8. Random sequences, incompleteness and information

8.1. What is a random sequence?

> When a rule is extremely complex, that which conforms to it passes for random. (Leibniz)

Another mathematically significant development in the research around the phenomenon of incompleteness on which we consider it important to focus attention, is that which has led to highlighting the link between this concept and that of randomness and information. Grigory Chaitin, starting in the 1970s (see the bibliography), reformulated the incompleteness theorems within the framework of algorithmic theory of information, presenting his results as a "dramatic extension" of the phenomenon already highlighted by Gödel. He claims that high complexity is the alleged reason of the unprovability of infinitely many true sentences: a true statement that can be expressed in the language of a theory is unprovable because its information content is greater than that of the axioms of that theory itself; or in other words, in a formal system no number can be proved random unless its complexity is less than that of the formal system itself.

At the heart of Chaitin's work, therefore, is the concept of the complexity of an object and the measure of the difficulty of describing it. *The mathematical concept of randomness* is an attempt to give an idealized model of *randomness*, as the recursive functions do in the case of computability.

When a sequence of numbers is random? Some solutions are not completely satisfactory: Borel (1909) introduced the notion of 'normality' of a sequence of decimal digits, namely the property for which the frequency of each block of digits of length $k \geq 1$ in each finite initial segment of length m of that sequence approximates to 10^{-k} as m goes to the infinity (see Calude (1994) for a thorough analysis). More precisely let us define a *basis*, namely a number $b \geq 2$ and let a *digit* in base b an element of $\{0, 1, 2..., b-1\}$.

A *block* in base b is a finite sequence w of digit and if we denote $w[i, j]$ a *subblock* from i to j of w (where $1 \leq i \leq j \leq |w|$), then let $occ(w, u)$ the cardinality of the set $\{i|w[i, i+|u|-1] = u\}$, i.e. the number of occurrences of the string u in w. For instance, if $w = 2122113211$ and $u = 211$ and $occ(w, u)$ =cardinality of the set $\{3, 8\}$, namely 2.

The expansion of $r \in [0, 1]$ in base b is given by $(r)_b = \sum_{i=1}^{\infty} a_i b^{-i}$, where the a_i are digits of the basis. The real number r is called *normal* for the basis b, if for each block u:

$$\lim_{n\to\infty} \frac{occ((r)_b[1, n], u)}{n} = \frac{1}{b^{|u|}}$$

It is *absolutely normal*, if this holds for every basis. Some important facts in this regard are the following:

Duccio Pianigiani, University of Siena, Italy, duccio.pianigiani@unisi.it, 0000-0001-9441-7226

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Book References DOI 10.36253/979-12-215-0778-2.references

(a) (Borel 1909). Almost all real numbers are absolutely normal.

(b) (Turing 1938). There is a computable absolutely normal number.

(c) (Champernowne 1933) The number:

$$0,1234567891011121314151617181920212223 24....$$

is normal in base ten.

Champernowne's number, in particulart, shows that normality is not a fully satisfactory definition of 'randomness', because its development is easily predictable. Some other attempts of characterizations of random sequences link randomness to the stability of frequency. However 1010101010... has this feature, while not seeming entirely random. In particular, notice that there is a subsequence (the one composed of only places even) that does not obey this law.

Randomness is associated with the idea of the growth of disorder, i.e. the decreasing of information, but even this is sometimes misleading. Indeed the string 01101010000010011110011 although apparently random, consists of the first 23-digit of the binary expansion of $\sqrt{2}-1$.

Already in Kolmogorov (1965), among the known approaches to the problem of the quantitative definition of information, besides that in terms of algorithmic complexity due to the author himself, is mentioned the concept of entropy introduced in Shannon (1948). The general concept of entropy in thermodynamics, as is well known, was introduced by Clausius in the middle of the nineteenth century, and its success is linked (also beyond its original scope) to names such as Boltzmann, Gibbs, Shannon and Von Neumann. The properties of Shannon's concept of entropy and those of Kolmogorov, Solomonov and Chaitin's concept of complexity are in many respects similar: both constitute bit-based measures of information; in both the information conveyed by an object depends on the length of its description. Propositions formulated in the terms of one can be reformulated in the terms of the other, and Romashchenko's theorem (see Hammer, Romashchenko, Shen and Vereshchagin (2000)) establishes, in very general terms, that any linear inequality true for Kolmogorov's complexity is also true for Shannon's entropy, and vice versa.

Considering infinite binary sequences, four principal characterizations have been provided:

(a) in terms of stability of frequency (Von Mises, Wald, Church),

(b) in terms of incompressibility (Solomonov, Kolmogorov, Chaitin)

(c) in terms of typicity (Martin-Löf)

(d) in terms of unpredictability (the theory of martingale).

As for unpredictability, a sequence $\sigma \in 2^{\omega}$ is called random, if given its initial segment $\sigma \upharpoonright n$ its first n bits we cannot predict the next bit. For example, the outcome of an ideal coin is unpredictable in the sense that the knowledge of the first n-outcomes do not helps to predict the next. We will not deal here of martingale theory. Jean Ville invented martingales in the 1930s in order to improve Richard Von Mises' concept of a collective, and Claus-Peter Schnorr made martingales algorithmic in the 1970 and characterized Martin-Löf randomness in terms of martingales. Random sequences, as equivalently defined in 2,3,4, although chaotic, nevertheless may have a strong computational power: we will see a particular sequence that computes the Halting Problem. More surprising is the result that was originally obtained from Gács (1986) and Kučera (1985) who proved the following result.

Theorem 120. *Any sequence is Turing-reducible to a random sequence.*

Contrary to what one might think, not only is it not true that no useful information can be extracted from random sequences, but on the contrary it seems that many random sequences are able to " calculate everything ":

Any type of information that can be coded into an infinite binary sequence, no matter how structured that might be, can be obfuscated into an algorithmically random infinite binary sequence, from which it is effectively recoverable (Barmpalias and Lewis-Pye (2018)).

