \bigwedge_{i<j}(\sigma_i^* \wedge \theta_i)) \geq \varepsilon\cdot(1/8)^j$ for all j. By Lemma 4.5(i), $\mathrm{Mod}(\varphi) \cap \bigcap_i \mathrm{Mod}(\sigma_i^*) \neq \emptyset$. Since σ_i^* is a Δ_{n+1}-function of i we get a Δ_{n+1}-world. Q.E.D.

THEOREM 4.8. *Let* Pr *be Δ_n-definable* ($n \geq 1$) *and let S_m be the intersection of all $\Sigma_m(L)$-sets of probability 1. Then*:

(i) *If $m \geq 1$ there exists for every $\varepsilon > 0$ a sentence $\lambda \in \Pi_{m+n-1}(L)$ such that* $\Pr(\lambda) > 1 - \varepsilon$ *and* $\mathrm{Mod}(\lambda) \subset S_m$.

(ii) *For all m, S_m contains as a member a Δ_{m+3n-2}-world. This is also true for* $S_m \cap \mathrm{Mod}(\varphi)$, *where φ is any $\Pi_{m+n-1}(L)$-sentence of positive probability.*

PROOF. (i) Let $\alpha(u) \in \Sigma_m(L)$ be such that every $\Sigma_m(L)$-sentence is equivalent to some $\alpha(i)$. Say $\alpha(u) = \exists \mathbf{v}\psi(u, \mathbf{v})$, $\psi \in \Pi_{m-1}(L)$. Use Theorem 3.3 and get a $\Delta_{m-1+n}(L_0)$-function g which approximates Pr over $\Pi_{m-1}(L)$. Let $k \in N$ be such that $\varepsilon > 1/k$. Let $\gamma(i, j)$ be the condition

$$g(\ulcorner \exists \mathbf{v} < j\, \psi(i, \mathbf{v})\urcorner, k\cdot 2^{i+2}) > 1 - k^{-1}2^{-(i+2)}.$$

Then $\gamma(i,j)$ implies that

$$\Pr(\exists \mathbf{v} < j\, \psi(i, \mathbf{v})) > 1 - k^{-1}2^{-(i+2)} - k^{-1}2^{-(i+2)} = 1 - k^{-1}2^{-(i+1)}.$$

Also

$$j < j' \Rightarrow \mathrm{Mod}(\exists \mathbf{v} < j\, \psi(i, \mathbf{v})) \subset \mathrm{Mod}(\exists \mathbf{v} < j' \psi(i, \mathbf{v})).$$

Hence the intersection of all $\mathrm{Mod}(\exists \mathbf{v} < j\, \psi(i, \mathbf{v}))$, where (i, j) ranges over all pairs satisfying $\gamma(i, j)$, has probability $> 1 - k^{-1} \cdot \Sigma_i\, 2^{-(i+1)} = 1 - k^{-1}$. The intersection is a subset of any $\mathrm{Mod}(\exists \mathbf{v} \psi(i, \mathbf{v}))$ of probability 1; because, if the probability is 1, there exists j satysfying $\gamma(i, j)$ and $\mathrm{Mod}(\exists \mathbf{v} < j\, \psi(i, v)) \subset \mathrm{Mod}(\exists \mathbf{v} \psi(i, \mathbf{v}))$. This intersection is defined by a $\Pi_{m+n-1}(L)$ sentence λ obtained as follows: γ is a $\Delta_{m+n-1}(L_0)$ condition on i, j, hence there is a formula $\tilde{\gamma}(u, w) \in \Sigma_{m+n-1}(L_0)$ that defines it. Let λ be

$$\forall u, w\, (\tilde{\gamma}(u, w) \rightarrow \exists \mathbf{v} < w\, \psi(u, \mathbf{v})).$$

(ii) If $m \geq 1$ use (i) and get $\lambda \in \Pi_{m+n-1}(L)$ such that $\mathrm{Mod}(\lambda) \subset S_m$ and $\Pr(\lambda) > 0$; or, given $\varphi \in \Pi_{m+n-1}(L)$ of positive probability, get λ such that $\Pr(\lambda) > 1 - \Pr(\varphi)$. Then apply Theorem 4.7, with $m + n - 1$ instead of n, to λ, or to $\lambda \wedge \varphi$. If $m = 0$ we imitate the construction of the proof of 4.7. Enumerate effectively all atomic empirical sentences: $\sigma_0, \sigma_1, \ldots, \sigma_i, \ldots$ and construct $\sigma_i^* \in \{\sigma_i, \neg\sigma_i\}$ such that, for all l, $\Pr(\varphi \wedge \bigwedge_{i<l} \sigma_i^*) > 0$, where φ is the given $\Pi_{n-1}(L)$ sentence of positive probability. Then, if $\mathrm{Mod}(\varphi) \cap \bigcap_i \mathrm{Mod}(\sigma^*(i))$ is not empty it consists of one world, which is Δ_k, provided that σ_i^* is a Δ_k-function of i. Evidently this world does not satisfy any conjunction of basic empirical sentences of probability 0, hence it is in S_0. If $n > 1$ we can, by Theorem 4.6, replace φ by $\hat{\varphi} \in \Pi_{2n-2}(L)$ such that $\mathrm{Mod}(\hat{\varphi})$ is compact. This ensures a nonempty intersection. Since $\hat{\varphi} \wedge \bigwedge_{i<l} \sigma_i^*$ is in $\Pi_{2n-2}(L)$, the function $i \rightarrow \sigma_i^*$ can be obtained as a $\Delta_{(2n-2)+n}$-function. If $n = 1$, $\varphi \in \Pi_0(L)$. In this case, to ensure a nonempty intersection, we use the same sort of construction that has been used in the proof of 4.7 for the case $m = 1$. Then the function $i \rightarrow \sigma_i^*$ is Δ_1. Q.E.D.

Question. Can the levels of the arithmetical worlds obtained in 4.7 and 4.8 be improved? For example, assuming a Δ_n-definable Pr, does the intersection of all $\Sigma_m(L)$ sets of probability 1 contain a $(\Sigma_{m+3n-3} + \Pi_{n+3n-3})$-world?

For $n = 1$ the answer is that, in general, the levels cannot be improved. For $n > 1$ the question is open.

A negative answer for 4.8 also implies a negative answer for 4.7, since any improvement in the level of w in 4.7 would lead to an improvement in 4.8. To show that for $n = 1$ these are the best general results, consider a probability that arises out of Bernoulli trials; say P is a monadic empirical predicate and

$$\Pr(\bigwedge_{i \in J} P(i) \wedge \bigwedge_{i \in J'} \neg P(i)) = p^j(1 - p)^{j'},$$

whenever J and J' are disjoint sets of cardinalities j and j', respectively. Here $0 < p < 1$. Note that if Pr is to be Δ_1-definable p should be chosen as a $\Delta_1(L_0)$-real number. The following claim is, however, true for all $p \in (0, 1)$:

For a Bernoulli distribution with $0 < p < 1$, *the intersection of all* $\Sigma_m(L)$-*sets of probability* 1 *contains no* $(\Sigma_m + \Pi_m)$-*world.* (Here $m \geq 1$.)

This is proved by showing that if P is interpreted either as a $\Sigma_m(L_0)$-set, or as a $\Pi_m(L_0)$-set, then this interpretation satisfies a $\Pi_m(L)$-sentence whose probability is 0. We leave it as an exercise for the reader.

§5. **Random worlds.** As stated in the introduction, a world w is ruled out by a probability Pr to the extent that it satisfies sentences whose probability is 0. In order to be possible a world should not satisfy such sentences, i.e., it should satisfy the sentences of probability 1. In general, this notion of possible world depends on the expressive power of the language. If $w \in \bigcap\{\mathrm{Mod}(\varphi): \Pr(\varphi) = 1\}$ but $\Pr^*(\{w\}) = 0$, then we can enrich L_0 so as to have a sentence ψ such that $\{w\} = \mathrm{Mod}(\psi)$. If $\Pr'$ is the unique extension of Pr to the enriched language then $\Pr'(\psi) = 0$; hence with respect to the enriched language w is impossible. This situation does not arise only if $\Pr^*(\{w\}) > 0$ for every $w \in \bigcap\{\mathrm{Mod}(\varphi): \Pr(\varphi) = 1\}$; in that case the whole probability is concentrated on a countable set of worlds. The right way of treating the notion of possible world is to relativize it, from the beginning, to sets of sentences. Following established terminology we shall henceforth use 'random worlds' instead of 'possible worlds'.

DEFINITION 5.1. A world w *is random with respect to a probability* Pr *and a set of sentences* Φ if $\Pr(\varphi) = 1 \Rightarrow w \models \varphi$ for all $\varphi \in \Phi$. We shall express this equivalently by saying that w is (Pr, Φ)-*random*. w is said to be (Pr, L)-*random* if it is random with respect to the set of all sentences of L. When the intended Pr is obvious we shall omit the reference to it and speak of Φ-randomness or L- randomness.

As remarked in the introduction and in §1, this approach enables us to speak of random binary sequences (with respect to a given Pr) as well as of random sequences of reals, random continuous functions, Brownian motions – everything that can be accommodated as "worlds". Also the probability is quite arbitrary.

Evidently, w is Φ-random iff it does not satisfy any of the sentences $\neg\varphi$ of probability 0 where $\varphi \in \Phi$. In [ML] the notion of a "test for randomness" is defined. These tests correspond, as we shall see, to sentences of probability 0. In general, we define:

A Ψ-*test* (*for randomness with respect to a probability* Pr) is a sentence $\psi \in \Psi$ such that $\Pr(\psi) = 0$. A world w *passes the test* if $w \not\models \psi$. A *universal* Ψ-*test* is a Ψ-test, ψ, such that passing ψ is equivalent to passing all the Ψ-tests. In other words, $\psi \in \Psi$ and $\mathrm{Mod}(\psi) = \bigcup\{\mathrm{Mod}(\psi'): \psi' \in \Psi\text{-tests}\}$.

