INFORMATION AND CONTROL 61, 15–37 (1984)

Randomness Conservation Inequalities; Information and Independence in Mathematical Theories*

LEONID A. LEVIN

Boston University, Boston, Massachusetts and
Massachusetts Institute of Technology, Cambridge, Massachusetts

The article further develops Kolmogorov's algorithmic complexity theory. The definition of randomness is modified to satisfy strong invariance properties (conservation inequalities). This allows definitions of concepts such as mutual information in individual infinite sequences. Applications to several areas, like probability theory, theory of algorithms, intuitionistic logic are considered. These theories are simplified substantially with the postulate that the objects they consider are independent of (have small mutual information with) any sequence specified by a mathematical property.

I. ALGORITHMIC INFORMATION

0. *Initial Remarks*

0.1. *Introduction*

Recursive function theory provides analogs of many concepts of classical analysis by requiring countable sets considered to be recursively enumerable (r.e.). The analogy is quite close: intrinsically nonalgorithmic methods are rare in mathematics. Moreover, the general theory of algorithms is very similar to descriptive set theory. There is, however, an important exception in the existence of universal algorithms. The set of all (countable) sets of integers is uncountable while the set of r.e. sets is r.e. This rather abstract difference opens, however, new analytical possibilities having no analogies in "nonalgorithmic" analysis. Let us illustrate this with a simple but important example.

Let $l_1 \subset \mathfrak{R}^{\mathbb{N}}$ be the space of all absolutely summable real sequences: $p \in l_1$, iff $\sum |p(x)| < \infty$. Its recursive analog $\overline{l_1} \subset l_1$ consists of elements of l_1 whose subgraph $\{(r, x): p(x) > r \in \mathbb{Q}\}$ is r.e. It is known in calculus that no element

* Supported in 1978–83 by NSF grants MCS 77-19754, 81-04211, and 83-04498. Correspondence should be addressed to the author, 150-3 Kenrick Street, Boston, Mass. 02135.

0019-9958/84 $3.00

643/61/1-2

is maximal in l_1 within a constant factor: $\forall p \in l_1\ \exists q \in l_1 \lim q(x)/p(x) = \infty$. In contrast to this $\overline{l_1}$ has an "absorbing" element $\mathbf{m}$ (*a universal measure*) such that

$$\forall q \in \overline{l_1}\ \sup(q(x)/\mathbf{m}(x)) < \infty. \qquad \text{Here } \mathbf{m}(x) = \sum r_i(x)/2i^2,$$

where $\{r_i\}$ is an r.e. family of all nonnegative r.e. sequences with the sum bounded by 1.

Complexity $\mathbf{K}(x) = -\lfloor \ln \mathbf{m}(x) \rfloor$ is closely related to the length of a shortest program generating x in an optimal language discovered in (Kolmogorov, 1965; Solomonoff, 1964). Their work originated an invariant approach to information theory, foundations of probability theory, inductive inference, and a number of other areas. Any function of integers, invariant with respect to all recursive transformations, is a constant. However $\mathbf{K}$ is *approximately* invariant, i.e., $\max(\mathbf{K}(\varphi(x)) - \mathbf{K}(x)) < \infty$ for any recursive φ. Concepts like *mutual information* $\mathbf{I}(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$ or *deficiency of randomness* (with respect to a measure μ): $|\ln \mu(x)| - \mathbf{K}(x)$ also have attractive invariance properties suitable for many applications. These ideas of algorithmic information theory, are based on analytical features arising in the recursive analogs of some spaces of classical analysis, as in the above example.

We hope to introduce the reader to the general spirit of the theory by choosing a particular problem and developing the concepts needed for its solution. The problem used for organizing this work is to formalize, justify, and apply the following physical principle:

Let R be the reference to a physical process generating sequence α_R. Let P be a (nonrecursive) mathematical property specifying sequence β_P. The lengths of R and P may be **negligible** compared to the informational content of α_R and β_P; e.g., R may be bibliographical reference to a book α_R and $P(\beta)$ may be "β is the first sequence which cannot be generated by a program of $<10^{10}$ bits." Then it is predicted:

INDEPENDENCE POSTULATE. *The sequences α_R and β_P are independent, i.e.,* $\mathbf{I}(\alpha_R : \beta_P) < |R| + |P|$.

A special case of $\alpha = \beta$ gives a "finitary Church's thesis": Every "physically existing" sequence must be "finitary recursive," i.e., have approximately as short recursive expression as any of its nonrecursive ones. For a more typical example, β might be specified as "the set of all true arithmetical assertions of $<10^{10}$ symbols" and α might be the library of all mathematical publications. To implement these ideas a function $\mathbf{I}: \mathbb{N}^{\mathbb{N}} \times \mathbb{N}^{\mathbb{N}} \to \mathfrak{R}^+$ must be defined satisfying various intuitive and technical requirements.

In Part II the random sequences, the intuitionistic free-choice sequences,

and the representatives of "regular" Turing degrees, respectively, are considered as α. In each of these three cases a formalization of the Independence Postulate is shown to simplify radically the corresponding theories.

0.2. *Brief References*

The following remarks do not present the history of the area and mainly concern the works directly used here. Algorithmic information theory originated with the discovery of universal coding and a recursively invariant approach to the concepts of complexity, information, randomness, and *a priori* probability (Kolmogorov, 1965; Solomonoff, 1964). *Uspekhi Mat. Nauk* announced Kolmogorov's talks on this subject in 1961 and following years. Some of Solomonoff's ideas were mentioned in (Minsky, 1962) and preprints. See also (Markov, 1964) and (Chaitin, 1966, 1969).

Despite the depth of the main idea, the technical expression of basic quantities was not accurate. Many important relationships hold only with an error such as the logarithm of complexity. This error rate is negligible in comparison to complexity itself, but can exceed such differences as mutual information, deficiency of randomness, etc. The errors distorted the picture and hindered the development of a transparent theory. The concept of randomness was improved in (Martin-Lof, 1966) for the case of recursive measures. It was not, however, extendible to other important cases nor expressible in terms of a measure-independent notion like complexity. Very interesting studies of randomness were made in (Schnorr). Some of its ideas proposed independently of (Zvonkin and Levin, 1970) are related to Propositions 4 and 5 below.

The problem of giving a precise expression for information proved to be more difficult. The first nontrivial results were obtained by Kolmogorov and Levin in 1967. The initial definition of $\mathbf{I}(x:y)$ from (Kolmogorov, 1965) was asymmetric and not monotone over y (with respect to projection $(y_1, y_2) \to y_1$). In (Kolmogorov, 1968; Zvonkin and Levin, 1970) this definition was demonstrated to coincide, within the logarithm of complexity, with a symmetric expression, and to be therefore approximately monotone over both arguments.

In (Zvonkin and Levin, 1970) the universal measure was introduced. Its logarithm (equal to the length of the shortest self-delimiting (or prefix) code) turned out to be a more satisfactory complexity measure on $\mathbb{N}$ than the original proposal from (Kolmogorov, 1965). It allowed improvement of the definitions of randomness (Levin, 1973a) and information (Levin, 1974). The new definition of information was monotonic within an additive constant (rather than logarithm) and extendible to the case of infinite sequences. This work is related to subtle results of (Gacs, 1974) concerning the differences between the symmetric and asymmetric expressions for information. A number of results of (Kolmogorov, 1968; Levin, 1970, 1973, 1974; Gacs,

1974) were also found independently in (Chaitin, 1975). A version of the present paper appeared as an MIT technical report MIT/LCS/TR-235 (1980) and some results were formulated in (Levin, 1970–77).

0.3. *Conventions*

$\mathbb{N}$, $\mathbb{Q}$, $\mathfrak{R}$ are the sets of natural, rational, and real numbers. T^+ is the subset $\{x: x \geqslant 0\}$ of an ordered set $T \supset \{0\}$, and $\bar{T}$ is $T \cup \{\infty\}$. The integer part of $x \in \mathfrak{R}$ is $\lfloor x \rfloor$. The number $m + ((m+n)(m+n+1)/2)$ is called *the pair* (m, n) of numbers $m, n \in \bar{\mathbb{N}}$. This enumeration of pairs is bijective on $\mathbb{N}^2 \subset \bar{\mathbb{N}}^2$. Cantor's perfect set is represented in the form $\Omega = \bar{\mathbb{N}}^{\mathbb{N}}$ which has simpler (than $\{0, 1\}^{\mathbb{N}}$) expression for pairs: $(\alpha, \beta)(i) = (\alpha(i), \beta(i))$, where $\alpha(i)$, $\beta(i) \in \bar{\mathbb{N}}$ are the ith terms of α and β. Let $S_k = \{0, 1, ..., k, \infty\}^k$ and $S = \bigcup S_k$; the *length* $l(x)$ of $x \in S_k$ is k. If $\alpha \in \Omega$ or $\alpha \in S_n$ and $k \leqslant n$, then $\alpha_k \in S_k$ is the initial k-segment of α with all terms $\alpha(i) > k$ replaced by ∞; $x \subset \alpha$ means $x = \alpha_{l(x)}$, $l(x) \leqslant l(\alpha)$.