8.2. From Von Mises to Martin-Löf

We start from Von Mises' *Kollektive* to understand how we arrive at the notion of the Martin-Löf test. In addition to Borel, the first remarkable attempt to formalize the 'random sequence' notion was made by Richard Von Mises in the 20s of '900 in the context of his investigation into the foundations of probability (see Zaffora Blando (2024), Van Lambalgen I (1987) and Van Lambalgen II (1987) among the extensive bibliography of this author for a detailed discussion). The problem of giving an axiomatization of the concept of probability is the sixth of the well-known list of twenty-three problems compiled by Hilbert in 1900. In 1919, Richard Von Mises, a member of the Vienna Circle and probabilist of the frequentist school, presented its axiomatization based on the concept of *Kollektiv*, i.e. on a certain type of abstract characterization of an infinite sequence of independent trials, that meets certain global and local regularities, about the extraction of infinite subsequences. We limit ourselves in this short exposition to the case where the results of a possible experiment are 0 or 1. We denote by $2^{<\omega}$ the set of finite binary strings and with 2^{ω} the set of *infinite* binary strings (Cantor space).

A binary sequnce $\sigma \in 2^{\omega}$ is a *Kollektiv* if and only if it satisfies the following conditions:

(a) if $S_n = \sum_{i \leq n} \sigma(i)$, then $\lim_{n\to\infty} \frac{S_n}{n}$ exists and is a real p in $[0, 1]$ (considering a uniform distribution generated by the toss of a fair coin then $p = \frac{1}{2}$).

(b) If R : is a "selection rule" extracting a subsequence $R(\sigma)$ of σ, then this subsequence satisfies point 1. with the same probability:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i \leq n} (R(\sigma)_i) = p$$

namely, extracting infinite subsequences, the stability of the frequency is preserved (impossibility of a *gambling system*).

The axiom 1. guarantees that the limiting relative frequencies of 1 and 0 exist. The meaning of point 2. is that a sequence should count as random if all infinite subsequences have the same frequency of 0's as 1's in the limit. A sequence such as 101010... clearly should not count as random in this sense. In other words, we want to guarantee that the sequence "cannot be gamed" by identifying a scheme that can be exploited to devise a successful gambling system. These 'collectives' are the formal counterpart of the notion of 'random sequences'.

However, it was necessary to better define the notion of "selection rule of a place". If we admit arbitrary selection rules, there is no collective *strictu sensu* (the so-called "Kamke's argument": take $R = \langle n | \sigma(n) = 1 \rangle$, extracting the sequence 1111...). It must nevertheless be emphasised that this objection is its strongly non-constructive character and nevertheless highlights the need to better denote the notion of a place-selection rule. Kamke (1932) therefore proposed to restrict the sequences which may represent admissible place selections to those which are lawlike; Wald (1937), accepting this criticism, demonstrated that a collective exists (thus the very notion of a collective is consistent), as long as it is restricted to a countable set of rules. Among the attempts to specify the notion of the 'collective' Church (1940) proposal to take as admissible place selections the computable functions $f : 2^{<\omega} \to \{0, 1\}$ actually constitutes the first definition of algorithmic randomness. However the most sharp criticism of Von Mises'

definition came from Ville (1939): he showed that in a sense the collective are not random enough, as there are some collectives with limit frequency of '1 'equal to 1/2, but $\forall n(\frac{S_n}{n} \geq \frac{1}{2})$. This collective violates a fundamental probabilistic law known as *Law of iterated logarithm*. This law implies that for almost all sequences frequency reaches its limit exhibiting large oscillations above and below the origin, while the Ville sequence reaches its limit "from above". Ville actually proved the following theorem.

Theorem 121. (Ville 1939) *For each countable set of rules of selection of a place, there is a collective that violates the law of the iterated logarithm.*

He suggested that a limit of the collectives was to obey a single law of chance, the law of large numbers, *and not to all these laws*, as we would expect from a random sequence. A random sequence has to be 'typical', in the sense of 'noexceptional ': we do not consider as 'typical' a sequence that differs from others for some peculiarities, such as contain a finite number of '1 '; a binary sequence is therefore considered *typical* , if it is not rare, it has no specific feature.

Martin-Löf proposed a satisfactory solution to the objections addressed to the notion of *Kollektiv* as a formal account of the concept of 'random sequence', establishing a relationship between stochastic computable properties and particular subsets of zero-measure of 2^ω: are considered *random* the sequences that do not fulfils property *effectively rare*. We distinguish between *typical* sequences and *special* sequences: *typical* means not exceptional, ordinary, as a synonym for *random*. More formally, a sequence $\sigma \in 2^\omega$ is *typical* if and only if every set of sequences each containing only a small number of sequences, does not contain σ (the statement should not be taken literally, since $\sigma \in \{\sigma\}$ and therefore each individual is not typical). It is not a typical sequence, the sequence that is distinguished by a specific property, for example, to possess a finite number of 1; the probability of obtaining a non-typical binary sequence with the launch of a correct coin is 0. Conversely, the probability of getting a typical sequence is 1. A typical sequence has all the opposite properties with respect to a non-typical sequence. It meets all the " laws of chance ", that is, the properties of which it is proved that they have the value 1.