Evidently, if $\Psi = \{\neg\varphi: \varphi \in \Phi\}$ then the existence of a universal Ψ-test is equivalent to the existence of $\varphi \in \Phi$ such that w is Φ-random iff w is $\{\varphi\}$-random.

A $\Sigma_n(L)$-*random world* is a Φ-random world where Φ consists of all $\Sigma_n(L)$ sentences. A $\Pi_n(L)$-random world is similarly defined. Note that we need only one of these notions, for w is $\Pi_n(L)$-random iff it is $\Sigma_{n-1}(L)$-random. This follows by observing that $\Pr(\forall u \varphi(u)) = 1$ iff, for all k, $\Pr(\varphi(k)) = 1$; thus, by satisfying all $\Sigma_{n-1}(L)$-sentences of probability 1 a world also satisfies all $\Pi_n(L)$-sentences of this property. $\Sigma_n(L)$-randomness is therefore randomness at the nth level of the arithmetical hierarchy. Evidently, w is $\Sigma_n(L)$-random iff it passes all $\Pi_n(L)$-tests.

When no confusion is likely we shall omit 'L' and use 'Σ_n-random' and 'Π_n-test' for '$\Sigma_n(L)$-random' and '$\Pi_n(L)$-test'. We shall use 'Pr' to denote also the corresponding measure Pr* over the Borel sets.

In [ML] Martin-Löf defines a test as a set U of pairs (k, s), where $k \in N$, $s \in \{0, 1\}^*$ = set of all finite binary sequences such that (i) U is recursively enumerable, (ii) the sets $U_k =_{\text{Df}} \{s: (k, s) \in U\}$ satisfy $\{0, 1\}^* = U_0 \supset U_1 \supset \cdots$, and (iii) $\mu([U_k]) \leq 2^{-k}$, where μ = the Lebesgue measure on $\{0, 1\}^N$ and $[U_k]$ = the open subset of $\{0, 1\}^N$ determined by U_k (i.e., $\xi \in [U_k]$ iff some initial segment of ξ is in U_k). An infinite binary sequence ξ passes the test at level of significance k if $\xi \notin [U_k]$. It is random if it passes all tests. Obviously, the tests of [ML] are Σ_1-describable families of open sets $(B(i))_i$ such that $\text{Mod}_L = B(0) \supset B(1) \supset \cdots$; all this for the special case where $\text{Mod}_L = \{0, 1\}^N$ and L_0 is the language of Peano's arithmetic (for then the $\Sigma_1(L_0)$-sets are exactly the recursively enumerable ones). Given such a family we have (by 4.2) a $\Sigma_1(L)$-formula $\phi(u)$ such that $B(i) = \text{Mod}(\phi(i))$ and, since $\text{Mod}_L = B(0) \supset B(1) \supset \cdots$, we also have $B(i) = \text{Mod}(\forall u < i\ \phi(u + 1))$. Vice versa, if $\phi(u) \in \Sigma_1(L)$ then $\text{Mod}(\forall u < i\ \phi(u))$ is Σ_1-describable (this is easily seen from the proof of 4.3(i)) and satisfies the above-mentioned conditions. Hence the tests in the sense of [ML] can be represented by the $\Pi_2(L)$-formulas $\forall u \phi(u)$, where $\phi \in \Sigma_1(L)$, for which $\Pr(\forall u < k\ \phi(u)) \leq 2^{-k}$ for all k. A world w passes the test at level k if $w \not\models \forall u < k\ \phi(u)$. The Pr of [ML] is the Lebesgue measure. It represents the set-up of Bernoulli trials with $p = \frac{1}{2}$. In [ML] this is also generalized to trials in which p is a recursive real. But the notion of a $\Pi_2(L)$-test, $\forall u \phi(u)$, such that $\Pr(\forall u < k\ \phi(u)) \leq 2^{-k}$ is meaningful for any language L and any probability Pr. Let us take a closer look at it.

Instead of requiring $\Pr(\forall u < k\ \phi(u)) \leq 2^{-k}$, we can require that there exist a recursive function g such that $\lim_k g(k) = \infty$ and $\Pr(\forall u < k\ \phi(u)) \leq (g(k))^{-1}$. For if g is such a function, let $h(k) =_{\text{Df}}$ smallest l such that $g(l) \geq 2^k$ and let

$$\phi'(u) = \forall u' < h(u)\ \phi(u)$$

where '$\forall u' < h(u)$' is formalized via the $\Delta_1(L_0)$ definition of h. Then $\phi' \in \Sigma_1(L)$, $\models \forall u \phi \leftrightarrow \forall u \phi'$ and $\Pr(\forall u < k\ \phi'(u)) \leq 2^{-k}$. (The same argument applies of course directly to the tests of [ML]: replace the test U by U' such that $U'_k = \bigcap_{i \leq k} U_{h(i)}$. Since h is recursive U' is recursively enumerable.) The fixing of $g(k)$ as 2^k is motivated in [ML] by the connection between the levels of passing tests and Kolmogoroff's proposed measurement of randomness [Ko 2]. This connection is valid if the levels are defined as in [ML], provided that we presuppose independent Bernoulli trials with $p = \frac{1}{2}$. The choice of 2^k is also handy in various calculations. But actually, it all amounts to a recursive estimate of the rate of convergence of $\Pr(\forall u < k\ \phi)$ to 0. This motivates the following definition.

DEFINITION 5.2. Let Q be either $\exists$ or $\forall$. The quantifier $Q\mathbf{u}$ is *effectively convergent in the formula* $Q\mathbf{u}\phi(\mathbf{u}, \mathbf{v})$ (*with respect to the probability* Pr) if there is a recursive function $g(x, \mathbf{y})$ such that, for each $\mathbf{j}$, $g(x, \mathbf{j}) \to \infty$ as $x \to \infty$ and

$$|\Pr(Q\mathbf{u}\phi(\mathbf{u}, \mathbf{j})) - \Pr(Q\mathbf{u} < k\ \phi(\mathbf{u}, \mathbf{j}))| \leq g(k, \mathbf{j})^{-1}.$$

A $\Sigma^e_n(L)$-*formula* ($\Pi^e_n(L)$-*formula*) is a fomula $\exists \mathbf{u}\phi$ ($\forall \mathbf{u}\phi$) such that $\phi \in \Pi_{n-1}(L)$ ($\phi \in \Sigma_{n-1}(L)$) and $\exists \mathbf{u}$ is effectively convergent in $\exists \mathbf{u}\phi$ ($\forall \mathbf{u}$ is effectively convergent in $\forall \mathbf{u}\phi$). Here we assume $n \geq 1$. It is $\Sigma^e_n(L)$, or $\Pi^e_n(L)$, *in normal form* if $\mathbf{u}$ consists of a single u and $|\Pr(Qu\phi(u, \mathbf{j})) - \Pr(Qu < k\ \phi(u, j))| \leq 2^{-k}$ (where ϕ is on a lower level in the hierarchy). The notion of an effectively convergent quantifier is extendible to a quantifier which is preceded by a string of bounded quantifiers: $Q_1v_1 < t_1 \cdots Q_mv_m < t_mQ\mathbf{u}\psi$, where each t_i is either a variable or some $n \in N$. It is easy to see that, for each assignment of values to the free variables, $\Pr(Qv_1 < t_1 \cdots Qv_m < t_mQ\mathbf{u} < k\,\psi)$ converges as $k \to \infty$ to $\Pr(Qv_1 < t_1 \ldots Qv_m < t_mQ\mathbf{u}\psi)$. If the rate of convergence can be effectively estimated, as above, by a recursive function of k and of the values of the free variables, we say that $Q\mathbf{u}$ *is effectively convergent in* $Qv_1 < t_1 \cdots Qv_m < t_mQ\mathbf{u}\psi$.

If $n \geq 2$ then every $\Sigma_n^e(L)$ (or $\Pi_n^e(L)$) formula can be put in equivalent normal form. The argument given before generalizes: Replace $Q\mathbf{u}\psi$ by $Qu\psi'$, where ψ' is $Q\mathbf{u} < h(u, \mathbf{v})\ \psi(\mathbf{u})$ and $h(k, \mathbf{j}) =_{\mathrm{Df}}$ smallest l such that $g(l, \mathbf{j}) \geq 2^k$. (For $n = 1$ this argument fails because the formalization of '$Q\mathbf{u} < h(u, \mathbf{v})$' will, in general, yield a $\Delta_1(L_0)$ but not a $\Sigma_0(L_0)$-formula). Obviously, a negation of a $\Sigma_n^e(L)$-formula is equivalent to a $\Pi_n^e(L)$-formula and vice versa. Hence $\Sigma_n^e(L)$-randomness is equivalent to the passing of all $\Pi_n^e(L)$-tests. As before we often omit 'L' and use 'Σ_n^e-random' 'Π_n^e-test' for '$\Sigma_n^e(L)$-random' and '$\Pi_n^e(L)$-test'. If $\varphi = \forall u\psi(u)$ is a Π_n^e-test in normal form, then we say that *w passes φ at level k* if $w \not\models \forall u < k\psi(u)$.