Any property P is identified with the set $\{x: P(x)\}$ and its characteristic function. An open subset of a topological space with natural countable basis is called *recursively enumerable* (*r.e.*) if it equals the union of an r.e. family of basis sets. A real function F on such set is called r.e. if its *subgraph*, i.e., the set $\{(x, r): r < F(x)\}$ is r.e. It is called *recursive* if F and $-F$ are r.e. The symbols $\prec$, $\succ$, and $\sim$ denote inequality and equality of functions within an additive constant; $\preccurlyeq$, $\succcurlyeq$, and $\simeq$ denote these relations within a constant factor. Such operations, as $\sum f$, $\sup f$, $\min f$, etc., are assumed taken over the values of *all* variables of the term f, not bounded in the context.

Conventions for Section 2. Let $\mathscr{F}$ be the space of continuous functions $f: \Omega \to \mathfrak{R}$ with the norm $\|f\| = \sup |f(x)|$; its countable dense subset $\mathscr{F}' \subset \mathscr{F}$ consists of the functions whose ranges are finite sets of rationals. Let $\mathscr{S}$ be the space of lower semicontinuous functions $\Omega \to \bar{\mathfrak{R}}$ and $\mathscr{M}$ be the set of positive linear functionals $\mathscr{F} \to \mathfrak{R}$. Any $\mu \in \mathscr{M}$ represents a measure on Ω and $\mu(f)$ is the average value of f. Let $\mathscr{N}$ be the set of positive linear operators $\mathscr{F} \to \mathscr{F}$. Any $A \in \mathscr{N}$ represents a continuous random transformation of Ω and $A(f)$ maps α to the average value of f on the image of α. We identify $\alpha \in \Omega$ with the measure $\mu_\alpha(f) = f(\alpha)$, and a deterministic transformation $A: \Omega \to \Omega$ with the operator $f \to g$, where $g(\alpha) = f(A(\alpha))$. Restricting $\mu \in \mathscr{M}$ to $S \subset \mathscr{F}$ gives $\mu': S \to \mathfrak{R}$, such that $\mu'(x) = \sum \mu'(y)$, $y \in \sigma(x)$, where $\sigma(x) = \{y: y \supset x, l(y) = l(x) + 1\}$ and μ is uniquely determined by such μ'. Any r.e. measure is recursive.

Random partial processes generate sequences which may stop after a finite number of terms. Their probability distributions satisfy only the inequality $\mu(x) \geqslant \sum \mu(y)$, $y \in \sigma(x)$. This leads to the space $\mathscr{H}$ of **semimeasures**, i.e., positive, uniform, concave functionals $\mu: \mathscr{F} \to \mathfrak{R}$, where $\mu(\mathscr{F}^+) \subset \mathfrak{R}^+$ and $\mu(\rho f + g) \geqslant \rho\mu(f) + \mu(g)$ for $\rho \in \mathfrak{R}^+$. $\mathscr{H}$ is normed by $\|\mu\| = |\mu(-1)|$ and ordered as functions on $\mathscr{F}^+$. Any semimeasure equals the infimum of the

measures majorizing it. Let $\bar{\mu}$ be the largest measure not exceeding μ. A semimeasure μ like a measure can be extended to $-\mathscr{S}$, as $\mu(f) = \inf\{\mu(f'): f \leqslant f' \in \mathscr{F}\}$ and to $(\bar{\mathfrak{R}}^+)^\Omega$, as $\mu(f) = \sup\{\mu(f'): f \geqslant f' \in -\mathscr{S}\}$. If $f \in \mathscr{S}$, then $\mu(f) = \sup\{\mu(f'): f \geqslant f' \in \mathscr{F}\}$. Random partial transformations are represented by positive uniform concave operators $A: \mathscr{F} \to \mathscr{S}$ forming the set $\mathscr{S}$. By $A(\mu)$ we mean $\mu' \in \mathscr{H}$ such that: $\mu'(f) = \mu(A(f))$. All these considerations can be justified using the Hahn–Banach theorem and related results of functional analysis.

The proofs are succinct and require slow reading with frequent reference to the present section. They are, however, independent; one may skip many of them and still understand the others.

1. *Discrete Case*

1.1. *Complexity, Randomness, and Information*

Complexity $\mathbf{K}(x)$ defined in Subsection 0.1 determines the length of the shortest codes of integers x. A *prefix* algorithm $A: \{0,..., k-1\}^* \to \mathbb{N}$ is one defined on at most one prefix of any string. Informally, A recognizes the end of input with no special mark and rejects any continuation. So it has a truly $k$ary input alphabet. The volume $|x|$ of $x \in \{0,..., k-1\}^n$ is $\lceil n \cdot \ln k \rceil$.

PROPOSITION 1 (Coding). *A prefix algorithm A (the Huffman code) exists generating any $x \in \mathbb{N}$ from some input of volume $\mathbf{K}(x)$; i.e., $\exists A,\ c\ \forall x\ \exists p\ (A(p) = x$ and $|p| = \mathbf{K}(x) + c)$.*

No prefix algorithm A can be better, i.e., $\forall A\ \exists c\ \forall x\ \forall p\ ((A(p) = x) \Rightarrow |p| \geqslant \mathbf{K}(x) - c)$.

It is not known how efficient Proposition 1 is. Is it easy to find (or at least to execute) a (generating x) program p of volume $|p|$ close to $|\ln \mu(x)|$, if measure $\mu \in l_1$ is computable in polynomial time?

Proof. The set $G = \{(x, n): n > \mathbf{K}(x)\}$ is r.e. Then a recursive bijection $f: \mathbb{N} \to G$ exists. Let $\lambda(x, n) = e^{-n}$ and $\mu(t) = \sum\{\lambda(f(t')): t' < t\}$. Then $\mu(\infty) = \sum \lambda(x, \mathbf{K}(x) + i) = \sum e^{-\mathbf{K}(x)} \sum e^{-i} < 1$. Let $f(t) = (x, n)$ and p be the shortest $k$ary fraction within $(\mu(t), \mu(t+1))$. Then $A(p) = x$.

Vice versa, A is defined on at most one prefix of any string, and is extendible to $\{0,..., k-1\}^{\mathbb{N}}$. The uniform measure on $\{0,..., k-1\}^{\mathbb{N}}$ is $B(\{\alpha: \alpha \supset q\}) = k^{-l(q)}$. If $K'(x) = \min\{|q|: A(q) = x\}$ then, obviously, $e^{-K'(x)} \leqslant B(A^{-1}(x)) < 1$ and $e^{-K'} \in \overline{l_1}$. Q.E.D.

So $\mathbf{K}(x)$ measures the minimal information needed to generate x. This makes definition of $\mathbf{I}(x:y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$ more intuitive and also agrees with Shannon's idea that the amount of information in an event equals the negative logarithm of its probability. Obviously, $\mathbf{I}(x:y) > 0$, because $\mathbf{m}^2(x, y) = \mathbf{m}(x)\,\mathbf{m}(y)$ is an r.e. measure and thus $\mathbf{m}^2 \leqslant \mathbf{m}$.

Let us arrive at the expression $\mathbf{I}(x:y)$ from another point of view: the

concept of randomness. Let μ be a recursive measure on $\mathbb{N}$ and $t: \mathbb{N} \to \bar{\mathfrak{R}}^+$ be an r.e. function with average value $\mu(t) \leqslant 1$. If $x \in \mathbb{N}$ appears randomly with probability $\mu(x)$ one may expect $t(x)$ to be not much larger then average. Then $\lfloor \ln t(x) \rfloor$ may serve as a randomness test (μ-test) for x.

Note 1. For any r.e. measure μ, $\mathbf{d}(x/\mu) = \lfloor \ln(\mathbf{m}(x)/\mu(x)) \rfloor$ is the largest (within an additive constant) μ-test.

So, $\mathbf{d}(x/\mu) \sim |\ln \mu(x)| - \mathbf{K}(x)$ is, in a sense, a universal characteristic of "nonrandomness," called the *randomness deficiency of x with respect to μ.* Motivations, some history, and the general formulation of the concept of randomness are discussed in Chapter 3.

Let two random variables be independent and have the same distribution μ. Then their joint distribution is $\mu^2(x, y) = \mu(x)\mu(y)$. Suppose a pair $(x, y) \in \mathbb{N}^2$ looks random for probability distribution $\mathbf{m}^2$, i.e., $\mathbf{d}((x, y)/\mathbf{m}^2)$ is small. This means that (1) x and y look independent and (2) each of them looks random for distribution $\mathbf{m}$. But (2) is vacuously true, since *all* numbers look random for the universal distribution $\mathbf{m}$: $\mathbf{d}(x/\mathbf{m}) \equiv 0$.

Therefore, the smallness of $\mathbf{d}((x, y)/\mathbf{m}^2)$ means only that x and y could be generated independently of each other. It is natural then to consider $\mathbf{d}((x, y)/\mathbf{m}^2) = \mathbf{I}(x : y)$ as the *deficiency of independence.* This reminds one of the theorem of classical probabilistic information theory in which two random variables are independent iff they have no mutual information. The difference is that the concepts given above are applicable to individual objects themselves, and not only to their probability distributions (i.e., random variables).