Let therefore $\sigma \in 2^{<\omega}$, where $2^{<\omega}$ is the set of *finite* binary strings, and let $[\sigma]$ the set of *infinite* binary sequences that begin with σ, namely that extend it, also called cylinders (the basic open of the product topology on 2^ω). We can read $[\sigma]$ as: "the first $|\sigma|$ flips of a coin gave σ". Mostly the *uniform measure* is used an it is given by:

$$\mu([\sigma]) = 2^{-|\sigma|}$$

(where $|\sigma|$ denotes the length of σ) i.e. the probability that a $\tau \in 2^\omega$ obtained flipping a fair coin will be in $[\sigma]$, or the probability that a $\tau \in 2^\omega$ begins with σ. The measure μ, defined on cilynders, can be extended to all sets $A \subseteq 2^\omega$. In general $\mu(A)$ is the probability that $\omega-$flips of a fair coin give rise to a binary sequence belonging to A. (see Nies (2009) or Downey and Hirschfeldt (2010) for all measure-theoretic detalis).

A probabilistic law, theoretically, is a set of measure-one and therefore, it will have the form $\mu\{\sigma \in 2^\omega | A(\sigma)\} = 1$. So, coming back to Ville's argument, if $A_0, A_1, A_2...$ is the collection of all sets such that $\mu(A_i) = 1$, then a sequence σ is typical, if belong to all these A_i, namely if $\sigma \in \bigcap_i A_i$. However notice that for the singleton $\{\sigma\}$, we have $\mu(\{\sigma\}) = 0$, since $\{\sigma\} \subseteq [\sigma_{\restriction n}]$, for all n and therefore $\mu(\{\sigma\}) \leq \mu([\sigma_{\restriction n}]) \leq 2^{-n}$, for all n. It follows that $\mu(\overline{\{\sigma\}}) = \mu(2^\omega \setminus \{\sigma\}) = 1$. But then each sequence is not in a set of measure 1, that is, the complement of its singleton. Ergo there are no " typical sequences " at all. We must restrict the collection of sets measure 1. We will say that a sequence that satisfies rare property (i.e. of zero measure) *effectively given*, is not random. Instead of belonging to all sets of measure one, we will say that a sequence is random, if avoids all null sets effectively given. A set X is *null*, if there is an infinite sequence of open sets $\{V_i\}_{i\in\mathbb{N}}$ such that $\mu(V_n) \leq 2^{-n}$ and $X \subseteq \bigcap_n V_n$. There are actually equivalent definitions of *null sets* which, being often used, are worth mentioning:

(a) A set A is *null*, if has measure zero.

(b) A set A is *null*, if there is an infinite sequence of open sets $V_0, V_1, V_2....$ such that $A \subseteq \bigcap_n V_n$ and $\mu(V_n) \leq 2^{-n}$.

(c) A set A is *null*, if there is an infinite sequence of open sets $V_0, V_1, V_2....$ such that $A \subseteq \bigcap_n V_n$ and $\lim_n \mu(V_n) = 0$.

What is meant by "effectively given"? Martin-Löf (1966) narrowed the concept of a "typical" sequence in order to avoid the difficulties that we have mentioned: we will not ask that a *random* sequence belongs to each set of measure one, i.e. that does not belong to no null set, but only that does not belong to null sets *effectively given*: this notion is, so to say, *costructivized*, considering only the laws of probability that can be proved in an effective way (it is provable that not all the laws of probability are effective in this sense). Hence, a sequence is random, if it meets all the probabilistic laws effectively given, i.e. belongs to all sets effectively given of measure one (i.e. all those we have called " typical " properties), or, equivalently, is *not* a member of any set of *zero measure* effectively given.

This brings us to the concept of *test*. If A is computably enumerable, that is, if $A = W_e$, then we speak of sets *effectively* open. The computably enumerable opens can be represented as:

$$X = [W_e] = [Dom(\phi_e)] = \{\tau \in 2^\omega | \sigma \subset \tau, \ for\ some\ \sigma \in W_e\}$$

for some index e. Formally a *Martin-Löf test* is a uniformly computably enumerable sequence of open sets $\{V_i\}_{i\in\mathbb{N}}$ such that for each $i \in \mathbb{N}$, $\mu(V_i) \leq 2^{-i}$. Thinking to an enumeration of computably enumerable sets $W_0, W_1, W_2...$, "uniformly" computably enumerable means that a test is univocally determined by an index e of a function $\phi_e(x)$ such that:

$$V_i = \bigcup_{\sigma \in W_{\phi_e(i)}} [\sigma] = [W_{\phi_e(i)}]$$

Notice that $\mu(\bigcap_n V_n) = 0$, namely is *effectively null* (sometimes is assumed that $V_i \supseteq V_{i+1}$, however this is not essential, since considering $U_n = \bigcup_{m>n} V_m$, we have that $\bigcap_i U_i = \bigcap_i V_i$ and the previous condition is satisfied for $\{U_i\}_i$).

Definition 61. *We say that:*

(a) An infinite sequence σ pass the test of $\{V_i\}_{i\in\mathbb{N}}$, *if* $\sigma \notin \bigcap_n V_n$.

(b) An infinite sequence σ *is* ML-random, *if pass all tests* $\{V_i\}_{i\in\mathbb{N}}$, *i.e. avoids all countable intersections of such null sets.*

This is the way in which in this formalism we express the fact that a sequence has no attribute "non-typical". An important concept is that of *Universal test.* It is possible to give an enumeration of the tests, starting from a enumeration of computably enumerable sets, discarding those that are too large:

$$W_{g(e,i),s} = \begin{cases} W_{\langle e,i\rangle,s} & \text{if } \mu\{[\sigma] | \sigma \in W_{\langle e,i\rangle,s}\} \leq 2^{-i} \\ \emptyset & otherwise \end{cases}$$

(where $n \in W_{e,s}$ if and only if $\phi_e(n)$ converges in at most s steps). If we take:

$$V_i = \bigcup_e \{[\sigma] | \sigma \in W_{g(e,i+e+1)}\}$$

we observe that $\mu(V_i) \leq \sum_e 2^{-(i+e+1)} \leq 2^{-i}$. The sequence $\{V_i\}_{i\in\mathbb{N}}$ costitutes a *universal* test, since for all other test $\{Z_i\}_{i\in\mathbb{N}}$, we have $\bigcap_i Z_i \subseteq \bigcap_i V_i$.