From all that has been said it follows that *randomness in the sense of* [ML] *is Σ_2^e-randomness, for the special case where the worlds are binary sequences, the probability a Bernoulli distribution with a recursive p and L_0 is a language of Peano's arithmetic.*

(Here by a *language of Peano's arithmetic* we mean a language containing 0, 1, +, $\cdot$ in which all predicates and function symbols refer to recursive relations.)

Note that the class of $\Sigma_n^e(L)$-formulas depends on the probability Pr. As it turns out the Σ_n^e-scale has, when randomness is concerned, some advantages over the Σ_n-scale. The $\Sigma_n^e(L)$ ($\Pi_n^e(L)$)) classes share the following properties of $\Sigma_n(L)$ ($\Pi_n(L)$) classes: They are closed up to equivalence under conjunctions and disjunctions (i.e., $\varphi_1, \varphi_2 \in \Sigma_n^e(L) \Rightarrow \varphi_1 \wedge \varphi_2$ is equivalent to a $\Sigma_n^e(L)$-formula, etc.). They are also closed, up to equivalence, under bounded quantification. Furthermore if $\exists\mathbf{u}$ is effectively convergent in $Qv_1 < t_1 \cdots Qv_m < t_m\exists\mathbf{u}\psi$ where $\psi \in \Pi_{m-1}(L)$, then the formula is equivalent to a $\Sigma_m^e(L)$ one. (The proof is by induction on m. One "pulls in" the bounded quantifiers observing that if $\exists\mathbf{u}$ is effectively convergent in $\forall v < t\exists\mathbf{u}\psi$ then $\exists u'$ is effectively convergent in $\exists u'\forall v < t\exists\mathbf{u} < u'\psi$.) Up to equivalence $\Sigma_n^e(L) \subset \Sigma_{n+1}^e(L)$ (given $\psi \in \Sigma_n^e(L)$, add between the right-most quantifier and the quantifier-free matrix another alternating quantifier over a dummy variable). Note that effective convergence can be also characterized by the following condition. There exists a recursive function h such that for all l if $k = h(l, \mathbf{j})$ then

$$|\Pr(Q\mathbf{u}\psi(\mathbf{u}, \mathbf{j})) - \Pr(Q\mathbf{u} < k\ \psi(\mathbf{u}, \mathbf{j}))| < 1/(l + 1).$$

Martin-Löf's construction of a universal test can easily be generalized to Π_n^e-tests:

THEOREM 5.1. *If* Pr *is Δ_1-definable and $n \geq 2$ then there exists a universal Π_n^e-test. Furthermore, it can be put in normal form ψ such that for every Π_n^e-test φ in normal form, the worlds that pass ψ at level k pass φ at level $k + d$, $k = 0, 1, 2, \ldots,$ where $d = d(\varphi)$ is an integer depending on φ only.*

PROOF. Let $\lambda(x, u) \in \Sigma_{n-1}(L)$ be such that every $\Sigma_{n-1}(L)$-formula with one free variable u is equivalent to some $\lambda(m, u)$. Say $\lambda = \exists v \lambda_0(x, u, v)$ where $\lambda_0 \in \Pi_{n-2}(L)$. Define $\tilde{\lambda}_0(x, u, v)$ as

$$\lambda_0(x, u, v) \wedge \forall u' \leq u[\Pr(\forall u'' < u'+1\, \exists v' < v\, \lambda_0(x, u'', v')) < 2^{-(u'+1)})]$$

where the expression within the square brackets is to be formalized in L_0 by a $\Sigma_{n-1}(L_0)$-formula (this makes $\tilde{\lambda}_0$ a $\Sigma_{n-1}(L)$-formula). This can be done because, by 3.2 and 3.3, the relation $\Pr(\varphi) < q$ where φ ranges over $\Pi_{n-2}(L)$-sentences and q over the rationals is $\Sigma_{n-1}(L_0)$ and we have an effective procedure for transforming $\forall u'' < i+1\, \exists v' < j\, \lambda_0(m, u'', v')$ into $\Pi_{n-2}(L)$ form. The second conjunct in $\tilde{\lambda}_0(m, i, j)$ asserts that for all $i' < i$ the probability of $\forall u'' < i'+1\, \exists v' < j\, \lambda_0(m, u'', v')$ is $< 2^{-(i'+1)}$. If this conjunct (which is in L_0) is true then $\models \tilde{\lambda}_0(m, i, j) \leftrightarrow \lambda_0(m, i, j)$, otherwise $\models \neg \tilde{\lambda}_0(m, i, j)$. Hence, for all m, $\Pr(\forall u < i + 1 \exists v < j\, \tilde{\lambda}_0(m, u, v)) < 2^{-(i+1)}$ for all i, j. It follows that $\Pr(\forall u < i \exists v\, \tilde{\lambda}_0(m, u, v)) \leq 2^{-i}$. Let $\psi_0(w)$ be

$$\exists x \forall u < w + x + 1\, \exists v\, \tilde{\lambda}_0(x, u, v).$$

Then $\psi_0 \in \Sigma_{n-1}(L)$ and we have

$$\Pr(\psi_0(k)) \leq \sum_{m=0}^{\infty} \Pr(\forall u < k + m + 1\, \exists v\, \tilde{\lambda}_0(m, u, v)) \leq \sum_{m=0}^{\infty} 2^{-(k+m+1)} = 2^{-k}.$$

Hence $\forall w \psi_0(w)$ is a Π_n^e-test in normal form. This is our ψ. Let φ be any Π_n^e-test in normal form. Say $\varphi = \forall u \varphi_0(u)$, where $\Pr(\forall u < k\, \varphi_0(u)) \leq 2^{-k}$. With no loss of generality we can, if $k > 0$, replace "$\leq 2^{-k}$" by "$< 2^{-k}$" (for let $\varphi_0'(u) = \forall u' \leq u + 1\, \varphi_0(u')$; then, for $k > 0$, $\Pr(\forall u < k\, \varphi_0'(u)) < 2^{-k}$ and a world that passes $\forall u \varphi_0'$ at level l passes $\forall u \varphi_0$ at level $l + 1$). There is $m \in N$ such that $\varphi_0(u)$ is equivalent to $\exists v \lambda_0(m, u, v)$. For all i, j $\Pr(\forall u < i + 1\, \exists v < j\, \lambda_0(m, u, v)) < 2^{-(i+1)}$, which implies $\models \lambda_0(m, u, v) \leftrightarrow \tilde{\lambda}_0(m, u, v)$. Hence $\models \varphi_0(u) \leftrightarrow \exists v \tilde{\lambda}_0(m, u, v)$. By definition, $\forall u < w + x + 1\, \exists v\, \lambda_0(x, u, v)$ logically implies $\psi_0(w)$. Substituting m for x we have $\models \forall u < w + m + 1\, \varphi_0(u) \to \psi_0(w)$; consequently $\models \forall u < k + m\, \varphi_0(u) \to \forall w < k\, \psi_0(w)$, i.e., every world that passes ψ at level k passes φ at level $k + m$. So put $d(\varphi) = m$. Q.E.D.

The case $n = 1$ is clarified by the following observation:

(5.2) *Suppose that L_0 is a language of Peano's arithmetic and* Pr *is Δ_1-definable. If* $\Pr(\forall \mathbf{u} \psi(\mathbf{u})) = 0$, *where* $\psi \in \Sigma_0(L)$, *then* $\forall \mathbf{u}$ *is effective in* $\forall \mathbf{u} \psi$. *Consequently, Π_1-tests are also Π_1^e-tests, implying that Σ_1-randomness and Σ_1^e-randomness are the same.*

The proof is by observing that we have a $\Delta_1(L_0)$-function for calculating to any desired accuracy the values $\Pr(\forall \mathbf{u} < k\, \psi(\mathbf{u}))$. By the assumption on L_0 this function is recursive. For each k set the error as $< 1/(k + 1)$, let ξ_k be the approximate value and let $g(k)$ be the largest l such that $\xi_k + 1/(k + 1) < 1/l$. Then $\Pr(\forall \mathbf{u} < k\, \psi) < g(k)^{-1}$ and $g(k) \to \infty$.

As we shall see, there is in general for any $n \geq 1$ no universal Π_n-test. This will show that Theorem 5.1 cannot be extended to $n = 1$. It will also show that Σ_n^e-randomness is strictly weaker than Σ_n-randomness, (That Σ_n-randomness implies Σ_n^e-randomness is trivial, for $\Sigma_n^e(L) \subset \Sigma_n(L)$.) On the other hand we have:

LEMMA 5.3. *If* Pr *is* $\Delta_1(L_0)$*-definable then every* Π_n*-test is equivalent to a* Π^e_{n+1}*-test; hence* Σ^e_{n+1}*-randomness implies* Σ_n*-randomness.*

PROOF. Let $\forall \mathbf{u}\varphi$ be a Π_n-test, $\varphi \in \Sigma_{n-1}(L)$. By Theorem 3.3 the condition $\Pr(\forall \mathbf{u} < k'\varphi(\mathbf{u})) < 2^{-(k+1)}$ is a $\Sigma_n(L_0)$-condition on k', k. Let $\alpha(v', v)$ be a formula defining it (where k', k correspond, respectively, to v', v). Let $\psi(v)$ be $\exists v'(\alpha(v', v) \wedge \forall \mathbf{u} < v'\varphi(\mathbf{u}))$.

Evidently $\Pr(\forall v < k\ \psi(v)) \leq 2^{-k}$. Also $\psi \in \Sigma_n(L)$. Since for every k there is k' such that $\Pr(\forall \mathbf{u} < k'\varphi) < 2^{-(k+1)}$ we have $\models \forall v\psi(v) \leftrightarrow \forall \mathbf{u}\varphi(\mathbf{u})$. Hence $\forall v\psi$ is an equivalent Π^e_{n+1}-test. Q.E.D.