1.2. *Conservation of Independence*

The information $\mathbf{I}(x : y)$ has a remarkable invariance; it cannot be increased by random or deterministic (recursive) processing of x or y. This is natural, since if x contains no information about y then there is little hope to find out something about y by processing x. (Torturing an uninformed witness cannot give information about the crime!)

PROPOSITION 2 (Independence Conservation). *Let $f: \mathbb{N} \to \mathbb{N}$ be a recursive function, and φ be an r.e. measure on $\mathbb{N}$. Then*

(1) $\mathbf{I}(f(x) : y) \prec \mathbf{I}(x : y)$,

(2) $\int \exp(\mathbf{I}((x, z) : y))\, d\varphi(z) \leqslant \exp(\mathbf{I}(x : y))$.

The linear scale (instead of the logarithmic one) strengthens (2) and is more natural with linear operator $\int$. Proposition 2 is an elementary version of the corollary of Theorem 1 below and also implies independence conservation in any combination of random and deterministic (recursive) processes. This supports the Independence Postulate in Subsection 0.1.

Proof. Lemma 1 (Gacs, 1974). $\mathbf{K}(x, \mathbf{K}(x)) \sim \mathbf{K}(x)$.

This elegant lemma has a short proof: let p be a ***shortest*** code for x. Obviously, both x and $\mathbf{K}(x) = |p|$ are computable from p. Therefore, the complexity of $(x, \mathbf{K}(x))$ equals $|p| = \mathbf{K}(x)$.

Let the *universal conditional* measure be the largest within a constant factor r.e. function $\mathbf{m}(/)\colon \mathbb{N}^2 \to \mathfrak{R}^+$ such that $\sup_y \sum \mathbf{m}(x/y) < \infty$, and $\mathbf{K}(x/y) = -\lfloor \ln \mathbf{m}(x/y) \rfloor$.

LEMMA 2. $\mathbf{K}(x, y) \sim \mathbf{K}(x) + \mathbf{K}(y/(x, \mathbf{K}(x)))$.

Let $m_\infty(y/x, n) = e^n \mathbf{m}(x, y)$. A nondecreasing by k, recursive sequence $m_k(y/x, n)\colon A_k \to \mathbb{Q}^+$ exists, such that $m_\infty = \sup m_k$, where $A_k \subset \mathbb{N}^3$ are finite. Let $k(x, n) = \sup\{k\colon \sum m_k(y/x, n) \leqslant 1\}$, and $\bar{m}(y/x, n) = m_{k(x,n)}(y/x, n)$. Obviously $\forall x,\ n \sum \bar{m}(y/x, n) \leqslant 1$ (thus $\bar{m}(/) \leqslant \mathbf{m}(/)$) and $\forall x, n$ if $\sum \mathbf{m}(x, y) \leqslant e^{-n}$, then $m_\infty(y/x, n) = \bar{m}(y/x, n)$. Therefore $\forall x, n$ if $\sum \mathbf{m}(x, y) \leqslant e^{-n}$ (and thus $\mathbf{m}(x) \leqslant e^{-n}$ or $n > \mathbf{K}(x)$) then $\mathbf{m}(y/x, n) \geqslant \bar{m}(y/x, n) = m_\infty(y/x, n) = e^n \mathbf{m}(x, y)$. Thus $\mathbf{K}(y/x, \mathbf{K}(x)) < \mathbf{K}(x, y) - \mathbf{K}(x)$.

It remains to prove that $\mathbf{K}(y/(x, \mathbf{K}(x))) > \mathbf{K}(x, y) - \mathbf{K}(x) \sim \mathbf{K}(x, y) - \mathbf{K}(x, \mathbf{K}(x))$. This follows from $\mathbf{K}(x, y) < \mathbf{K}(y, x, \mathbf{K}(x))$, $\mathbf{K}(x) \sim \mathbf{K}(x, \mathbf{K}(x))$, and $\mathbf{K}(y, t) < \mathbf{K}(t) + \mathbf{K}(y/t)$. The latter inequality holds since $m'(y, t) = \mathbf{m}(t)\, \mathbf{m}(y/t)$ is obviously an r.e. measure and then $m'(y, t) \leqslant \mathbf{m}(y, t)$.

Now, $\mathbf{K}(x, y, z) < \mathbf{K}(x, \mathbf{K}(X)) + \mathbf{K}(y/(x, \mathbf{K}(x))) + \mathbf{K}(z/(x, \mathbf{K}(x))) \sim \mathbf{K}(x, y) + \mathbf{K}(x, z) - \mathbf{K}(x)$, since $\mathbf{K}(y, z/t) < \mathbf{K}(y/t) + \mathbf{K}(z/t)$. Therefore $\mathbf{I}((z, x) : y) > \mathbf{I}(x : y)$ and (1) follows by noting that $\mathbf{I}(z : y) \sim \mathbf{I}((z, x) : y)$ for $x = A(z)$, since z and $(z, A(z))$ are computable from each other.

To prove (2) we need to show that $\mathbf{m}(x, y)/(\mathbf{m}(x)\, \mathbf{m}(y)) \geqslant \int (\mathbf{m}(x, y, z)/(\mathbf{m}(y)\, \mathbf{m}(x, z)))\, d\varphi(z)$, or $\int (\mathbf{m}(x, y, z)/\mathbf{m}(x, z))\, d\mathbf{m}(z) \leqslant \mathbf{m}(x, y)/\mathbf{m}(x)$, since $\mathbf{m}(z) \geqslant \varphi(z)$. Rewrite it: $\sum_z \mathbf{m}(z)\, \mathbf{m}(x, y, z)/\mathbf{m}(x, z) \leqslant \mathbf{m}(x, y)/\mathbf{m}(x)$ *or* $\sum_z \mathbf{m}(z)\, \mathbf{m}(x)\, \mathbf{m}(x, y, z)/\mathbf{m}(x, z) \leqslant \mathbf{m}(x, y)$. The latter is obvious since $\mathbf{m}(z)\, \mathbf{m}(x) \leqslant \mathbf{m}(x, z)$ and $\sum \mathbf{m}(x, y, z) \leqslant \mathbf{m}(x, y)$. Q.E.D.

1.3. *Time of Computation*

The speed of generating various r.e. sets is, as a rule, ignored in this work. Now we touch this question briefly. Let $t_{A(p)}$ mean the running time of $A(p)$. If A is the optimal algorithm from Proposition 1 and $R(p)$ is $(A(p), |p|)$ then $\forall x\, \exists p\colon R(p) = (x, \mathbf{K}(x))$. Exhaustive search for such p takes exponential time, even when $R(p)$ is fast. Let us give a fastest algorithm (storage modification machine) finding p.

Let $Kt_B(x/y) = \min\{(|p| + \ln t_{B(p,y)})\colon B(p, y) = x\}$, where p is a string without termination mark: the algorithm B receives, upon request, the digits of p in order until p is ended; in case of further requests B gets no reply and gives no output. $Kt_B(x) = Kt_B(x/0)$. Analogously to Proposition 1, an optimal B exists such that Kt_B is minimal within an additive constant, and Kt_B is denoted by Kt. There exists an algorithm $G(n, y)$ generating the list

$\{x: Kt(x/y) = n\}$ in time e^n; and within a constant factor, Kt is a minimal function with this property. (The asymptotically minimal one is $kt(x) = \min\{Kt(a): x \in a \subset \mathbb{N}\}$). Let R be some function computable in polynomial time. A problem of finding $q \in R^{-1}(s)$ (when it exists) is called a search problem. An NP-problem is to find whether (short) such a q exists. W.l.o.g., t_R can be made linear by "padding" q with zeros. Searching through all q in the order of increasing $Kt(q/s)$ (**rather than** $|q|$) gives a fastest (within a constant factor) algorithm for solving any search problem (see Levin, 1973a; related ideas were also expressed by L. Adleman).

Functions like Kt are of a particular interest for the case of randomized algorithms. For $f: \mathbb{N} \to \bar{\mathfrak{R}}^+$ let $c(f)$ be the expected value of $1/(t_{B(\alpha)} + f(B(\alpha)))$, where α is the random variable and, like above, B is the optimal algorithm. Let $C(f) = \lfloor -\ln c(f) \rfloor$. For $F \subset \mathbb{N}$, $C(F)$ means $C(f)$, where $f(x) = 0$, if $x \in F$, else $f(x) = \infty$. The above algorithm $G(n, y)$, generates numbers $x \in F \subset \mathbb{N}$ from random $y = \alpha$ in time $< e^n$ with probability $p > 1 - e^{-k}$, where $\ln k$ equals asymptotically $n - C(F)$. For any other algorithm, $p < e^{n - C(F)}$ (otherwise $C(f)$ could be improved). Thus $C(F)$ determines the time needed to "hit" F. A function f, with range other than just $\{\infty, 0\}$, can be interpreted as a "price" (e.g., time) needed to establish $x \in F = f^{-1}(\mathfrak{R}^+)$.

Everything is analogous for $C(f/y) = -\lfloor \ln \int d\alpha/(t_{B(\alpha, y)} + f(B(\alpha, y), y)) \rfloor$. C. Bennett proposed an interesting electrical interpretation of $C(F/y)$. Let us take the circuit of parallelly connected wires, each of which corresponds to a particular finite α such that $B(\alpha, y) \in F$. The length and the probability of a computation gives the length and the section of the wire. Then $C(F/y)$ is the logarithm of the resistance of the circuit.