We will go no further in our discussion of the Martin-Löf approach because the work we will refer to is based on another of the previously mentioned equivalent notions of randomness: that of incompressibility.

8.3. Kolmogorov-Solomonoff-Chaitin's theory and incomputability

To account for the definition given by Solomonoff (1964) and Kolmogorov (1965), later adopted with some variations by Chaitin, let's start highlighting a few fundamental underlying ideas. Reasoning in general terms, certain seemingly random sequences can be described by relatively simple means, while the other, as for example 1001110111010110 does not seem have other description, except that by the mere repeatition *verbatim*: we will say that such sequences are *incompressible*, because their descriptive complexity is at least equal to their length. Reasoning more abstractly, if $\tau_0\tau_1\tau_2...\tau_n$ is a binary string (i.e. a member of the set $2^{<\omega}$ of finite binary sequences), it might be generated by a "shorter program" of the string itself, coded for example with k bits, for $k < n+1$: a finite string will be considered *random*, if cannot be generated by a program shorter than itself.

From this arises the following preliminary definition, relative to a Turing machine $\mathcal{M}$ (remember that $|w|$ denotes the length of the binary string w):

$$K_{\mathcal{M}}(x) = \begin{cases} |w| & w = \text{the shortest sequence such that } \mathcal{M}(w) = x, \text{ if exists} \\ \infty & \text{otherwise} \end{cases}$$

However, in order to disengage the definition from the particular machine $\mathcal{M}$, this was formulated rather in terms of a universal Turing machine U. For example $U(0^e1\sigma) \simeq \mathcal{M}_e(\sigma)$. Let therefore U be a universal machine; the Kolmogorov complexity of a finite stringe w with respect to U is given by the following function:

$$K_U(w) = min\{|z| \mid U(z) = w\}$$

The function $K_U(w)$ has to be readen "the length of the shortest program that (with respect to U) gives a description of w". This definition is related to the previous one by the fact that for each machine $\mathcal{M}$ there is a constant $c_{\mathcal{M}}$ such that $K_U(w) \leq K_{\mathcal{M}}(w) + c_{\mathcal{M}}$. Hence for two universal machines the Kolmogorov complexity is the same up to an additive constant, in the sense that if U, U' are two universal machines, we have that for some constant c, $K_U(w) \leq K_{U'}(w) + c$. We can therefore fix once and for all U and simply write $K(w)$. It is generally true that $K(w) \leq |w| + s$, for some constant s, and then we are talking of a total function.

Incompressible strings really exist? Yes!

Theorem 122. *There are incompressible strings of any fixed length n.*

Proof. Observe that there are $\sum_{k=0}^{n-1} 2^k = 2^n - 1$ programs of length minus than n, while the programs of length n are exactly 2^n:

$$\begin{array}{cccccccc} \emptyset & 0 & 1 & 00 & 01 & 10 & 11... \\ 0 & 1 & 2 & 3 & 4 & 5 & 6... \end{array} \tag{1}$$

If the programs σ of length n were all compressible, we would have $K(\sigma) < n$: hence for all σ of length n we would have a τ of length less than n that print it. There would therefore be a τ of this kind that prints two different programs (a contradiction). QED

Lemma 36. *There is a constant e such that $K_U(\sigma) \leq |\sigma| + e$.*

Proof. Let us take $\mathcal{M}_e(\sigma) = \sigma$ the *copying-machine*; hence σ is a program of $\mathcal{M}_e$ that prints σ: at worst, therefore, the minimum program of U that prints σ will have length $0^e 1\sigma$. QED

Lemma 37. *For all partial recursive function ϕ_e, $K(\phi_e(\tau)) \leq K(\tau) + c$.*

Proof. Let σ be of minimal length such that $U(\sigma) \simeq \tau$ and let:

$$Z(0^{\log(e)}1e\sigma) \simeq \phi_e(U(\sigma)) \simeq \phi_e(\tau)$$

Hence $K(\phi_e(\tau)) \leq |0^{\log(e)}1e\sigma| \leq 2|e| + 1 + |\sigma| \leq K(\tau) + c$. QED

An element of surprise and dissatisfaction with respect to this complexity measure, however, is its *character not additive*, that is, the complexity of the concatenation $\sigma = \tau\alpha$ may be bigger than the sum of the complexity of σ plus that of α. From an algorithmic point of view, a fact even more troublesome is the following.

Theorem 123. *$K(x)$ is not a computable function.*

Proof. Suppose that on the contrary it iscomputable. Suppose that we have ordered lexicographically the finite binary sequences $\emptyset < 0 < 1 < 00 < 01 < 10 < 11...$ and let:

$$y_m = min\{\sigma | K(\sigma) > m\}$$

where the minimum is taken with respect to the order $<$; let us consider all these $y_0, y_1, y_2.....$

(a) if $K(x)$ were computable, then there would be c such that $K(y_m) < |m| + c$. Indeed, let M be a machine that on n, first generates the strings $\sigma_0, \sigma_1, \sigma_2...$ in lexicographic order and the computes $K(\sigma_0), K(\sigma_1), K(\sigma_2)...$:

   i. if $K(\sigma_i) > n$, it prints σ_i and then halts.
   ii. Otherwise analyzes σ_{i+1}

Sooner or later will come a σ_i such that $K(\sigma_i) > n$; so, on input n, the program will produce y_n

(a) Hence $K_M(y_n) \leq |n|$ and we know thata in this case $K(y_n) \leq |n| + c$.