Thus Σ^e_{n+1}-randomness is between Σ_n-randomness and Σ_{n+1}-randomness. It will turn out that this "between" is a strict one on both sides.

In [Sc] Schnorr argues that the tests used in statistics are more restricted than the general tests of [ML]. In order to capture the additional restriction he adds the requirement that (in the terminology of [ML]) $\mu([U_k])$ be a recursive function of k—i.e., calculable effectively to any desired accuracy. Thus he gets a narrower class of tests and therefore a weaker notion of randomness. (Apparently he does not require $U_k \supset U_{k+1}$, but this is inessential, for, under his assumptions, an equivalent test that does satisfy this requirement can be constructed.) His setting is the same as in [ML]—binary sequences and Bernoulli distributions with a recursive $p \in (0, 1)$. We indentify his tests with $\Pi_2(L)$-formulas and we get Π^e_2-tests in normal form, $\forall u\, \psi(u)$, such that $\Pr(\forall u < k\ \psi(u))$ is a recursive real-valued function of k. In order to compare this notion with others we note the following.

LEMMA 5.4. *Let* Pr *be* Δ_1*-definable, where* L_0 *is a language of Peano's arithmetic. Let* $\psi(\mathbf{u}, \mathbf{v}) \in \Sigma_0(L)$. Then $\Pr(\forall \mathbf{u} < k\exists \mathbf{v}\psi(\mathbf{u}, \mathbf{v}))$ *is a recursive real-valued function of* k *iff* $\exists \mathbf{v}$ *is effectively convergent in the formula* $\forall \mathbf{u} < w\exists \mathbf{v}\psi(\mathbf{u}, \mathbf{v})$ *(where* w *is a variable not occurring in* ψ*).*

PROOF. Suppose that $\Pr(\forall \mathbf{u} < k\exists \mathbf{v}\psi)$ is a recursive function of k. Calculate the approximate value of $\Pr(\forall \mathbf{u} < k\exists \mathbf{v}\psi)$ with error $< 1/(l + 1)$ and also the approximate value of $\Pr(\forall \mathbf{u} < k\exists \mathbf{v} < l\ \psi)$ with error $< 1/(l + 1)$. Both calculations can be done recursively. Let $\eta(k, l)$ be the absolute difference of the approximate values; then the absolute difference of the real values is in

$$(\eta(k, l) - 2/(l + 1), \eta(k, l) + 2/(l + 1)).$$

Hence $\lim_{l\to\infty} \eta(k, l) = 0$. Let $g(k, l) =$ largest n such that $\eta(k,l) + 2/(l + 1) < 1/n$. Then g is recursive, $\lim_{l\to\infty} g(k,l) = \infty$ and

$$|\Pr(\forall \mathbf{u} < k\exists \mathbf{v}\psi) - \Pr(\forall \mathbf{u} < k\exists \mathbf{v} < l\ \psi)| < g(k, l)^{-1}.$$

In the other direction, suppose that $\exists \mathbf{v}$ is effectively convergent in $\forall \mathbf{u} < w\exists \mathbf{v}\psi$; then we have a recursive g satisfying these properties. Let $h(k, i) =$ smallest l such that $g(k, l) > 2(i + 1)$. Given k, i let $m = h(k, i)$ and calculate $\Pr(\forall \mathbf{u} < k\exists \mathbf{v} < m\ \psi)$ with error $< 1/2(i + 1)$. The resulting value approximates $\Pr(\forall \mathbf{u} < k\exists \mathbf{v}\psi)$ with error $< 1/(i + 1)$. Q.E.D.

We are thus led to Π^e_2-tests $\forall \mathbf{u}\exists \mathbf{v}\psi$, $\psi \in \Sigma_0(L)$, such that $\exists \mathbf{v}$ is effectively convergent in $\forall \mathbf{u} < w\exists \mathbf{v}\psi$. We can generalize this to Π^e_n-tests where $n > 2$ by taking $\psi \in \Sigma_{n-2}(L)$. The condition that the test also be in normal form is not essential, for an equivalent normal form satisfying the additional requirement of effective convergence of $\exists\mathbf{v}$ can be obtained as follows. First convert it into the form $\forall u'\forall\mathbf{u} < h(u')\exists\mathbf{v}\psi$, where h is a suitable recursive function. If $n > 2$ pull $\forall\mathbf{u} < h(u')$ inside and get $\forall u'\exists v\forall\mathbf{u} < h(u')\exists\mathbf{v} < v\,\psi$, where $\forall\mathbf{u} < h(u')\exists\mathbf{v} < v\,\psi \in \Pi_{n-2}(L)$. It is not difficult to verify that $\exists v$ is effectively convergent in $\forall u' < w\exists v\forall\mathbf{u} < h(\mathbf{u}')\exists\mathbf{v} < v\psi$. If $n = 2$ we pull $\forall\mathbf{u} < h(u')$ inside and formalize $h(u')$ using additional existentially quantified variables. We get the form $\forall u'\exists\mathbf{v}'\psi'$. The point is that $\exists\mathbf{v}'$ is effectively convergent in $\forall\mathbf{u}' < w\exists\mathbf{v}'\psi'$. Roughly speaking, since h is a total recursive function, the values of the additional variables in $\mathbf{v}'$ that are needed in the computation of $h(x)$, as $x = 0, 1, \ldots, k - 1$, are bounded from above by a recursive function of k. We leave the details to the reader.

All this leads to the following definition.

DEFINITION 5.3. Let φ be $Q_1\mathbf{u}_1Q_2\mathbf{u}_2 \cdots Q_j\mathbf{u}_j\psi$ where $Q_1\mathbf{u}_1 \cdots Q_j\mathbf{u}_j$ are j alternating quantifiers. Then *φ is $\Sigma^e_{n,j}(L)$ ($\Pi^e_{n,j}(L)$), with respect to* Pr, if it is $\Sigma_n(L)$ ($\Pi_n(L)$) and for all $i = 1, \ldots, j$, $Q_i\mathbf{u}_i$ is effectively convergent in

$$Q_1\mathbf{u}_1 < w_1Q_2\mathbf{u}_2 < w_2 \cdots Q_{i-1}\mathbf{u}_{i-1} < w_{i-1}Q_i\mathbf{u}_i \cdots Q_j\mathbf{u}_j\psi.$$

(Here $w_1, \ldots, w_j$ are distinct variables not occurring in φ. For $i = 1$ the condition is that $Q_1\mathbf{u}_1$ be effectively convergent in φ.) We define $\Sigma^e_{n,0}(L)$ ($\Pi^e_{n,0}(L)$) as $\Sigma_n(L)$ ($\Pi_n(L)$).

Our previous $\Sigma^e_n(L)$ ($\Pi^e_n(L)$) are, in this notation, $\Sigma^e_{n,1}(L)$ ($\Pi^e_{n,1}(L)$). Roughly speaking, the concept of $\Sigma^e_{n,j}(L)$ ($\Pi^e_{n,j}(L)$) means that the effective convergence of the outer quantifier is also carried inside to a depth of j alternating quantifiers. For a fixed j this hierarchy shares with the Σ^e_n-hierarchy the same elementary properties such as closure under conjunctions, disjunctions and bounded quantification, the effect of negation and others. We get:

Randomness in the sense of [Sc] *is $\Sigma^e_{n,2}$-randomness for the case of worlds that are infinite binary sequences, a probability that is a Bernoulli distribution* (*with a recursive p*) *and L_0 which is a language of Peano's arithmetic.*

The concept of [Sc] can be generalized in other ways. For example, consider Π_n-tests $\forall\mathbf{u}\varphi$ such that $\Pr(\forall\mathbf{u} < k\,\varphi)$ is a recursive real-valued function of k (this implies also that the test is Π^e_n). For $n = 2$, L_0 = language of Peano and Pr recursive, these coincide with $\Pi^e_{2,2}$-tests. The coincidence breaks down for $n > 2$ and the relation between the passing of all these tests and $\Sigma^e_{n,2}$-randomness is not clear. It appears, however, that the notions of $\Pi^e_{n,2}$-test and $\Sigma^e_{n,2}$-randomness yield better insight into the situation. The relation between randomness in the sense of [ML] and randomness in the sense of [Sc] (in Schnorr's terminology between hyperrandomness and randomness) is seen as a special case of the relation between $\Sigma^e_{n,1}$ and $\Sigma^e_{n,2}$-randomness. A noneasy generalization of a result of Schnorr shows that $\Sigma^e_{n,2}$-randomness is, in general, strictly weaker than $\Sigma^e_{n,1}$-randomness.

Various easily established inclusions hold between the $\Sigma^e_{n,j}(L)$ and $\Pi^e_{n,j}(L)$ classes (all modulo the identification of equivalent formulas): $n \le n'$ and $k \ge k' \Rightarrow \Sigma^e_{n,k} \subset \Sigma^e_{n',k'}$ (and $\Pi^e_{n,k} \subset \Pi^e_{n',k'}$); $\Sigma^e_{n,k} \subset \Pi^e_{n+1,k+1}$ and $\Pi^e_{n,k} \subset \Sigma^e_{n+1,k+1}$ (by adding a left-most alternating quantifier with a dummy variable); $\Sigma^e_{n,k} \subset \Sigma^e_{n+2j,\,k+2j}$ and $\Pi^e_{n,k} \subset \Pi^e_{n+2j,\,k+2j}$ (by adding $2j$ alternating dummy quantifiers); $\Sigma^e_{n,n} \subset \Sigma^e_{n+j,n+j}$ and $\Pi^e_{n,n} \subset \Pi^e_{n+j,n+j}$ (by adding j alternating dummy quantifiers between the right-most and the $\Sigma_0(L)$ matrix). These imply trivial implications between the corresponding concepts of randomness. The general picture is, however, far from clear.