A number of search (NP) problems are known, which are easy for probabilistic algorithms, but seem hard for deterministic ones; e.g., constructing "incompressible" words x, of high $K_c(x)$, where c is a constant and $K_c(x)$ (computable in polynomial time) is the minimal length of p, with $Kt(x/p) + Kt(p/x) < c \log |x|$. The complexity of a search problem R for probabilistic algorithms is characterized by $C(f_s/s)$, where $f_s(q) = t_{R(q)}$ if $R(q) = s$, and $f_s(q) = \infty$ otherwise. The relationship of this complexity with $|s|$ is a "randomized" version of the P = NP problem. But its relationship with the "complexity of obtaining s" looks even more interesting. More accurately, how does $C(\{s: \infty > C(f_s/s) > n\})$ grow with n, polynomially or logarithmically? Short s may exist for which it is very difficult to find $q \in R^{-1}(s)$, but to find such s may be even more difficult.

2. *Continuous Case*

2.1. *Universal Semimeasure*

Now let us extend **m** to the case of Ω.

PROPOSITION 3. *There exists a largest within a constant factor (universal) r.e. semimeasure*: $\exists \mathbf{M}\, \forall \mu\, \exists c\, \forall f > 0 : \mu(f) \leqslant c \cdot \mathbf{M}(f)$.

Proof. Let $(H, <)$ be an ordered set with a monotone operation $a: H^2 \to H$ ("averaging") and a family $Y: \mathbb{N} \to H$, containing the smallest element 0, closed under a and such that "$<, a$" are r.e. on the indexes. Let the supremums of all directed sets $Y(A)$ with r.e. A exist and be called the r.e. elements. Then $(H, Y, <, a)$ will be called a *numbered convex body*.

Note 2. Any numbered convex body $(H, Y, <, a)$ has a universal r.e. element, i.e., largest in any weaker than "$<$" order "$\lessdot$" such that $x \lessdot a(x, y) \gtrdot y$.

This element is the supremum of $\bigcup a_k(0)$, where $a_0(x) = \{x\}$, $a_{k+1}(x) = \bigcup \{a_k(a(r, x)) : r \in t_k\}$, and t_k is the kth enumerable directed subfamily of Y.

Proposition 3 is a special case, where $H = \mathscr{K}$, $a(\mu, \varphi) = (\mu + \varphi)/2$ and Y is the family of "elementary" semimeasures in a natural numeration. An elementary semimeasure is the minimal one satisfying a finite set of inequalities $\mu(f) > r$, where $f \in \mathscr{F}'$, $r \in \mathbb{Q}$. Q.E.D.

This $\mathbf{M}$ is the central technical concept of the work. Being largest, it determines the broadest class of sets $A \subset \Omega$ of positive probability. In statistics one tries, given α, to get a probability distribution μ with respect to which α may be reasonably considered "random"; i.e., $\mu(A) \gg 0$ for "standard" properties A satisfied by α. But this assertion is the weakest with $\mu = \mathbf{M}$. So $\mathbf{M}$ can be taken *a priori*, before studying what the properties of α really are. In other words. If α occurs randomly with probability distribution μ then it has properties A, for which $\mu(\neg A) = 0$. The class of such properties A is narrowest for $\mu = \mathbf{M}$, and they can be predicted *a priori* before finding out what μ really is. This justifies using $\mathbf{M}$ as *a priori* probability.

The distribution $\mathbf{M}$ is suitable for other applications, as *a priori* probability (e.g., for inductive inference in accordance with the ideas of Solomonoff, 1964), but these questions are not considered here. Let us note that $\mu(\alpha_n) \simeq \mathbf{M}(\alpha_n)$ for any r.e. semimeasure μ and μ-almost all α. This property of α can be used (see Proposition 5) as a definition of the concept of a μ-random sequence.

2.2. *Randomness and Information*; *Conservation Laws*

Now the second half of the conventions is very essential. Let us extend to Ω the concept of randomness tests considered in Subsection 1.1. For any set A of μ-measure 0 there is a lower semicontinuous function $t \in \mathscr{T}^+$, with average value $\mu(t) \leqslant 1$ and $t(A) = \{\infty\}$. Only for recursive μ, r.e. tests t are natural to consider. For a general case let $t^\rho \in \mathscr{T}^+$ be the function whose subgraph is enumerated by $\rho \in \Omega$. In Definition 2 we will average this over all ρ generated "arbitrarily," i.e., randomly with universal distribution $\mathbf{M}$. We

eliminate the case of $\mu(t^p) > 1$ defining $t^p_\mu = t^p$ if $\mu(t^p) \leqslant 1$, and $t^p_\mu = 0$ otherwise. To deal with random transformations which may turn an individual sequence α into a measure φ, we will extend tests $t^p_\mu(\alpha)$ to distributions as $\varphi(t^p_\mu)$. Let $\binom{\varphi}{\mu} \in \bar{\mathfrak{R}}^\Omega$ maps ρ to $\varphi(t^p_\mu)$.

DEFINITION 2. $\mathbf{D}(\varphi/\mu) = \lfloor \ln \mathbf{M}\binom{\varphi}{\mu} \rfloor$ is the deficiency of randomness of φ with respect to μ.

Most important is the special case $\mathbf{D}(\alpha/\mu)$ when φ is concentrated in a single point α. It is a generalization of $\mathbf{d}(x/\mu)$ from Note 1, since its "average" is $\mathbf{D}(\mu/\mu) \leqslant 0$. For motivations of the notion of randomness one may look in Sections 3 and 1.1. Now the central fact is

THEOREM 1 (Randomness Conservation). $\mathbf{D}(A(\mu_1)/A(\mu_2)) \prec \mathbf{D}(\mu_1/\mu_2)$, *where* $A \in \mathscr{S}$ *is r.e.,* $\mu_1, \mu_2 \in \mathscr{M}$.

Proof. Let $A'(\rho)$ be a sequence, enumerating the subgraph of $A(t^p) \in \mathscr{T}^+$. Then $\exp \mathbf{D}(A\varphi/A\mu) = \mathbf{M}\binom{A\varphi}{A\mu} = \mathbf{M}(A'\binom{\varphi}{\mu}) = (\mathbf{M} \circ A')\binom{\varphi}{\mu} \leqslant \mathbf{M}\binom{\varphi}{\mu}$, since $\mathbf{M}$ is universal. Q.E.D.

So, $\mathbf{D}(\alpha/\mu)$ is invariant (within a constant) with respect to all r.e. operators preserving μ.

DEFINITION 3. $\mathbf{I}(\alpha : \beta) = \mathbf{D}((\alpha, \beta)/\mathbf{M}^2)$ is called the amount of information in α about β or the deficiency of their independence. Here α, β may be sequences or semimeasures.

COROLLARY (A Generalization of Proposition 2). *Suppose* $\alpha, \beta \in \Omega$, $A \in \mathscr{S}$. *Then* $\mathbf{I}(A(\alpha) : \beta) \prec \mathbf{I}(\alpha : \beta)$.

The proof follows by noting that $A(\mathbf{M}) \leqslant \mathbf{M}$ and $\mathbf{D}$ is monotone.

This justifies the Independence Postulate, from the Introduction, since one can usually "explain" known physical processes reducing them to simpler ones in combination with recursive and random transformations (considered in the above corollary). The Universe, on the whole, is also assumed to evolve according to the (recursive) equations of physics from a state of random movement of hot plasma (additional randomness appears in the observation processes). Of course, not being a mathematical assertion (the physical world is not chosen mathematically), the Independence Postulate (like, e.g., Church's thesis) cannot be proven.

2.3. *Complete Sequences*

Any r.e. measure is recursive, i.e., computable with any accuracy, in contrast to semimeasure $\mathbf{M}$ for which any r.e. lower bound of $(\max f(x))/\mathbf{M}(f)$ is bounded by a constant. But it may be known about some

$\alpha \in \Omega$ that on its segments $\mathbf{M}$ agrees with some r.e. *measure* μ within a constant factor. Then, computing μ gives $\mathbf{M}(\alpha_n)$ within a constant. Such α is called *complete*, denoted $\alpha \in C$ or $C(\alpha) \Leftrightarrow \exists \mu \sup(\mathbf{M}(\alpha_n)/\mu(\alpha_n)) < \infty$. Its segments contain all the information needed to compute their complexity $-\lfloor \ln \mathbf{M}(\alpha_n) \rfloor$. According to Proposition 4, C is very wide. By virtue of its item 2, any sequence α satisfying the Independence Postulate has a completion $(\alpha, \tau) \in C$, satisfying this postulate as well.

PROPOSITION 4. (1) *For any recursive measure μ and total recursive operator A, $\mu(C) = \mu(\Omega)$ and $A(C) \subset C$.*

(2) *Let β be a sequence to which a universal r.e. set is Turing reducible and α be independent of β. Then $\tau \in \mathbb{N}^{\mathbb{N}}$ exists such that (α, τ) is complete and independent of β.*

This α comes from (α, τ) by a partial recursive (but not total) projection operator. So incompletable partial operators can lead out of C. In Section 4 an axiom is used which means intuitively that every "physical" sequence is a projection of a complete one.