(b) Ergo $n < K(y_n) \leq |n| + c$, from which $n < |n| + c$, contradiction, because asymptotically n grows more than $|n| + c$.

QED

Although incomputabile, the function $K(x)$ is nevertheless *approximable*, i.e. there is a computable function g such that:

(a) $g(s + 1, x) \leq g(s, x)$ (decreasing in s)

(b) $\lim_{s\to\infty} g(s, x) = K(x)$ (computable from above, or right-computably enumerable)

Recall that in general if g is of this kind and $f(x) = \lim_{s\to\infty} g(s, x)$, then the set:

$$X = \{\langle\sigma, n\rangle \in 2^{<\omega} \times \mathbb{N} | f(\sigma) < n\}$$

is computably enumerable (and viceversa). Hence the set:

$$\{\langle\sigma, n\rangle \in 2^{<\omega} \times \mathbb{N} | K(\sigma) < n\}$$

is computably enumerable non computable.

Hence the incomputability of $K(x)$ can also be seen from an angle more abstract appealing to the notion of *simple set*, due to Post (see p. 49). Recall that a simple set can not be computable.

Theorem 124. *The set* $X = \{\sigma \in 2^{<\omega} | K(\sigma) < |\sigma|\}$ *is simple.*

Proof. By contradiction let $Z \subseteq \overline{X}$ computably enumerable and infinite. Remember that each computably enumerable set and infinite contains a computable set (Post), say $A = \{z_0, z_1, z_2....\}$. Notice that $K(z_i) \leq |i| + c$, because the program that prints z_i is obtainable from the machine that generates A and from the index i, that in the binary representation has length $\log(i)$. But for a z_i big enough this contradicts the fact that it is incompressible, as an element of $\overline{X}$. QED

We would like now to extend the notion of *incompressible string* to infinite sequences (i.e. real numbers between 0 and 1), that is, we would like to say that σ is random, if and only if for every n, $K_U(\sigma \upharpoonright n) \geq n$, but we will point out in this regard that machines then it must be understood as *prefix-free*, i.e. their domain will consist of binary strings, none of which is an initial segment of the other: for a result of Martin-Löf, dropping the restriction on machines to be *prefix-free*, we would have an empty definition, since no infinite sequence would satisfy the above condition.

Theorem 125. *For all* α *and infinite* n, $K(\alpha \upharpoonright n) < n - c$ *(where* $w \upharpoonright n =$ *the prime* n *bits of* w*).*

Proof. Here we see a version due to Katseff (1978): we write $\log(n)$ for abbreviating $\lfloor \log_2(n+1) \rfloor$="the integer part of $\log_2(n+1)$"; let f be the canonical correspondence between binary sequences and numbers that we have already used (a string σ is identified with the number n such that 1σ is the base-two representation of $n+1$), observing that $|f(n)| = \log(n)$. We consider a Turing machine $\mathcal{M}$ doing this:

(a) on input σ, look if $|\sigma|$ has the shape $k - \log(k)$, namely $k - |f(k)|$ and if the answer is yes, then it returns as output $f(k) \frown \sigma$ (recall that we used $\frown$ for concatenation of binary strings).

For instance if $\sigma = 0001 = f(16)$, then $|\sigma| = k - \log(k) = 4$, where $k = 7$, $\log(k) = 3$ and therefore $k - \log(k) = 4$. In this case the machine returns $f(7) \frown \sigma = 000 \frown 0001$.

Consider now that $\sigma = f(n)$ can be seen as the initial segment $f(n) = \sigma \upharpoonright m$ of an infinite sequence, fors ome m (in the previous example, $m = 4$ and $n = 16$); therefore take $\sigma \upharpoonright n = f(n) \frown \tau$ where $|\tau| = n - |f(n)| = n - \log(n)$ (once more following the previous example, $\sigma \upharpoonright n = \sigma \upharpoonright 16 = 0001 \frown \tau$, where τ contains 12 bits). Ergo, the machine $\mathcal{M}$ on input τ will return $f(n) \frown \tau$.

In conclusion, considering that:

$$K_{\mathcal{M}}(\sigma \upharpoonright n) = min\{|\alpha| | \mathcal{M}(\alpha) \simeq f(n)\tau\} =$$

$$= min\{|\alpha| | \mathcal{M}(\alpha) \simeq \sigma \upharpoonright n\} \leq |\tau|$$

for infinite n we will have $K_U(\sigma \upharpoonright n) \leq |\tau| + e \leq n - |n| + c$. QED

The problem inherent in the definition of $K(x)$ we used hitherto, has been highlighted by Chaitin with this example, consider a machine $\mathcal{M}$ that on input σ, first scans it for determine its length $|\sigma| = n$, then go back to start the computation of σ bit by bit. If now $\mathcal{M}(\sigma) \simeq \tau$, then the information necessary to get τ, intended as $K_M(\tau)$, is not dependent only by σ, but also by $|\sigma|$ and will be then encoded encoded by a string whose length is of order $n + \log(n)$. The problem does not occur if, for example, we consider machines *self-delimiting*: think to a Turing device where there is a single input tape of only reading, whose head flows only from left to right, a certain finite number of working tapes and an output tape.

The *self-delimiting* machines are not authorized to affix a symbol $*$ of "white" at the end of the input, which therefore is not delimited by any *end marker*; the input tape contains therefore only digit 0 and 1: a machine M on input σ, gives as a result τ, if after reading all σ, but the next bit, print τ; it is *self-delimiting* in the sense that it can determine where the input terminates, without the need to read the next symbol. The machine converges on $\sigma \in 2^{<\omega}$, if after a finite number of steps halts reading the last bit of the input tape and produces output. It is interesting to note that if $\mathcal{M}$ is *self-delimiting*, the set:

$$Dom(\mathcal{M}) = \{\sigma \in 2^{<\omega} | \mathcal{M}(\sigma) \downarrow\}$$

is *prefix-free*. For domains without prefixes, an important inequality, known as "Kraft's inequality", applies. The following, more refined version is actually due to Chaitin and it is the one which we use in Levin and Schnorr's theorem.