The notions of [ML] and [Sc] employ $\Sigma_1(L_0)$-describable sequences of open sets (where L_0 = language of Peano's arithmetic); randomness is defined as not belonging to intersections of such sequences of a certain type. A scale of randomness-notions can be based on describable sequences of open sets as follows:

Say that a world w is *O_n-random* if it does not belong to any intersection of open sets $\bigcap_{i\in N} A(i)$ whose probability is 0, where $(A(i))_i$ is a $\Sigma_{n-1}(L_0)$-describable family. Say that w is *O^e_n-random* if it does not belong to any intersection of this form for which there exists a recursive g, such that $g(x) \to \infty$ as $x \to \infty$ and $\Pr(\bigcap_{i<k}A(i)) < g(k)^{-1}$. Here we assume $n \geq 2$.

O_2-randomness is the same as Σ_2-randomness and O^e_2-randomness is the same as Σ^e_2-randomness (i.e. $\Sigma^e_{2,1}$-randomness); because if $\varphi(\mathbf{u}) \in \Sigma_1(L)$ then $(\mathrm{Mod}(\forall \mathbf{u} < i\, \varphi))_i$ constitutes a $\Sigma_1(L_0)$-describable indexed family of open sets, and every descending $\Sigma_1(L_0)$-describable family of open sets can be given this form. Also, $\Sigma^e_{2,2}$-randomness can be characterized by adding a requirement on the indexed families which serve as tests. The situation is not clear if $n > 2$. (For the tests of the form $\bigcap_i A(i)$ can be represented as Π_n-tests but not vice versa.) If the probability is Δ_1-definable then again Σ_n-randomness and O_n-randomness are equivalent and so are Σ^e_n-randomness and O^e_n-randomness. This follows from Theorem 4.3. We do not know whether, for Δ_1-definable probabilities, $\Sigma^e_{n,2}$-randomness can be characterized in terms of the describable open-sets scale. Attempts to find such a characterization lead to certain open questions but we shall not go into it here. There are situations in which the open-sets scale is the more convenient one.

The relative strength of various notions of randomness. If $\Phi' \supset \Phi$ then Φ'-randomness implies Φ-randomness. It is to be expected that, as we increase the class Φ along some natural scale, randomness becomes strictly stronger; e.g., there are Σ_n-random worlds which are not $\Sigma_{n'}$-random if n' is sufficiently larger than n. This is indeed the case, but we have to make a certain assumption concerning Pr. To see that an additional assumption is necessary suppose that the empirical part consists of one monadic predicate P and that $\Pr(P(i)) = 1$ for all $i \in N$. Then $\Pr(\{w_0\}) = 1$ where w_0 is the world in which P is interpreted as N. Hence w is $\{\forall v P(v)\}$-random iff $w = w_0$ and the same is true for every Φ which contains $\forall v P(v)$ as a member (or which contains all sentences $P(i)$). In general, if $\Pr(\{w\}) > 0$ then w is random for any concept of randomness. This indicates that when comparing various notions such worlds play a special role.

DEFINITION. A world w is a *discrete point* (*of the probability* Pr) if $\Pr(\{w\}) > 0$. Evidently the set of discrete points is countable. If $A \subset \mathrm{Mod}_L$ is a Borel subset then Pr *is discrete over* A if $\Pr(A) = \sum_{w\in A}\Pr(\{w\}) > 0$, where the summation is over all discrete points in A. Pr is *discrete* if it is discrete over Mod_L.

The following two easy observations are essential.

(5.5) If Pr is Δ_n-definable, $n \geq 1$, then *there is a* $\Pi_n(L)$*-formula* $\psi(u)$ *such that* $\mathrm{Mod}(\psi(i)) = \{w\colon \Pr(\{w\}) \geq 1/i\}$. *Consequently the set of all discrete points is defined by* $\exists u\psi(u)$ *and is* $\Sigma_{n+1}(L)$.

(5.6) *If* w *is a* Δ_m*-world*, $m \geq 1$, *then there exists a* $\Pi_m(L)$*-sentence* φ *such that* $\mathrm{Mod}(\varphi) = \{w\}$. *If* w *is not a discrete point then* $\Pr(\varphi) = 0$, *implying that* w *is not* Σ_m*-random.*

PROOF OF (5.5). It is easy to see that $\Pr(\{w\}) \geq \varepsilon$ iff $\Pr(\varphi) \geq \varepsilon$ for every elementary sentence φ such that $w \in \mathrm{Mod}(\varphi)$. This means that $\{w\colon \Pr(\{w\}) \geq 1/i\}$ is defined by the condition:

$\forall\varphi\,[\varphi$ is elementary $\wedge$ True$(\varphi) \to \Pr(\varphi) \geq 1/i]$ where True( ) is a $\Delta_1(L)$ truth definition for elementary sentences. Since Pr is Δ_n-definable, the relation $\{(\varphi, i)\colon \varphi\text{-elementary}, \Pr(\varphi) \geq 1/i\}$ is $\Pi_n(L_0)$. Hence the condition is formalizable by a $\Pi_n(L)$-formula.

PROOF OF (5.6). If w is a Δ_m-world then for every empirical predicate R we have a $\Delta_m(L_0)$-formula ψ such that $w \models \forall\mathbf{u}(R(\mathbf{u}) \leftrightarrow \psi(\mathbf{u}))$. This formula is $\Pi_m(L)$. A similar $\Pi_m(L)$-sentence characterizes the interpretation of any empirical function symbol. The finite conjunction of all these sentences is a $\Pi_m(L)$-sentence that defines $\{w\}$. Q.E.D.

Putting together (5.5) and (5.6) with Theorem 4.8 we arrive at the following argument. If Pr is Δ_n-definable let α be the $\Sigma_{n+1}(L)$-sentence that defines the set of discrete points. If Pr is discrete then $\Pr(\alpha) = 1$ and the set of discrete points is the set of all $\{\alpha\}$-random worlds. Since every discrete point is Φ-random, $\{\alpha\}$-randomness is the strongest randomness that we can have. Once Φ includes α there is no further strengthening of Φ-randomness. If Pr is not discrete, $\Pr(\neg\alpha) > 0$ and $\neg\alpha$ is $\Pi_{n+1}(L)$. If $m \geq 2$ then $m + n - 1 \geq n + 1$; hence by 4.8, the intersection of all $\Sigma_m(L)$ sets of probability 1 with $\mathrm{Mod}(\neg\alpha)$ contains a Δ_{m+3n-2}-world. By definition this world is Σ_m-random. Since it is not a discrete point it is not Σ_{m+3n-2}-random. Thus, for each $m \geq 2$, Σ_m-randomness is strictly weaker than Σ_{m+3n-2}-randomness. Instead of $\neg\alpha$ we can use any $\Pi_{m+n-1}(L)$-sentence ψ such that $\Pr(\psi) > 0$ and $\mathrm{Mod}(\psi)$ contains no discrete points. Hence if we have such a ψ in $\Pi_n(L)$ the conclusion holds also for $m = 1$, and if we have one in $\Pi_{n-1}(L)$ the conclusion holds for $m = 0$. All in all we get:

THEOREM 5.7. *Let* Pr *be* Δ_n*-definable*, $n \geq 1$. *Then*:

(i) Pr *is discrete iff there exists* $\alpha \in \Sigma_{n+1}(L)$ *such that* $\{\alpha\}$*-randomness is equivalent to L-randomness* (*i.e., with respect to all sentences of* L). *In this case a world is L-random iff it is a discrete point.*

(ii) *If* Pr *is not discrete then for every* $m \geq 2$, *there exists a* Σ_m*-random world which is not* Σ_{m+3n-2}*-random.*

(iii) *The conclusion in* (ii) *is true for* $m = 1$ *or* $m = 0$, *provided that there is a* $\Pi_{n+m-1}(L)$*-set of positive probability containing no discrete points.*

Question. Is (ii) the best possible result for $n > 1$? That is to say, is $3n - 2$ the smallest increase necessary for strengthening Σ_m-randomness?

Note that the world that we get in (ii) is Δ_{m+3n-2}. It is conceivable that a different a better estimate. Using a different, more complicated, argument, one can improve (iii) by replacing the requirement that Mod(ψ) contains no discrete point by the weaker requirement: for every elementary φ, Pr is not discrete over Mod($\psi \wedge \varphi$). This can be further relaxed, but we do not know whether the nondiscreteness of Pr is sufficient by itself. For Δ_1-definable probabilities we have:

THEOREM 5.8. *If* Pr *is Δ_1-definable and not discrete, then for each $m \geq 3$ there exists no universal Π_m-test. This is also true for $m = 2$ or $m = 1$, provided that there is a $\Pi_{m-1}(L)$-set of positive probability that contains no discrete points.*

REMARKS. If there is no universal Π_m-test, where Pr is Δ_1-definable and $m \geq 2$, then Σ_m-randomness is strictly stronger than Σ_m^e-randomness; for otherwise the universal Π_m^e-test (which exists by 5.1) would also be a universal Π_m-test. Combining this with Theorem 5.7 we see that, for Δ_1-definable Pr's and for $m \geq 3$, Σ_m^e-randomness is equivalent to Σ_m-randomness iff Pr is discrete. We know by 5.2 that for $m = 1$ the two notions are equivalent. This leaves $m = 2$ as the case where the result does not settle the question completely. If Pr is discrete then we have a universal Π_2-test, but this need not be equivalent to the universal Π_2^e-test (we have not checked this thoroughly but apparently examples of such discrete Pr's exist). We do not know if for nondiscrete Pr's the claim of the theorem can be extended to $m = 2$ or $m = 1$ without additional assumptions. But as in 5.7, using a more sophisticated argument these assumptions can be improved in the same way noted in the remark to 5.7.