Proof. Let $r < 1$, $\delta_\mu(\alpha) = \ln(1-r) + r \ln \sup(\mathbf{M}(\alpha_n)/\mu(\alpha_n)) < \infty$ and $\mu'(x) = \mu\{\alpha: A(\alpha) \supset x\}$. Then μ' is also an r.e. measure, and $\delta_{\mu'}(A(\alpha))$ is a μ-test. Then by virtue of Proposition 5, $\delta_{\mu'}(A(\alpha)) < \infty$ and $A(\alpha) \in C$. Obviously $\mu(C) = 1$ because $\delta_\mu(\alpha)$ is a randomness test.

It remains to prove (2). In Subsection 3.2 of (Zvonkin and Levin, 1970) it is shown that $\mathbf{M}$ (like any other r.e. semimeasure) is generated by a partial recursive operator A from a recursive measure μ: $\mathbf{M}(x) = \mu\{\alpha: A(\alpha) \supset x\}$. Let $A'(\alpha) = (A(\alpha), \tau_{A(\alpha)})$, where $\tau_{A(\alpha)}$ is the sequence of values of the time of computation of terms of $A(\alpha)$. The operator A' is total and, hence, $\mu'(x) = \mu\{\alpha: A'(\alpha) \supset x\}$ is a recursive measure. $\mathbf{M}$ is generated from μ' by the projector $(\alpha, \tau) \to \alpha$. And $\mu'\{(\alpha, \tau): \mathbf{I}((\alpha, \tau): \beta) = \infty\} = 0$, by Note 3. Also $\mu'(\Omega - C) = 0$. Therefore, $\bar{\mathbf{M}}\{\alpha: \forall \tau \in \Omega((\alpha, \tau) \notin C) \vee \mathbf{I}((\alpha, \tau): \beta) = \infty)\} = 0$. By Note 3, for any set A such that $\bar{\mathbf{M}}(A) = 0$, a sequence β' exists on which all elements of A depend. The same is true for any sequence to which β' is reducible. Using reducibility to β of the universal r.e. set, one can routinely check that the necessary β' is computable with respect to β. Thus β depends on all sequences α not completable for a complete, independent of β sequence (α, τ). Q.E.D.

II. Applications

3. *Foundations of Probability Theory*

3.1. *Foundational Difficulties (A Historical Digression)*

Hilbert's sixth problem suggests "**To treat** in the same manner" (as geometry), "**by means of axioms, those physical sciences, in which mathematics plays an important part; in the first rank are the Theory of Probabilities** and mechanics." (see Hilbert, 1902). The probability theory is considered there as a physical science. But its physical nature was almost forgotten after the analytical explosion followed Kolmogorov's book (Kolmogorov, 1933), where probability theory obtained a powerful mathematical foundation. And some difficulties in the relation of this mathematical apparatus to the probabilistic natural phenomena were left aside. Kolmogorov noted this in the foreword to the second Russian edition (1974) of the book, where he refers to (Kolmogorov, 1965; Zvonkin and Levin, 1970) for his new approach. The well-known previous attempts to overcome these difficulties by von Mises and Church turned out to be imperfect (Ville, 1939).

The difficulties lie in the gap between intuitive probabilistic ideas and those methods which are justifiable theoretically. Probabilistic considerations start with an assumption that a sequence x is generated randomly with probability distribution μ. This μ is discovered or hypothesized, e.g., by analogy with other processes and statistical data about them, considerations of symmetry, etc. Then, according to the naive ideas, those properties of x are indicated as probabilistic laws whose μ-probability is 1 (approximately, in the finite case); e.g., the law of large numbers plays an important role when $x = x_1,..., x_n$ are independent and identically distributed trials (i.e., $\mu(x_1,..., x_n) = \mu'(x_1)\,\mu'(x_2)\cdots\mu'(x_n)$). For each set B it predicts $x \in LLN$, which means that the frequency of i: $x_i \in B$ is close to the probability $\mu'(B)$ (and $\mu(LLN) \approx 1$). In general, x is predicted to have the properties whose probabilities are close to 1.

The problem is that *jointly the properties of probability* 1 *have probability* 0! One cannot guarantee all of them simultaneously, but should choose one or a few. Thus, the outcome should not be expected to withstand statistical tests chosen afterwards. So classical theory provides no rigorous basis to doubt the honesty of the lottery director whose son won the main prize in ten consecutive years, if this is discovered "post factum"! One cannot criticize an election when the share of votes for the ruling party in a series of consecutive years formed a sequence $0.99k_i$, even if k_i turn out to be the decimal digits of π! Of course, one can select few "standard laws" and presume them always to be chosen. *However, the classical probability theory*

has no ideas for selection of such standard laws. Besides, this would not justify applying probability theory to events preceding such a standardization (e.g., in cosmology, history, geology, etc.).

Kolmogorov's idea for solving this paradox is to select those properties of probability close to 1, which are "simply expressible." The objects not satisfying such a property form a simple set of small measure and correspondingly small cardinality. Then any such object is simple itself, being specifiable by its number (smaller than the cardinality of the set) with the simple description of the set. This allows substitution of many simple properties by a single one, "not to be a simple object." Kolmogorov's algorithmic information theory was a surprising discovery which provided a rigorous basis for the obscure notion of simplicity. In the infinite case the corresponding property is "to be random with respect to distribution μ." Then only this property is expected from the objects occurring randomly with distribution μ. This property is of μ-measure 1 and implies all other "good" properties of μ-measure 1. Attempts to introduce such universal concepts of randomness were undertaken previously (see von Mises and Church) for Bernoulli distributions. However, it was found (Ville, 1939) that even such standard properties as the law of iterated logarithm do not follow from their concept of being a "collective."

3.2. *The Laws of Randomness and Independence*

Any set A is of μ-measure 0 iff there is a lower semicontinuous function $t \in \mathcal{T}^+$, with average value $\mu(t) \leqslant 1$ and $t(A) = \{\infty\}$. For a typical result α of a μ-distributed random process, $t(\alpha)$ should not exceed by much the average, and the probability of large deviations is small. This justifies the following modification of a definition from (Martin-Lof, 1966).

DEFINITION 4. A randomness test with respect to a recursive measure μ (or a μ-test) is a function $\delta(\alpha) = \lfloor \ln t(\alpha) \rfloor$, where $t \in \mathcal{T}^+$ is r.e. and $\mu(t) \leqslant 1$.

Definition 4 is a formalization of the concept of a "good" law of probability theory. The degree $\delta(\alpha)$ of deviation from such a law is absolute when α fails the test, i.e., $\delta(\alpha) = \infty$, the probability of which is 0. The deviations can be effectively discovered since δ is r.e. The logarithmic scale is chosen for convenience.

PROPOSITION 5. *The following properties of $\alpha \in \Omega$ are equivalent for any recursive measure μ and true for $\mu = \mathbf{M}$:*

(1) $\mathbf{D}(\alpha/\mu) < \infty$. (*see Definition* 2.)

(2) $\sup(\mathbf{M}(\alpha_n)/\mu(\alpha_n)) < \infty$.

(3) *For any randomness test δ: $\delta(\alpha) < \infty$.*

Item (2) means that μ on segments of α is not much smaller than the *a priori* probability $\mathbf{M}$ (i.e., the assumption that α has occurred randomly with probability distribution μ is at least as consistent with reality as the *a priori* idea about it occurring with distribution $\mathbf{M}$). This property is of μ-probability 1 and implies all other effective probabilistic laws.

Proof. For any recursive measure μ and $r < 1$, it is easy to see that $r \ln t + \ln(1-r)$ is a μ-test itself, where $t = \sup\{\mathbf{M}(\alpha_n)/\mu(\alpha_n)\}$ and thus (3) $\Rightarrow$ (2). For any semimeasure μ, (1) $\Rightarrow$ (3) since the sequence ρ enumerating the subgraph of $\exp\delta$ is r.e. (thus $\mathbf{M}(\{\rho\}) > 0$) and $\mu(\exp\delta) \leqslant 1$. For $\mu = \mathbf{M}$, (2) is obvious and it remains to prove (2) $\Rightarrow$ (1) for any r.e. semimeasure μ. Let μ'_ρ be an r.e. family of r.e., with respect to ρ, normalized semimeasures, such that $\mu'_\rho = t^\rho \cdot \mu$ if $\mu(t^\rho) \leqslant 1$. Let $\tau^f_\mu(\rho) = \mu'_\rho(f)$. Then $\mu''(f) = \mathbf{M}(\tau^f_\mu)$ is obviously an r.e. semimeasure and $\mathbf{M}(f) \geqslant \mu''(f) \geqslant \exp\mathbf{D}((\mu \cdot f)/\mu)$. Let $\varphi_x(f) = \min\{f(\alpha): x \subset \alpha\}$. Let $g_x(\alpha) = 1$ if $x \subset \alpha$, and 0 otherwise. Since $\mu \cdot g_x \geqslant \mu(x) \cdot \varphi_x$, we have $\mathbf{M}(x)/\mu(x) \geqslant \exp\mathbf{D}(\varphi_x/\mu)$ and $\mathbf{D}(\alpha/\mu) = \sup\{\mathbf{D}(\varphi_x/\mu): x \subset \alpha\} < \sup\{\mathbf{M}(\alpha_n)/\mu(\alpha_n)\}$. Q.E.D.