Theorem 126. (Kraft-Chaitin inequality) *Suppose we have an effective list:*

$$\langle n_0, \sigma_0 \rangle, \langle n_1, \sigma_1 \rangle, \langle n_2, \sigma_2 \rangle ...$$

(called bounded request), such that $\sum_{k \in \mathbb{N}} 2^{-n_k} \leq 1$. *Then it is possible to define a prefix-free machine such that for all* k, *there is a* τ_k *such that* $|\tau_k| = n_k$ *and the machine on input* τ_k, *outputs* σ_k.

Theorem 127. (Levin-Schnorr (1973)) *Let* $\sigma \in 2^{\omega}$. *Then the following are equivalent:*

(a) σ *is* ML*-random.*

(b) There exists a number b *such that for all numbers* n, $K(\sigma \restriction n) > n - b$

Proof. The theorem was proven independently in Levin (1973) and Schnorr (1973). $1 \Rightarrow 2$ follows from the simple observation that the cylinders $R_b = [\{\tau | K(\tau) \leq |\tau| - b\}]$ give rise to Martin-Löf test $\{R_b\}_{b \in \mathbb{N}}$. Actually $K(\tau) \leq |\tau| - b$ if and only if $\exists \sigma \exists s (U_s(\sigma) \simeq \tau \wedge |\sigma| \leq |\tau| - b)$; it is therefore a Σ^0_1 formula and the sequence $\{R_b\}_{b \in \mathbb{N}}$ can be expressed in a uniformly computably enumerable way.

Let now X_b be the set of σ such that $K(\sigma) \leq |\sigma| - b$. Hence $\mu(R_b) = \sum_{\sigma \in X_b} 2^{-|\sigma|}$. Notice that if $\sigma \in X_b$, then $|\sigma| \geq K(\sigma) + b$. Moreover using Kraft's inequality (since $Dom(U)$ is prefix-free):

$$2^{-b} \cdot \sum_{\sigma \in X_b} 2^{-K(\sigma)} \leq 2^{-b} \cdot \sum_{\sigma \in Dom(U)} 2^{-|\sigma|} \leq 2^{-b} \cdot 1$$

Hence:

$$\sum_{\sigma \in X_b} 2^{-|\sigma|} \leq 2^{-b} \cdot \sum_{\sigma \in X_b} 2^{-K(\sigma)} \leq 2^{-b} \cdot 1$$

Namely $\mu(R_b) \leq 2^{-b}$. Now, if σ does satisfy 1. then $\sigma \notin \bigcap_b R_b$ and therefore there is a b such that $K(\sigma \restriction n) > n - b$ for all n; hence does satisty 2.

As regards instead $2 \Rightarrow 1$, suppose that $\sigma \in \bigcap_m V_m$, for some test $\{V_m\}_{m \in \mathbb{N}}$. Remember that each open computably enumerable can be represented (uniformly in m) in the form:

$$V_m = [\{\alpha_{i,m} | i < k_m\}]$$

where $\{\alpha_{i,m} | i < k_m\}$ is a *prefix-free* set and $k_m \in \mathbb{N} \cup \{\infty\}$; recall the notion of *bounded request*, i.e. a set of pairs $W \subseteq \mathbb{N} \times 2^{<\omega}$ such that $\sum \{2^{-x} | \langle x, y \rangle \in W\} \leq 1$. This sum is called the "weight" of W.

Let us consider therefore the *bounded request*

$$W = \{\langle |\alpha_{i,m}| - m + 1, \alpha_{i,m}\rangle | m \in \mathbb{N}, i < k_m\}$$

We can assume w.l.o.g. that $\mu(V_m) \leq 2^{-2m}$ and check that the contribution of V_m to the weight of W is at most 2^{-m-1}. Actually $\sum_{i \leq k_m} 2^{-|\alpha_{i,m}|+m-1} = \mu(V_m) \cdot 2^{m-1} \leq 2^{-m-1}$. The sum of these values, that is, $2^{-1} + 2^{-2} + 2^{-3} + ...$ for each m is equal to 1, i.e. W is a bounded request. Kraft and Chaitin's theorem states that we can effectively build a machine M_d such that $\langle x, \alpha\rangle \in W$ if and only if there is a string τ such that $|\tau| = x$ and $M_d(\tau) = \alpha$, where $x = |\alpha_{i,m}| - m + 1$ and $\alpha = \alpha_{i,m}$ (and therefore $K_{M_d}(\alpha_{i,m}) \leq |\alpha_{i,m}| - m + 1$) for some $m \in \mathbb{N}, i < k_m$. QED

8.4. Incompleteness and randomness

Chaitin's basic idea was to measure the information content of a theory using the notion of algorithmic complexity. This idea turned out to have strong implications in the analysis of the phenomenon of incompleteness. Here we shall analyse in particular the relationship with Gödel's first incompleteness theorem (for the second, see Kritchman and Raz (2010)). Chaitin proved a version of the first incompleteness theorem which says that, among true, but *unprovable* formulas there are all true statements $K(u) > c$ for a certain constant c (for all finite binary strings u). According to Chaitin this constant is somehow a measure of the information content of the theory. This very elegant version of Chaitin's result is attributed in Van Lambalgen III (1987) to Albert Visser and Dick de Jongh. Recall that to define $K(x)$ we used the universal machine U defined on input of the form $0^e 1\sigma$, that simulates ϕ_e on σ. Hence if $\phi_e(e) = n$, then $K(n) \leq K_{\phi_e}(n) + e + 1 \leq 2e + 1$.