Note also that the theorem implies that for Δ_1-definable Pr's there is, in general, no universal Π_1-test, i.e., 5.1 cannot be extended to $m = 1$.

PROOF OF 5.8. Let α be the $\Sigma_2(L)$-sentence which defines the set of discrete points. Let $\forall u \psi(u)$ be a Π_m-test where $\psi \in \Sigma_{m-1}(L)$. Then $\Pr(\exists u \neg\psi(u)) = 1$; hence $\Pr(\neg\alpha \wedge \exists u \neg\psi(u)) > 0$, implying that for some i, $\Pr(\neg\alpha \wedge \neg\psi(i)) > 0$. If $m \geq 3$ then, modulo equivalence, $\neg\alpha \wedge \neg\psi(i) \in \Pi_{m-1}(L)$. By 4.7 there is a Δ_m-world, w, satisfying $\neg\alpha \wedge \neg\psi(i)$. There is a $\Pi_m(L)$-sentence φ such that $\{w\} =$ Mod(φ). Since $\Pr(\{w\}) = 0$, φ is a Π_m-test. Evidently Mod(φ) $\not\subset$ Mod($\forall u \psi$). If $m = 1$ or $m = 2$ use the same argument with $\neg\alpha$ replaced by β, where $\beta \in \Pi_{m-1}(L)$, $\Pr(\beta) > 0$ and Mod(β) has no discrete points. Q.E.D.

We conclude this section with some results proved recently by Gaifman concerning the concepts of Martin-Löf [ML] and Schnorr [Sc]. The proofs shall not be included here. Recall that for Δ_1-definable probabilities, Σ_{n+1}^e-randomness implies Σ_n-randomness.

THEOREM. *Let* Pr *be Δ_1-definable*:

(I) *If* Pr *is not discrete then for $m \geq 3$ there is a Σ_m-random world which is not Σ_{m+1}^e-random.*

(II) *The same holds for $m = 1, 2$, provided that there is $\beta \in \Pi_{m-1}(L)$ of positive probability such that for each $\varphi \in \Sigma_0(L)$,* Pr *is not discrete over* Mod($\beta \wedge \varphi$).

The proof is totally different from the previous ones. The world is no longer obtained as an arithmetical world on the corresponding level. A variant of the method also proves:

THEOREM. *Let* Pr *be* Δ_1-*definable*:

(I) *If* Pr *is not discrete and* $m \geq 3$ *then there exists a* $\Sigma^e_{m,2}$-*random world which is not* $\Sigma^e_{m,1}$-*random.*

(II) *The same holds for* $m = 2$, *provided that there exists* $\beta \in \Pi_1(L)$ *of positive probability such that for all* $\varphi \in \Sigma_0(L)$, Pr *is not discrete over* $\mathrm{Mod}(\beta \wedge \varphi)$.

Schnorr [Sc] has shown that there are hyperrandom sequences (i.e., random in the sense of [ML]) which are not random in his sense. Thus he has proved the special case where $m = 2$ and Pr is a Bernoulli distribution over binary sequences with $p = \frac{1}{2}$ (in which case there are, of course, no discrete points). His proof uses martingales and the method is extendible to any recursive $p \in (0, 1)$. But, apparently, there is no generalization of it to arbitrary Δ_1-definable Pr's. Indeed, the proof of this theorem rests on a different approach. The construction is not at all easy.

Randomness with respect to a family of probability distributions. As explained in the introduction, there are many situations in which we are not asking whether certain empirical data is compatible with a particular probability distribution, but whether it is compatible with a certain *type* of distribution, e.g., is this binary sequence a typical case of Bernoulli trials (where the probability, say of 1, is left unspecified)? In general, let **Pr** be a family of probability distributions. To say that a world is possible, or random, with respect to **Pr** is to say that there exists Pr $\in$ **Pr** with respect to which the world is random.

DEFINITION. Let (...) be a notion of randomness. Say that a world w is (...)-*random with respect to* **Pr**, *or* **Pr**, (...) -*random*, if there exists a probability Pr $\in$ **Pr** such that w is (Pr,(...))-random.

Here the notion (...) can be Ω-randomness, where Ω is either some fixed set of sentences or a set that depends on the probability such as Σ^e_n or $\Sigma^e_{n,2}$. Thus to say that a world is Σ^e_n-random with respect to **Pr** is to say that there exist Pr $\in$ **Pr** such that w passes every test which is Π^e_n with respect to this Pr.

The definition uses existential quantification over the probabilities in **Pr**. Usually this is second-order quantification, for **Pr** can be naturally parameterized by a real number or by a real vector. For example, a Bernoulli distribution is determined by the value p, say of getting 1; consequently Bernoulli, (...)-random is the requirement that there exists $p \in [0, 1]$ such that the world is (Pr_p, (...))-random, where Pr_p is the distribution determined by p. Or a finite-state Markov process is determined by a real matrix; so again the condition is that there exists a real matrix such that .... The problem of **Pr**,(...)-randomness is to dispense with the second-order quantification and to characterize the notion by means of sentences in L:

Does there exist a set Φ *of sentences of* L *such that* $\mathrm{Mod}(\Phi)$ *consists exactly of all* **Pr**, (...)-*random worlds*? *If so, find a* Φ *which is as simple as possible, in particular, one which is definable in* L_0.

Equivalently, the problem is to find a set of tests Ψ, as simple as possible, such that **Pr**, (...)-randomness is characterized as the passing of all tests in Ψ. Observe that if Mod(φ) includes all **Pr**, (...)-random worlds, then $\Pr(\varphi) = 1$ for all $\Pr \in$ **Pr**. Hence, in looking for a characterizing set Φ we should restrict ourselves to sentences whose value is 1 under all $\Pr \in$ **Pr**. (Equivalently, the tests should have value 0 for all $\Pr \in$ **Pr**.) Now let Ω be some class of sentences in a natural hierarchy, such as $\Sigma_n(L)$, including also cases where the class depends on Pr, such as $\Sigma_n^e(L)$ or $\Sigma_{n,i}^e(L)$; so in general, $\Omega = \Omega(\Pr)$. If w is **Pr**, Ω-random, then, for some $\Pr \in$ **Pr**, w satisfies all $\varphi \in \Omega(\Pr)$ such that $\Pr(\varphi) = 1$. A fortiori w satisfies all $\varphi \in \Omega^*$, where, by definition, $\varphi \in \Omega^* \Leftrightarrow_{\mathrm{Df}}$ for all $\Pr \in$ **Pr**, $\varphi \in \Omega(\Pr)$ and $\Pr(\varphi) = 1$. But, on the other hand, there is no general reason why the satisfaction of all $\varphi \in \Omega^*$ should imply the existence of $\Pr \in$ **Pr** such that w is Pr, $\Omega(\Pr)$-random. The problem of characterizing **Pr**, Ω-randomness by a set of sentences of L—in particular, the question whether Ω^* is such a characterizing set—has been posed by Gaifman, who has proved positive answers in the case where **Pr** = Bernoulli (i.e. family of Bernoulli distributions over $\{0, 1\}^N$) and $\Omega = \Sigma_n^e(L)$, $n \geq 2$, or $\Omega = \Sigma_{2,2}^e(L)$. (Recall that $\Sigma_n^e(L) = \Sigma_{n,1}^e(L)$.) Furthermore, in these cases one gets a simple characterizing set of sentences which is a subset of Ω^*, definable in L_0. Using this one can construct a single sentence φ such that Mod(φ) is the set of all Bernoulli, $\Sigma_n^e(L)$-random worlds (or Bernoulli, $\Sigma_{2,2}^e(L)$-random worlds). Hence the hypothesis that w is Bernoulli, $\Sigma_n^e(L)$-random (or Bernoulli, $\Sigma_{2,2}^e$-random) is expressible in L and, given some a priori probability distribution, is itself accorded a numerical value. We present the results without proofs. $\Pr_p$ is the Bernoulli distribution with the parameter p.

THEOREM. *For $n \geq 2$ the following are equivalent:*

(I) *There exists $p \in [0, 1]$ such that w is Σ_n^e-random with respect to* $\Pr_p$.

(II) *If $(A(i))_i$ is a Σ_{n-1}-describable family of open sets such that* $\Pr_p(\bigcap_{i<k} A_i) \leq 2^{-k}$ *for all k and all $p \in [0, 1]$ then $w \notin \bigcap_{i\in N} A_i$.*

(III) *If* $\Pr_p(\forall u < k\psi(u)) \leq 2^{-k}$ *for all k and all $p \in [0, 1]$, where $\psi \in \Sigma_{n-1}(L)$, then $w \notin$* Mod($\forall u\psi(u)$).