Now we are ready to formulate the first of the two distinguished probabilistic laws.

THE LAW OF RANDOMNESS. Let $\alpha \in \Omega$ be taken randomly with a probability distribution μ. Then $\mathbf{D}(\alpha/\mu) \ll \infty$.

This law was shown to imply all r.e. probabilistic laws. What about the other ones? This is interesting for clarifying the relation between the algorithmic and classical approaches to probability theory. Let us give an important example of nonrecursive laws.

THE LAW OF INDEPENDENCE. Let $\beta \in \Omega$ be chosen. A random $\alpha \in \Omega$ must be independent of β, i.e., $\mathbf{I}(\alpha : \beta) \ll \infty$.

Note 3. The assertions: "$\bar{\mu}(A) = 0$ for all r.e. semimeasures μ" and "there exists β, such that $\mathbf{I}(\alpha : \beta) = \infty$ for all $\alpha \in A$" are equivalent for any $A \subset \Omega$, as it will follow from Lemma 3.

3.3. *Covering the Classical Formulation of Probability Theory*

The Law of Independence (as well as of Randomness) is violated only with probability 0, and thus it is a law of probability theory in the customary "classical" sense. This law varies only with the parameter β, and in its formulation ($\mathbf{I}(\alpha : \beta) < \infty$) the probability μ is not mentioned at all.

The Independence Postulate in Subsection 0.1 extends this law from the usual random processes to any physically realizable ones. This suggests bringing this law outside the bounds of probability theory and considering other probabilistic laws only for sequences which satisfy the Law of Indepen-

dence. This turns out reducing any other probabilistic law (recursive or not) to the Law of Randomness. Let $\mathbf{I}_\beta = \{\alpha: \mathbf{I}(\alpha:\beta) = \infty\}$, $\mathbf{D}_\mu = \{\alpha: \sup(\mathbf{M}(\alpha_n)/\mu(\alpha_n)) = \infty\}$ $(= \{\alpha: \mathbf{D}(\alpha/\mu) = \infty\}$ for an r.e. $\mu)$.

THEOREM 2. *Let $A \subset \Omega$ and $\mu(A) = 0$, then such β exists that $A \subset \mathbf{D}_\mu \cup \mathbf{I}_\beta$, i.e., any probabilistic law follows from the Laws of Randomness and Independence.*

Proof. Since $\mu(A) = 0$ and $\alpha \notin \mathbf{D}_\mu \Rightarrow \mathbf{M}(\alpha_n) \simeq \mu(\alpha_n)$, we have $\bar{\mathbf{M}}(A - \mathbf{D}_\mu) = 0$. Then there exists $\tau \in \mathscr{T}^+$ such that $\bar{\mathbf{M}}(\tau) < \infty$ and $\tau(A') = \{\infty\}$, and Theorem 2 follows from

LEMMA 3. *For each borel $\tau \in (\bar{\mathfrak{R}}^+)^\Omega$: $\bar{\mathbf{M}}(\tau) < \infty$ iff $\beta \in \Omega$ exists such that $\mathbf{I}(\alpha:\beta) > \ln \tau(\alpha)$.*

"If" follows from Theorem 1 (for $A: \beta \to \mathbf{M} \otimes \beta$) and item 1 of Proposition 5, since for any β:

$$\ln \bar{\mu}(\exp \mathbf{I}(\alpha:\beta)) \leqslant \mathbf{D}(\bar{\mu} \otimes \beta/\mathbf{M}^2) \leqslant \mathbf{D}(\mathbf{M}(\beta/\mathbf{M}^2) \leqslant \mathbf{D}(\beta/\mathbf{M}) < \infty.$$

On the other hand, $\mathbf{M}(\tau) < \infty$, then $\beta \in (\mathbb{Q}^+)^{\mathbb{N}}$ exists such that $\mathbf{M}(f) > 1 \Rightarrow \beta(f^*) = 0$ (for a natural effective enumeration $f \to f^*$ of $\mathscr{T}^+$) and $\sigma = \sum \beta(i) < \infty$, $\tau < \sum f \cdot \beta(f^*)$. Let L be the uniform measure on $[0, \sigma]$ and, $s(\rho, \beta)$ be the largest integer for which $\sum\{\beta(i): i < s(\rho,\beta)\} < \rho \in [0, \sigma]$. Let $A(\rho, f, k)$ be a sequence enumerating the subgraph of θ, where $\theta(\alpha, \beta) = e^k f(\alpha)$, if $f^* = s(\rho, \beta)$, otherwise $\theta(\alpha, \beta) = 0$. Obviously $\mathbf{M}^2(\theta) \leqslant \mathbf{M}(f) \cdot e^k \cdot \mathbf{m}(f^*/\rho)$ and thus, for some constant c, if $\mathbf{K}(f^*/\rho) \geqslant k + c$, $\mathbf{M}(f) \leqslant 1$, then $\binom{\alpha,\beta}{M^2}(A(\rho, f, k)) = \theta(\alpha, \beta)$. As in Lemma 1, $\mathbf{m}((f^*, \mathbf{K}(f^*/\rho))/\rho) \simeq e^{-\mathbf{K}(f^*/\rho)}$ and then: $\exp \mathbf{I}(\alpha, \beta) = \mathbf{M}\binom{\alpha,\beta}{M^2} \geqslant \sum f(\alpha) \cdot L\{\rho: s(\rho, \beta) = f^*\}$. And $\mathbf{I}(\alpha, \beta) > \ln \sum(f(\alpha) \cdot \beta(f^*)) = \tau(\alpha)$, since $L\{\rho: s(\rho, \beta) = f^*\} = \beta(f^*)$. Q.E.D.

4. *Intuitionistic Mathematics*

4.1. *A Digression*

The second-order theories (permitting quantification over functions or sequences) are much more complicated logically than the first-order ones. Some mathematicians (like intuitionists) considered these complications dangerous in terms of possible paradoxes. In particular, they assume sequences to be formed by sequential "free choices" thus resulting from physical (or mental) events rather than from logical definitions. Therefore, applicability of usual logical operations to them is not *a priori* obvious when these operations have no physical analogies. For example, classical universal quantification assumes the unrealistic ability to scan all conceivable

sequences. The hope was to get a less suspicious mathematics restricting the logical procedures and postulates to only those closer related with physical intuition.

However, obscurity of notions like "free choice" and of the physical intuition makes it difficult to choose formal principles reflecting adequately the nature of physically generated sequences. The result is a great variety of intuitionistic principles and theories that strengthen, weaken, or contradict each other. These theories are often so strong that the relation of their principles to physical intuition stops being obvious. In fact, they are often equiconsistent to the corresponding classical theories. On the other hand, they are too weak, leaving independent many other principles of intuitionistic reasoning. This provides room for creativity in extending these theories, but eliminates the hope to get a "canonical" theory with some kind of completeness.

To deal with these difficulties an axiom schema is introduced below formalizing the Independence Postulate from the Introduction. The intuitionistic *second*-order arithmetic obtained is equiconsistent to (is a *conservative* extension of) the classical *first*-order arithmetic, formulated without disjunction and existential quantifier. So its new principles are "purely logical," i.e., imply no new facts of classical number theory. On the other hand, it is *complete*, i.e., is a *maximal* conservative extension. The Independence Postulate brings these "virtues" by excluding sequences with unbounded information about the validity of mathematical statements. It is natural to attribute the usual troubles of second-order theories to such fancy "logical" sequences which, in fact, cannot physically exist.

4.2. *The Preliminary Calculus A*

Our theory AI will be constructed in Section 4.3 by extending the basic calculus A, described below. The language of A is second-order arithmetic. It contains first-order arithmetic (see Kleene, 1967, Sect. 38), a countable list of second-order variables (denoting sequences or functions of natural numbers), terms $\alpha(t)$, and formulas $\forall\alpha F$ and $\exists\alpha F$ for all terms t, formulas F, and second-order variables α. A formula is called *absolute* if it does not contain $\exists$, $\vee$ and quantification over *second-order* variables. Absolute formulas have identical meaning and equivalent provability in intuitionistic and classical theories. Pairs (and tuples) of integers, sequences, and terms are defined the same way as in subsection Notation. Abbreviation $s = \alpha(n, \downarrow)$ mean $\exists k(\alpha(n, k) = s + 1 \;\&\; (\forall k' < k\ \alpha(n, k') = 0))$.

The postulates of A consist of the ones of first-order arithmetic (see Kleene, 1967, p. 387, List of Postulates, Schema 8 taken in the intuitionistic version 8′) and three second-order postulates:

Schema of Choice: $(\forall n(\neg A \Rightarrow \exists k\, B(k))) \Rightarrow \exists\alpha\, \forall n(\neg A \Rightarrow B(\alpha(n, \downarrow)))$ (4.2.1)

Markov Principle: $(\neg\forall n\, \alpha(n) = 0) \Rightarrow \exists n\, \alpha(n) \neq 0$ (4.2.2)

Axiom of Countability: $\exists\alpha\, \forall\beta\, \exists k\, \forall n\, \beta(n) = \alpha(k, n, \downarrow)$. (4.2.3)

The axioms of A are not new and need no detailed discussion. Still in any complete (in the sense of Theorem 3) theory these axioms must be provable or refutable or equivalent to some undecidable absolute statements of number theory. The last two possibilities seem to be less natural. It is known that (4.2.1–4.2.3) are inconsistent with the principle of continuity.