Theorem 128. *Let $w \in 2^{<\omega}$. There exist a constant c such that* PA *does not prove any statement of the kind of '$K(w) > c$'.*

Proof. Fix an enumeration of the derivations of PA; let ϕ_e the following partial recursive function:

$\phi_e(m) = n$ if and only if n is that n occurring in the first proof in PA of a sentence of the form "$\phi_m(m) \neq n$".

It is provable that $\phi_e(e)$ is not definite: indeed, if it were defined, for instance $\phi_e(e) = n$, then PA would prove $\phi_e(e) \neq n$ (contradiction, under the hypothesis of soundness). Hence, notice that PA + $\{\phi_e(e) = n\}$ is consistent. But this implies that PA + $\{K(n) \leq 2e + 1\}$ is consistent and therefore PA cannot prove statements of the form $K(n) > 2e + 1$. QED

The minimal c such that PA $\nvdash K(\sigma) > c$ is called by Chaitin "characteristic constant"; according to Chaitin it depends only by the complexity of the axioms and is related to the information content. Hence true statements like the above, expressed in the language of a theory as PA, are unprovable because their information content is *higher than that of the axioms* of the theory itself. This interpretation has been strongly questioned in Van Lambalgen III (1987), Franzen (2005) and in particular Raatikainen (1998) and Raatikainen (2000). For example, we can collapse the characteristics constant to zero, or we can make it arbitrarily large. These constants are dependent on factors quite different from the information content. Raatikainen's criticism, in particular, is that the above results due to Chaitin depend essentially on the *acceptable system of indexes* adopted. Let the standard indexing be $\phi_0, \phi_1, \phi_2....$ An indexing $\psi_0, \psi_1, \psi_2...$ is called acceptable, if there are computable functions f, g such that $\phi_e \simeq \psi_{f(e)}$ and $\psi_e \simeq \phi_{g(e)}$.

The acceptable indexing meet in particular the fixed point theorem, i.e. for each total recursive function f, there is a number e such that $\phi_e \simeq \phi_{f(e)}$. It is precisely through a simple

application of fixed point that we can collapse to zero the characteristic constant: in fact we take a theory T sufficiently strong a let π an acceptable coding of Turing machines; let us define *ad hoc* the following indexing:

$$\pi^n(x) = \begin{cases} 0 & \text{if } x = n \\ x+1 & \text{if } x < n \\ x & x > n \end{cases}$$

Raatikainen uses this equivalent definition of the Kolmogorov complexity:

$$K(\sigma) = min\{e|\phi_e(0) \simeq \sigma\}$$

namely the smallest description of σ on a fixed input 0. On the basis of this new indexing we can define the algorithmic complexity relative to it as:

$$K^n(\sigma) = min\{k|\exists y(\pi^n(y) = k \wedge \phi_y(0) \downarrow = \sigma)\}$$

Now we define a program P_m in this way: on input 0 look for the minimum x such that $\mathsf{Prf}_{\mathsf{T}}(x, \overline{\ulcorner K^n(\sigma) > \overline{0}\urcorner})$, for some σ; if you find it, print σ.

With the help of the fixed point theorem, we can make sure that the code of this program coincides with the parameter n in $K^n(x)$: if $f(n) = m$ is the program code, there will be an e such that $P_{f(e)} \simeq P_e$.

However P_e never halts: indeed it looks for the minimum x such that:

$$\mathsf{Prf}_{\mathsf{T}}(x, \overline{\ulcorner K^e(\sigma) > \overline{0}\urcorner})$$

and if it finds such a minimum, it prints σ. If there was a similar x with a minimal length of proof, then P_e would print σ: ergo $K(\sigma) \leq e$. But $\pi^e(e) = 0$ and therefore $K^e(\sigma) = e$; hence if hypothetically we had proved $K^e(\sigma) > 0$, thanks to the *soundness* we would have a contradiction. Since P_e never halts, there is no proof of $K^e(\sigma) > \overline{0}$, i.e. $c_T = 0$.

With similar arguments Raatikainen shows that c_T can be made arbitrarily large. What is therefore the true source of "characteristic constants"? According to him the value of c_T is actually determined simply by the smallest (by its code) Turing machine which does not halt, but for which this cannot be proved in T.

In defence of Chaitin's argument Ferbus-Zanda and Grigorieff (2014) point out, however, that modelization rarely rules out all pathological cases and it is intended to be used in "reasonable" cases (perfect modelization is illusory). Mathematically more significant is the defence in Calude and Jürgensen (2005): Chaitin's "heuristic principle", i.e. the thesis that the theorems of a theory cannot be significantly more complex than the theory itself, is defensible as long as the measure of complexity is further specified. By changing the measure of complexity, from program-size $K(x)$ to the measure $\delta(x) = K(x) - |x|$, it is proved that for any sound, consistent theory strong enough to formalize arithmetic and for any Gödel numbering of its well-formed formulas, we can compute a bound N such that no sentence x with complexity $\delta(x) > N$ can be proved in the theory and this phenomenon is independent on the choice of the Gödel numbering.

In Chaitin (2007) it is introduced a particular infinite binary sequence (i.e. a real), called Ω, with singular characteristics. With emphasis the autor says that he is able to construct a much more uncomputable real than Turing does. In this section we will attempt to provide an introduction to what is considered one of the most important concepts in Algorithmic Information Theory. It is necessary to recall a few notions first.