Evidently, 2^{-k} can be replaced by $g(k)^{-1}$, where g is any recursive function tending to ∞. Thus, (III) means the passing of all Π_n-tests in which the convergence to 0 is at some recursive rate which is equally good for all $p \in [0, 1]$. It means the satisfaction of all $\Sigma_n^e(L)$-sentences in which the convergence to 1 is at the same recursive rate for all $p \in [0, 1]$. Obviously these sentences belong to Ω^* (Bernoulli) for the case $\Omega = \Sigma_n^e(L)$. If $p \in [0, 1]$ is not a recursive real then $\Pr_p(\sigma)$, as σ ranges over elementary sentences, is not a recursive function; however, $\operatorname{Sup}_{p\in[0,1]}\Pr_p(\sigma)$ and $\operatorname{Inf}_{p\in[0,1]}\Pr_p(\sigma)$ are recursive functions. From this it follows that if λ is a formula of L_0 which describes an open set A_λ, then $\operatorname{Sup}_{p\in[0,1]}\Pr_p(A_\lambda)$ and $\operatorname{Inf}_{p\in[0,1]}\Pr_p(A_\lambda)$, as λ ranges over $\Sigma_m(L_0) \cup \Pi_m(L_0)$, are $\Delta_{m+1}(L_0)$-functions. Using this we can enumerate on the nth level a family of $\Pi_n(L)$-tests which characterizes Bernoulli, Σ_n^e-randomness. We can piece them together, as in Martin-Löf's construction, in order to get one universal test; so we get:

For $n \geq 2$ there exists a $\Sigma_n(L)$-formula φ such that Mod(φ) *consists of all Bernoulli, Σ_n^e-random worlds.*

In [ML] Martin-Löf defines Bernoulli sequences (i.e., sequences random with respect to Bernoulli) as sequences that pass all Bernoulli tests, where these are defined as recursively enumerable families $U = (U_i)_{i \in N}$ such that $U_0 \supset U_1 \supset \cdots$, each U_i is a set of finite binary sequences, and for all i, n, k, $U_i(n, k)$ has $< 2^{-i}\binom{n}{k}$ members, where $U_i(n, k)$ = set of all sequences with an initial in U_i which consist of k 1's and $n-k$ 0's. He argues intuitively in favour of this definition. Yet, once we have a notion of $\Pr_p$-randomness, the notion of Bernoulli randomness is already determined, with no extra definition: w is Bernoulli random iff it is $\Pr_p$-random for some $p \in [0, 1]$. It is far from clear that this notion is characterized as the passing of all Bernoulli tests. One can, however, show that the passing of all these tests is equivalent to the passing of all tests of (II) for $n = 2$. So, using the theorem, one concludes that Martin-Löf's definition does in fact define the desired notion, given his previous definition of randomness with respect to specific Bernoulli distributions.

With respect to Bernoulli, $\Sigma^e_{2,2}$-randomness the result is:

THEOREM. *The following are equivalent*:

(i) *For some* $p \in [0, 1]$, *w is* $\Pr_p$, $\Sigma^e_{2,2}$-*random.*

(ii) *If* $(A(i))_i$ *is a family of open sets such that* $\Pr_p(\bigcap_{i<k} A_i) \le 2^{-k}$ *for all* $p \in [0, 1]$ *and all k and if, furthermore, for all i,* $A_i = \bigcup_j A_{i,j}$, *where* $(A_{i,j})_{i,j}$ *is a* $\Delta_1(L_0)$-*describable family of clopen sets (i.e., each* $A_{i,j}$ *is a finite union of basic sets) such that* $|\Pr_p(\bigcap_{i<k} A_i) - \Pr_p(\bigcap_{i<k} \bigcup_{j<l} A_{i,j})| \le 2^{-l}$ *for all* $p \in [0, 1]$ *and all k, l, then* $w \notin \bigcap_i A(i)$.

(iii) *If* $\phi(u, \mathrm{v}) \in \Sigma_0(L)$ *and we have* $\Pr_p(\forall u < k \exists \mathrm{v} \phi) \le 2^{-k}$ *and there exists a recursive g such that* $g(x, y) \to \infty$ *as* $y \to \infty$ *and, for all p,* $|\Pr_p(\forall u < k \exists \mathrm{v} \phi) - \Pr_p(\forall u < k \exists \mathrm{v} < l\, \phi)| \le g(k, l)^{-1}$, *then* $w \notin \mathrm{Mod}(\forall u \exists \mathrm{v} \phi)$.

In [Sc] Schnorr modifies Martin-Löf's notion of Bernoulli tests and obtains a more effective notion which he uses in order to define his version of Bernoulli-random sequences (in his terminology: Zufalsfolgen). This definition corresponds to his version of $\Pr_p$-randomness, for $p = \frac{1}{2}$ (or where p is recursive). As in the previous case his definition can be justified by the theorem, for one can show its equivalence to (ii). (In [Sc] a partial justification is given by showing that for a recursive real $p \in [0, 1]$ the $\Pr_p$-random sequences in his sense, i.e., the Zufällige folgen, are exactly the Zufalsfolgen in which the relative frequency tends to p.)

We do not know whether the last theorem holds in its general version for $n > 2$. In this version $\Sigma^e_{2,2}$ is replaced by $\Sigma^e_{n,2}$, in (ii) $\Delta_1(L_0)$ is replaced by $\Delta_{n-1}(L_0)$ and in (iii) – $\Sigma_0(L)$ by $\Sigma_{n-2}(L)$. For $n > 2$ the relations between the three conditions are at present unclear. (It seems that the generalized (ii) characterizes what can be described as Bernoulli, $O^e_{n,2}$-randomness, a notion formed by using the "open scale" O_n mentioned earlier.)

Results similar to the last two theorems can be proved for certain subfamilies of Bernoulli distributions of the form $\{\Pr_p: p \in X\}$ where $X \subset [0, 1]$. For example, the results are true if X is any interval (open, closed or half-open) with recursive endpoints. These analogous theorems are obtained by having, everywhere, p range over X instead of over [0, 1]. The results obtain also for Bernoulli distributions over $\{0, \ldots, k-1\}^N$, where k is fixed.

The concept of randomness with respect to a family of distributions opens several interesting broad lines of investigation. Of what other natural families of probability distributions is it true that Ω^* characterizes randomness? Markov chains with a given fixed number of states seem a natural candidate.

We conclude by pointing out the program of investigating the notion of randomness for various natural families of distributions and trying to characterize them through tests. With respect to the particular family Bernoulli, the questions of characterizing Σ_n-randomness and, in particular, Σ_2-randomness remain open.

Random worlds and testing procedures. Recall that a testing procedure t is a function from sentences to sentences and that for $w \in \mathrm{Mod}_L$, $\bar{t}(w, n)$ is the outcome of the first $n + 1$ tests obtained by carrying out procedure t in the world w, i.e., $\bar{t}(w, n) = \bar{t}(w, n - 1) \wedge \varphi(w, n)$ where $\varphi(w, n)$ is either $t(t(w, n - 1))$ or its negation, whichever holds in w. In §2 we have seen that if t is a.e. separating then, for almost all w, $\Pr(\psi|\bar{t}(w, n)) \to [\psi](w)$, for all sentences ψ, where $[\psi](w)$ is 1 if $w \models \psi$, 0 otherwise. Say that a testing procedure is *definable* if it is definable (as a function from Gödel numbers to Gödel numbers) by a formula of L_0. If t is definable and all sentences $t(\sigma)$ are of bounded quantifier depth and if, furthermore, Pr is definable, then for each ψ the statement that $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$ is expressible by a sentence in L. If t is a.e. separating this sentence has probability 1. Consequently all (Pr, L)-random worlds satisfy it. Vice versa, if for all ψ, $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$ then w must satisfy all ψ of probability 1 (for $\Pr(\psi) = 1$ implies $\Pr(\psi|\bar{t}(w, n)) = 1$ for all n). Hence we get:

THEOREM. *If* Pr *is definable and t is an a.e. separating definable testing procedure whose values are sentences within some fixed $\Sigma_i(L)$, then the* (Pr, L)*-random worlds are exactly those that satisfy* $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$ *for all* ψ.

Each of the two assumptions that Pr is definable and that t is definable is essential. If either is dropped one can construct a counterexample to the conclusion. (We leave these constructions as exercises for the reader.) But we do not know whether the additional assumption that the values of t have bounded quantifier depth is essential, and we pose it as a problem.

By checking carefully the level at which we get a sentence asserting that $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$, one can arrive at the following estimate:

Let Pr *be Δ_k-definable and let t be an a.e.-separating testing procedure definable by a $\Sigma_j(L_0)$-formula, such that t chooses sentences from $\Sigma_i(L)$. Then in order that* $\Pr(\psi|t(w, n)) \to_n [\psi](w)$ *be true for all $\psi \in \Sigma_m(L)$ it suffices that w be Σ_l-random, where* $l = \mathrm{Max}(k + m, k + i + 1, j)$.

We omit the details.

On the other hand, the argument given before shows that in the converse direction, if $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$ for all $\psi \in \Sigma_m(L)$, then w is Σ_m-random, without any assumptions on Pr and t. We do not know whether, for the nontrivial direction, $\mathrm{Max}(k + m, k + i + 1, j)$ is the best possible estimate and we pose it as a problem. In particular, we pose it for the case where Pr is Δ_1-definable and where t tests all atomic sentences in some fixed recursive order. In this case the maximum is simply $m + 1$ and the question is whether Σ_m-randomness is sufficient to ensure $\Pr(\psi|\bar{t}(w, n)) \to_n [\psi](w)$ for all $\psi \in \Sigma_m(L)$.

Appendix.