Of course, the calculus A is still too weak. Nevertheless,

PROPOSITION 6. *For any formula F an absolute P exists such that $A \vdash F \Leftrightarrow \forall\alpha\, \exists\beta\, P$.*

Proof. Axioms of A allow a construction analogous to Kleene's recursive realizability, using the universal sequence α from axiom (4.2.3). Namely, a formula $P_F(x, \alpha)$, meaning "a number x realizes a formula F with respect to a sequence α" can be defined as in Kleene's "Introduction to Metamathematics," Chapter 2, except that recursiveness of all functions used is considered with respect to α. Then any formula F is equivalent to the existence of a realization of F with respect to a universal α. It is easy (though bulky) to check that A contains all the axioms necessary for formalizing these arguments, i.e., the deduction of $F \Leftrightarrow \forall\beta\, \exists\alpha\, P_F(\alpha(0), (\alpha, \beta))$. Q.E.D.

4.3. *The Calculus AI; its Relative Consistency and Completeness*

Let $P(n)$ be an absolute formula with a single free variable n. A finite binary sequence p is *compatible* with P (denoted $p \subset P$), if $\forall n \leqslant l(p)$: $(p(n) = 0 \Leftrightarrow P(n))$. The abbreviation $\mathbf{I}(\alpha : P)$ means $\sup\{\mathbf{I}(\alpha : p) : p \subset P\}$. For a given P, the statement $\mathbf{I}(\alpha : P) \leqslant c$ can be easily expressed by an absolute formula with free variables α, c. This is used in the following axiom schema with a parameter P:

Independence Postulate: $\exists c\, \mathbf{I}(\alpha : P) \leqslant c$. (IP)

The property of completeness (defined in Subsection 2.3) is expressible by an absolute formula $C(\alpha)$. The following last axiom of AI asserts implementability of completion of sequences mentioned in Proposition 4: $\exists\tau\, C(\alpha, \tau)$. The double negations of this axiom and of (IP) would be sufficient for our purposes, but the chosen ones are simpler.

DEFINITION 5. A theory T is called *absolute* if for every closed formula F an absolute (see Subsection 4.2) formula P exists such that $T \vdash \neg F \Leftrightarrow P$.

Replacing (4.2.3) in A with *Church's thesis* (CT): $\exists k\, \forall n\colon \beta(n) = U(k, n, \downarrow)$ (where U is a universal recursive function) one gets the theory of recursive realizability of Kleene. It is a known example of an absolute theory. Our theory AI is not, of course, absolute, inasmuch as (CT) is independent of it and cannot be reduced to an absolute formula. Then to get an absolute theory one needs an axiom implying either (CT) or $\neg$(CT). It turns out that this is sufficient as well.

LEMMA 4. *For any closed formula F, four absolute formulas P_1, P_2, P_3, P_4 exist such that these statements are deducible in AI*:

$$\neg\neg(P_1 \vee P_2 \vee P_3 \vee P_4); \qquad P_1 \Rightarrow \neg\neg F; \qquad P_2 \Rightarrow \neg F;$$
$$P_3 \Rightarrow (\neg F \Leftrightarrow (CT)); \qquad P_4 \Rightarrow (\neg F \Leftrightarrow \neg(CT)).$$

Church's thesis (CT) is a very strong axiom. It excludes any nonrecursive sequences, e.g., random ones. The axiom $\neg(CT)$ is, inversely, very weak. But, unexpectedly, AI + $\neg(CT)$ *is also absolute*:

THEOREM 3. *The absolute closed theorems of $AI + \neg(CT)$ are the same as ones of the classical first-order arithmetic. No essential (i.e., containing new theorems of the form $\neg F$) extension of $AI + \neg(CT)$ has this property.*

Thus $AI + \neg(CT)$ is a maximal conservative extension of classical arithmetic and is, relatively in a sense, consistent and complete. The basic goal of introducing this theory was to study the effects of the axiom schema (IP).

Proof of Theorem 3. We need for each closed formula F to establish a corresponding absolute formula $\bar{F}$ such that: $(AI + \neg(CT)) \vdash \neg F \Leftrightarrow \bar{F}$, and if F is absolute itself, then $\neg F \Leftrightarrow \bar{F}$ is deducible in first-order arithmetic. Besides, we need to show that all deduction rules and axioms of $(AI + \neg(CT))$ will be converted into derivative deduction rules and theorems of first-order arithmetic. We shall indicate the transformation $\neg F$ into $\bar{F}$ and explain its meaning without writing out all routine formal deductions. Due to Proposition 6 one may restrict himself to formulas of the kind $F = \forall\alpha\, \exists\beta\, P(\alpha, \beta)$, where P is absolute. We say that F is *rejected* on $\gamma \in \Omega$ if for any recursive function $r\colon \mathbb{N} \to \mathbb{N}$ it is false that for any recursive operator $k\colon \Omega \to \Omega$, applicable to γ, $P(\alpha, \beta)$ holds, where $\alpha = k(\gamma)$, and $\beta = k'(\gamma)$, $k' = r(k)$. Let μ be a recursive continuous measure. We will see that the equivalence of $\neg F$ to the formula "F is rejected for μ-almost all γ" is deducible in $AI + \neg(CT)$.

The latter formula can be written in an absolute form and chosen as $\bar{F}$. The point is that the quantifier "for almost all γ" in contrast to the quantifier "for all γ" is expressible in the first-order language. Obviously the formula

"F is rejected on γ," being absolute, can be presented in the form of $\forall n_k \neg \forall n_{k-1} \neg \cdots \forall n_0 \neg R(\gamma, n_0, n_1, \ldots, n_k)$, where R is a recursive predicate, monotonic on each of the arguments n_i (up for the even i and down for the odd ones). Let us show by means of recursion, how the predicate $\mu\{\gamma: \forall n_{i-1} \neg \forall n_{i-2} \neg \cdots \forall n_0 \neg R(\gamma, n_0 \cdots n_k)\} \geqslant r$ is expressed by an absolute formula. For $i = 0$ it is trivial. Now let, at the given i, the predicate be expressed in the form of $S_i(r, n_k \cdots n_i)$. Then $\forall n_i \forall r' > (1 - r) \neg S_i(r', n_k \cdots n_i)$ can serve as $S_{i+1}(r, n_k \cdots n_{i+1})$. Thus, it remains to show that $\neg F$ is equivalent in $AI + \neg(CT)$ to the assertion "F is rejected for μ-almost all γ."

LEMMA 5. *Let μ and μ' be r.e. measures and μ be continuous. Then recursive deterministic reciprocal in the domains operators P and P' on Ω exist defined on μ' (resp. μ)-almost all nonrecursive sequences and such that $\mu' = P(\mu)$.*

The proof of this lemma follows from Theorem 3.1(b) in (Zvonkin and Levin, 1970). Since the property "F is rejected on γ" is invariant with respect to any recursive reversible transformation of γ, it is sufficient to prove the equivalence of $\neg F$ to "F is rejected for B-almost all γ," where B is the uniform measure on $\{0, 1\}^{\mathbb{N}}$. By virtue of the same invariance and 0-1 law (Kolmogorov, 1933), the set A of all γ, on which F is rejected, can be only of measure 0 or 1 with respect to B. Hence if R is the set of all recursive sequences, the measure of either $(A \cap \neg R)$ or $(\neg A \cap \neg R)$ equals 0 with respect to all other recursive μ as well. Then by virtue of Theorem 2, a sequence exists (and can easily be defined by an absolute formula), upon which all **complete** γ from this set depend. The axioms of $AI + \neg(CT)$ imply that any universal sequence (from axiom 4.2.3) is nonrecursive, equivalent to a complete one, and independent of sequences defined by absolute formulas. Therefore in the case $\mu(A) = 0$, F is not rejected on a universal γ and $\neg\neg F$ holds. In the opposite case $\neg F$ holds by analogous reasons. These reasonings can be easily transformed to formal proofs in $AI + \neg(CT)$. Each of the two cases gives implication in one of the directions between $\neg F$ and "F is rejected for μ-almost all γ." Q.E.D.

5. *Theory of Turing Degrees*

5.1. *Independence and Negligible Sets*

A natural field for application of algorithmic information theory is the theory of Turing degrees. One may interpret the recursive reducibility of α to β as β contains all (but a finite amount of) information about α. However, the informational concepts are subtler and more elegant than reducibility degrees. In particular, they are invariants applicable to finite objects as well

(Proposition 2 shows that $\mathbf{I}(x, y)$ is left invariant within a constant by any recursive reversible transformation of $\mathbb{N}$).

One of the new possibilities is the informational approach to the concept of independence in addition to the concept of reducibility. In terms of reducibility degrees one can also say that α and β are independent if any sequence reducible to both of them is trivial (recursive). But a simple example shows that this definition does not always agree with the intuition. Let α and γ be random 0, 1 sequences, of independent trials, with the probability of $\alpha_n = 0$ being $\frac{1}{2}$, and the probability of $\gamma_n = 0$ being 0.99. Let $\beta_n = \alpha_n \oplus \gamma_n$. Then, α and β are almost always such that only recursive sequences are reducible simultaneously to both of them, though 99% of the (random) contents of α and β coincide (so it is hard to consider them independent).