PROOF OF REMARK (II) TO THEOREM 3.3. We construct a Δ_n-definable Pr such that, for all m, Pr is not Δ_{m+n-1} over $\Sigma_m(L)$. For $n = 1$ this is true for all Pr (cf. Remark (I)). Assume $n \geq 2$. We construct Pr so as to have the following property: With every $\varphi(\mathbf{v})$ in L_0 a formula $\tilde{\varphi}(\mathbf{v})$ in L can be associated such that $\forall \mathbf{v}(\varphi(\mathbf{v}) \leftrightarrow \tilde{\varphi}(\mathbf{v}))$ has probability 1 and $\varphi \in \Sigma_{m+n}(L_0) \Rightarrow \tilde{\varphi} \in \Sigma_{m+1}(L)$, $\varphi \in \Pi_{m+n}(L_0) \Rightarrow \tilde{\varphi} \in \Pi_{m+1}(L)$. Now take $\varphi(v) \in \Sigma_{m+n-1}(L_0)$ which is not equivalent to any $\Pi_{m+n-1}(L_0)$ formula. Then $\{k \in N\colon \Pr(\tilde{\varphi}(k)) \geq \frac{1}{2}\}$ is not $\Pi_{m+n-1}(L_0)$ (because $\Pr(\tilde{\varphi}(k)) \geq \frac{1}{2} \Leftrightarrow \models \varphi(k)$). Since $\tilde{\varphi} \in \Sigma_m(L)$, Pr is not Δ_{m+n-1} over $\Sigma_m(L)$. To construct Pr take $\gamma(u, v) \in \Sigma_{n-1}(L_0)$ such that every $\varphi(v) \in \Sigma_{n-1}(L_0)$ is equivalent to some $\gamma(i, v)$. Let $R(u, v)$ be an empirical predicate. Define Pr so that $\Pr(R(i, j))$ is 1 or 0 according as $\gamma(i, j)$ is true or false. This condition is recursive in $\{(i, j)\colon \models \gamma(i, j)\}$, hence Pr is Δ_n. (For the conjunctions of basic empirical sentences not containing R use an arbitrary Δ_n-definition.) Let $\langle v_0, \ldots, v_{i-1}\rangle$ be the term describing the encoding of j-tuples. If $\varphi(v_0, \ldots, v_{j-1}) \in \Sigma_{n-1}(L)$, let i be such that $\models \varphi(v_0, \ldots, v_{j-1}) \leftrightarrow \gamma(i, \langle v_0, \ldots, v_{j-1}\rangle)$ and let $\tilde{\varphi}(v_0, \ldots, v_{j-1}) =_{\mathrm{Df}} R(i, \langle v_0, \ldots, v_{j-1}\rangle)$. Extend the definition of $\tilde{\varphi}$ to higher levels by putting $(\exists \tilde{u}\varphi) =_{\mathrm{Df}} \exists u(\tilde{\varphi})$, $(\forall \tilde{u}\varphi) =_{\mathrm{Df}} \forall u(\tilde{\varphi})$. Instead of a 2-place empirical R we could have used a 1-place predicate for we can employ a coding of pairs.

PROOF OF REMARK (III) TO THEOREM 3.3. Given $n \geq 1$ we construct a Δ_n-definable Pr which is not Δ_{n-1}-definable such that, for each m, Pr is Δ_n over $\Sigma_m(L)$ if $m < n$ and is Δ_{m+1} over $\Sigma_m(L)$ if $m \geq n$. This is interesting only if $n \geq 2$, for if $n = 1$ take any $\Delta_1(L_0)$-definable probability that cannot be defined by a L_0-formula with bounded quantifiers. So let $n \geq 2$. We shall choose Pr as a conditional probability $\Pr_1(\ |\psi)$ where $\Pr_1$ is a certain Δ_1-definable probability and ψ is a certain fixed $\Sigma_{n-1}(L)$ sentence such that $\Pr_1(\psi) > 0$. By 3.3, $\Pr_1$ is Δ_{m+1} over $\Sigma_m(L)$. From Lemma 3.1(ii) it follows that $\Pr_1(\varphi_1|\varphi_2)$ is Δ_{m+1} over $\{(\varphi_1, \varphi_2)\colon \varphi_1, \varphi_2 \in \Sigma_m(L), \Pr_1(\varphi_2) > 0\}$. Hence, Pr is $\Delta_{m'+1}$ over $\Sigma_m(L)$ where $m' = \max\{m, n-1\}$. In particular, over all $m' < n$, Pr is Δ_n. It remains to fix $\Pr_1$ and ψ so that Pr is not Δ_{n-1}-definable; this is the main point of the construction. We take $\Pr_1$ as the Lebesgue measure over the infinite binary sequences, where these are the worlds of L. (With no loss of generality we can assume that the empirical part consists of one monadic predicate, for if there are others, extend this probability in any arbitrary Δ_1-definable way.) For every finite binary sequence $s = s_0, \ldots, s_{j-1}$ let $[s]$ = set of all $s' \in \{0, 1\}^N$ such that $s'_i = s_i$, $i = 0, \ldots, j-1$, and let $[S] = \bigcup_{s \in S}[s]$. Construct a sequence $s_0, \ldots, s_i \ldots$ of finite binary sequences such that length$(s_i) = 2(i+1)$, $[s_i] \cap [s_j] = \varnothing$ for $i \neq j$ and $\{s_i\colon i \in N\}$ is Σ_{n-1} but not Π_{n-1}. Let

$$S = \{s\colon \exists i \in N\,([s] \subset [s_i])\}.$$

Then S is seen to be Σ_{n-1} but not Π_{n-1}, $\Pr_1([S]) = \frac{1}{4} + \frac{1}{16} + \cdots = \frac{1}{3}$ and for each s of even length, $s \notin S \Rightarrow \Pr_1([s] \cap [S]) \leq \frac{1}{3} \Pr_1([s])$. It follows that, for each s, $s \notin S \Rightarrow \Pr_1([s] \cap [S]) \leq (\frac{1}{2} + \frac{1}{3}) \Pr_1([s])$. Hence $s \in S \Leftrightarrow \Pr_1([s]|[S]) \geq 3 \Pr_1([s])$. This implies that $\Pr_1(\ |[S])$ is not Δ_{n-1}. Take $\psi \in \Sigma_{n-1}(L)$ such that $\mathrm{Mod}(\psi) = [S]$ (cf. 4.2).

REFERENCES

[Br] LEO BREIMAN, ***Probability***, Addison-Wessly, London, 1968.

[Ca 1] RUDOLF CARNAP, *Remarks on induction and truth*, ***Philosophy and Phenomenological Research***, vol. 6 (1945–46), pp. 590–602.

[Ca 2] ———, ***The logical foundations of probability***, Chicago University Press, Chicago, Ill., 1951.

[Ch] G.J. CHAITIN, *Algorithmic information theory*, ***IBM Journal of Research and Development***, vol. 21 (1977), pp. 350–359.

[Chr] J.P.R. CHRISTENSEN, ***Topology and Borel structure***, ***North-Holland Mathematical Studies***, no. 10, North-Holland, Amsterdam, 1974.

[Fe] J.E. FENSTAD, *Representations of probabilities defined on first order language*, ***Sets, models and recursion theory*** (Crossley, editor), North-Holland, Amsterdam,1967, pp. 156–172.

[Ga 1] HAIM GAIFMAN, *Probability models and the completeness theorem*, ***International Congress of Logic Methodology and Philosophy of Science***, 1960, ***Abstracts of Contributed Papers***, pp. 77–78, mimeographed.

[Ga 2] ———, *Concerning measures in first order calculii*, ***Israel Journal of Mathematics***, vol. 2, no. 1 (1964), pp. 1–18.

[Ga 3] ———, *Subjective probability, natural predicates and Hempel's ravens*, ***Erkenntnis***, vol. 14 (1979), pp. 105–147.

[H-H] PETER HÀJEK and TOMÁŠ HAVRÀNEK, ***Mechanizing hypothesis formation, mathematical foundations for a general theory***, Springer-Verlag, Berlin and New York, 1978.

[Ho] HOSIASSON JANINA, *On confirmation*, this JOURNAL, vol. 5 (1940), pp. 133–148.

[Ko 1] A.N. KOLMOGOROFF, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, ***Ergebnisse der Mathematik und ihrer Grenzgebiete***, no. 2, Berlin, 1933.

[Ko 2] ———, *Three approaches to the definitions of the concept "amount of information"*, ***Problemy Peredači Informacii***, vol. 1 (1965) (Russian); English translation in ***Selected Translations in Mathematical Statistics and Probability***, vol. 7 (1968), American Mathematical Society, Providence, R.I., pp. 293–302.

[Koo] B.O. KOOPMAN, *The axioms and algebra of intuitive probability*, ***Annals of Mathematics*** (2), vol. 41 (1940), pp. 269–292.

[Lo] JERZY ŁOS, *Remarks on foundations of probability*, ***Proceedings of the 1962 International Congress of Mathematics***, Stockholm, 1963.

[ML] PER MARTIN-LÖF, *The definition of random sequences*, ***Information and Control***, vol. 9 (1966), pp. 602–619.

[Pu] HILARY PUTNAM, *Degree of confirmation and inductive logic*, ***The philosophy of Rudolf Carnap*** (Schilp, editor), Library of Living Philosophers, Open Court, 1963, pp. 761–784.

[Ra] F.D. RAMSEY, *Truth and probability* (1926) and *Further considerations* (1928), ***The foundations of mathematics and other essays***, Harcourt-Brace, 1931.

[Sa] LEONARD SAVAGE, ***The foundations of statistics***, Dover, New York, 1972.

[Sc] CLAUS P SCHNORR, *Zufälligkeit und Wahrscheinlichkeit*, ***Lecture Notes in Mathematics***, vol. 218, Springer-Verlag, Berlin and New York, 1971.

[S-K] DANA SCOTT and PETER KRAUSS, *Assigning probabilities to logical formulas*, ***Aspects of inductive logic*** (Hintikka and Suppes, editors), North-Holland, Amsterdam, 1966, pp. 219–264.

HEBREW UNIVERSITY
JERUSALEM, ISRAEL