Many exotic types of Turing degrees are known, e.g., "minimal" degrees containing indivisible information (any part of the information of such a degree β, i.e., a degree $\alpha < \beta$, is equivalent to 0 or to β). The existence of such degrees is proven by diagonal methods. Such sequences cannot appear in any combination of random and recursive processes (see Rogers, 1967). One may hope that many complications of the theory of Turing degrees are caused by exotic examples of this kind and the theory of "realistic degrees" is simpler. We shall see that this is only partially so.

Let us use the concept of independence to define the notion of "negligible sets" of sequences and study properties of Turing degrees "within this negligibility." A set $A \subset \Omega$ is called *inaccessible*, if its complement is closed with respect to any recursive operator F (i.e., $\alpha \notin A \Rightarrow F(\alpha) \notin A$).

PROPOSITION 7. *The following properties of a set $A \subset \Omega$ are equivalent:*

(1) *A sequence $\alpha \in \Omega$ exists on which all $\beta \in A$ are dependent (i.e., $\exists \alpha\, \forall \beta \in A\colon \mathbf{I}(\alpha : \beta) = \infty$).*

(2) *A is a subset of an inaccessible set, whose any (or some continuous) r.e. measure is* 0.

(3) $\bar{\mathbf{M}}(A) = 0$.

Proof. (1) $\Leftrightarrow$ (3) follows from Note 3. It is obvious that $F(\mathbf{M})$, the image of $\mathbf{M}$ at an arbitrary recursive mapping $F\colon \Omega \to \Omega$, is an r.e. semimeasure and hence $F(\mathbf{M}) \leqslant \mathbf{M}$. Therefore, if $\bar{\mathbf{M}}(A) = 0$, then $\bigcup F^{-1}(A) = A_1 \supset A$ is inaccessible and $\bar{\mathbf{M}}(A_1) = 0$. This gives (3) $\Rightarrow$ (2). Lemma 5 implies equivalence of "some" and "any" in (2). Any r.e. semimeasure is the image of a recursive measure at a recursive mapping $\Omega \Rightarrow \Omega$ (see Zvonkin and Levin, 1970, Sect. 3.2). This gives (2) $\Rightarrow$ (3). Q.E.D.

The sets with any of these three properties are called *negligible* (this neglect is, of course, based on the belief in the Independence Postulate). Two

sets A and B are called *i-equivalent* if their symmetric difference is negligible. "A property of Turing degrees" means a Turing-invariant set $A \subset \Omega$. Studying them to within i-equivalence is simpler, since some properties of "exotic" degrees are excluded. The Boolean algebra of Borel sets of Turing degrees is denoted by K, and **L** is its factor algebra with respect to the i-equivalence.

5.2. *Types of Turing Degrees*

In Subsection 2.3 the concept of "sequence completeness" was considered. The set of incomplete sequences has a property very close to negligibility. Namely, (2) in Proposition 7 is obtained from (1) of Proposition 4 by omitting the word "total." Thus, incomplete sequences cannot arise in a process running in time bounded by a total recursive operator. Let us call *regular* a sequence that is Turing-equivalent to a complete one. It is natural to consider properties of the Turing degrees of regular sequences. It turns out that only four of them are not equivalent.

THEOREM 4. *Any Turing-invariant Borel set of regular sequences is i-equivalent to*:

(1) *the empty set,*

(2) *the set of recursive sequences,*

(3) *the set of all regular sequences, or*

(4) *the set of all regular nonrecursive sequences.*

Thus, the properties of a regular sequence (to within i-negligible sets) depend only on its recursiveness, and these sequences form the two most natural elements (atoms) of the algebra **L**.

Proof. As it follows from Lemma 5, any set A of nonrecursive sequences, invariant with respect to Turing equivalence, either is of measure 0 at any recursive measure μ, or (for any μ) contains μ-almost all nonrecursive sequences. Then, by virtue of Theorem 2, a β exists such that all the complete nonrecursive sequences either from A, or from the complement of A, respectively, depend on β. And A is i-equivalent to one of the four sets, mentioned in Theorem 4, since any invariant set contains either all recursive sequences, or none. Q.E.D.

Other Turing degree types consist of nonregular sequences. It is difficult even to prove that their union is not negligible. Nevertheless **L** contains (see V'yugin, 1982; Levin and V'yugin, 1977) an infinitely divisible element and a countable number of atoms. Only two of them (namely, (2) and (4) of Theorem 4) contain complete sequences.

ACKNOWLEDGMENTS

I am deeply grateful to A. R. Meyer for improvements to this text, for financial and moral support, and to P. Elias and G. E. Sacks who agreed to read the first draft of the work. My wife Larissa actually wrote this paper (before I spoiled it). Many colleagues had a hard time convincing me to omit my bravest discoveries in English. I am grateful to all of them and sorry that their mission was so difficult.

RECEIVED: January 7, 1984; ACCEPTED: July 5, 1984

REFERENCES

CHURCH, A. (1940), On the concept of random sequence, *Bull. Amer. Math. Soc.* **46**, 254–260.

CHAITIN, G. J. (1966, 1969), On the length of programs for computing finite binary sequences, I, II, *J. Assoc. Comput. Mach.* **13**, 547–569; **16**, 145–159.

CHAITIN, G. J. (1975), A theory of program-size formally identical to information theory, *J. Assoc. Comput. Mach.* **22**, 329–340.

GACS, P. (1974), On the symmetry of algorithmic information, *Soviet Math. Dokl.* **15**, 1477.

HILBERT, D. (1902), Mathematical problems, *Bull. Amer. Math. Soc.* **2**, No. 8, 437–479.

KLEENE, S. C., AND VESLEY, R. E. (1965), "The Foundations of Intuitionistic Mathematics," North-Holland, Amsterdam.

KLEENE, S. C. (1967), "Mathematical Logic," Wiley, New York.

KOLMOGOROV, A. N. (1933), "Grundbegriffe der Wahrscheinlichkeitrechung," Berlin; 2nd Russian ed. (1974), "Osnovnye Poniatija Teorii Verojatnostej," Nauka, Moskow.

KOLMOGOROV, A. N. (1965), Three approaches to the concept of *The Amount of Information*, *Problems Inform. Transmission* **1**, No. 1.

KOLMOGOROV, A. N. (1968), Talk resume, *Uspekhi Mat. Nauk* **2**, 201.

ZVONKIN, A. K., AND LEVIN, L. A. (1970), The complexity of finite objects and the algorithmic concepts of information and randomness, *Russian Math. Surveys* **25**, No. 6, 83–124.

LEVIN, L. A. (1973a), On the notion of a random sequence, *Soviet Math. Dokl.* **14**, No. 5, 1413.

LEVIN, L. A. (1973b), Universal sequential search problems, *Problems Inform. Transmission* **9**, No. 3, 265–266.

LEVIN, L. A. 1973c), On storage capacity for algorithms, *Soviet Math. Dokl.* **14**, No. 5, 1464–1466.

LEVIN, L. A. (1974), Laws of information conservation (non-growth) and aspects of the foundations of probability theory, *Problems Inform. Transmission* **10**, No. 3, 206–210.

LEVIN, L. A. (1976a), On the principle of conservation of information in intuitionistic mathematics, *Soviet Math. Dokl.* **17**, 601–605.

LEVIN, L. A. (1976b), Various measures of complexity for finite objects (axiomatic descriptions), *Soviet Math. Dokl.* **17**, No. 2, 522–526.

LEVIN, L. A. (1976c), Uniform tests of randomness, *Soviet Math. Dokl.* **17**, No. 2, 337–340.

LEVIN, L. A., AND V'YUGIN, V. V. (1977), "Invariant Properties of Informational Bulks," Lecture Notes on Computer Science No. 53, pp. 359–364, Springer, Berlin/New York.

MARKOV, A. A. (1964), On normal algorithms which compute boolean functions, *Soviet Math. Dokl.* **5**, 922–924.

MARTIN-LOF, P. (1966), The definition of random sequences, *Inform. and Control* **9**, 602–619.

MINSKY, M. L. (1962), Problems of formulation for artificial intelligence, *in* "Proc. Sympos. Appl. Math. No. 14," Amer. Math. Soc., Providence, R.I.

ROGERS, H. (1967), "Theory of Recursive Functions and Effective Computability," McGraw–Hill, New York.

SCHNORR, C. P. (1971), "Zufaelligkeit und Wahrscheinlichkeit," Lecture Notes in Math. No. 218, Springer, Berlin/New York.

SOLOMONOFF, R. J. (1964), A formal theory of inductive inference, *Inform. and Control* **7**, No. 1, 1–22.

VILLE, J. (1939), "Etude critique de la concept de Collectif," Gauthier–Villars, Paris.

VON MISES, R., AND GEIRINGER, H. (1964), "The Mathematical Theory of Probability and Statistics," Academic Press, New York.

V'YUGIN, V. V. (1982), The algebra of invariant properties of binary sequences, *Problems Inform. Transmission* **18**, No. 2, 147–161.