Definition 62. *The Turing computable reals are define as follows:*

(a) *A real α is* computable*, iff is the limit of a computable sequence of rationals*[1] *$r_0, r_1, r_2...$ for which there exists a computable function f such that for all n, $|\alpha - r_{f(n)}| < 2^{-n}$ (i.e. has a computable rate of convergence). E.g. the rationals, $\sqrt{2}$, π, e.*

(b) *A real r is called* recursively enumerable*, if can be approximated in an effective way: $r = \lim_n r_n$, for a computable non decreasing sequence of rationals $r_0, r_1, r_2....$ Equivalently, if it is* left-computable *(or* lower-semicomputable*), i.e. if $L(r) = \{q \in \mathbb{Q} | q < r\}$ is computably enumerable*

Recall that any real in $[0, 1]$ con be associated with a binary sequence: if $\varepsilon = \varepsilon_0\varepsilon_1\varepsilon_2...$ is a binary sequence we can associate to it the real:

$$r_\varepsilon = \varepsilon_0 2^{-1} + \varepsilon_1 2^{-2} + \varepsilon_2 2^{-3} + ...$$

So, a real r in the interval $[0, 1]$ will be written as $\sum_i \epsilon_i 2^{-(i+1)}$.

There are other characterizations of computable reals. E.g.

(a) $r \in [0, 1]$ is computable iff $L(r) = \{q \in \mathbb{Q} | q < r\}$ is computable;

(b) Recall that r can be written as $\sum_i \epsilon_i 2^{-(i+1)}$. We can identify a real $r \in [0, 1]$ with the set A such that the $n - th$ bit of the binary expansion $\epsilon_0, \epsilon_1\epsilon_2....$ of r is 1, iff $n \in A$ (we write $r = 0.A$ to mean that $r = \sum_{i \in A} 2^{-(i+1)}$); then we can say that r is computable iff A is computable. Therefore a real is computable if and only if its binary expansion is computable, i.e. if there is an algorithm to calculate its bits.

Theorem 129. *A real is computable iff it is the characteristic function of some computable set A.*

This notion can be relativised. Hence $r = 0.A$ is computable in $\emptyset'$ (hence is in Δ_2, by Post's results) iff A is computable in $\emptyset'$, iff the left-cut $L(r)$ is computable in $\emptyset'$. Lastly, real is computable if and only if its binary expansion is computable, but a real may be computably enumerable even if its binary expansion *is not* computably enumerable Actually we must distinguish the weaker property of being an computably enumerable real from the stronger one of being a real whose binary expansion is computably enumerable:

(a) A real r is left-recursively enumerable, iff $L(r) = \{q \in \mathbb{Q} | q < r\}$ is computably enumerable

(b) However a real left-computably enumerable *is not* one of the form $0.A$ where A is r.e, namely where $A(0), A(1), A(2), ...$ is the "characteristic function" of a set computably enumerable

(c) We call *strongly computably enumerable* the reals *left computably enumerable* in which this happens, namely whose binary expansion is the" characteristic function " of a set computably enumerable (in the sense that for some e, $\phi_e(n) \downarrow$ iff $A(n) = 1$).

Only the converse is true: if r is strongly computably enumerable then is also left-computably enumerable.

We therefore arrive at a central result.

Theorem 130. *There are reals left computably enumerable that are not strongly computably enumerable.*

[1] A sequence $\{r_i\}_i$ of rationals is computable iff there are computable functions f, g such that $r_i = \frac{f(i)}{g(i)}$.

Typical example of this kind of reals are the constants Ω_U introduced by Chaitin. There are real left-computably enumerable random; they are all and only those of the form Ω_U, for some universal machine U. The number Ω_U depends on U as the halting-set K depends on the enumeration of partial recursive functions; for a theorem due to Myhill however, all K are equivalent, up to a computable permutation. Similarly, all versions of Ω have similar properties that in this phase allow us sometimes to leaving aside from the particular machine. If for example we define:

$$\Omega_s = \sum\{2^{-|\sigma|} | U(\sigma) \; converges \; in \; at \; most \; s - steps\}$$

then Chaitin's Ω is the limit of this non-decreasing computable sequence of rational $\Omega = \lim_s \Omega_s$ (and is therefore computably enumerable in the first sense), however, its binary representation is *random* (and therefore not computably enumerable). An incompressible binary sequence σ, *cannot be* computably enumerable, if for computably enumerable we mean in the second sense, namely that there is a number e such that $\phi_e(i) \downarrow$ if and only if $\sigma(i) = 1$ and where $\sigma(i)$ is the $i - th$ bit of σ. This is the content of the following result.

Lemma 38. (Barzdins 1968) *If σ is computably enumerable in the strong sense, then $K(\sigma \upharpoonright n) \leq 2|n| + e$.*

Proof. For printing $\sigma \upharpoonright n$ we must know: 1) the length n (coded by $\log(n)$-bits), 2) the number k of elements of W_e less than n (coded also with at most $\log(n)$-bits) and 3) the index e. If we provide these informations we consider these steps of computation: generate W_e, until k elements have been enumerated. Then observe that the input $\langle n, k \rangle$ is coded by at most $2 \cdot \log(n) + c$-bits. QED

As we have said, this is the case of numbers Ω_U, whose binary expansion is incompressible. These numbers may be seen as expressing the probability that a universal Chaitin machine U halts when it receives as input a binary string, determined for example through the launch of a coin. Consider that the probability of getting the program u by launching a coin is $2^{-|u|}$, and therefore the probability of fish a program u such that $U(u) = n$ is:

$$\sum\{2^{-|u|} | U(u) = n\}$$

From this comes the definition of the probability that a universal machine U sooner or later halts, on programs selected by lot:

$$\Omega_U = \sum\{2^{-|p|} | U(p) \downarrow\}$$

The Ω_U are real numbers with very peculiar properties. Recall that they depend on the choice of U and then there are, not just one, but a class. In Kučera and Slaman (2001) is proved that each *random left computably enumerable* real coincides with some Ω_U. To summarise, the Ω_U also fulfils these properties:

(a) in terms of binary expansion, they are "chaotic", and therefore incompressible and a fortiori not computably enumerable

(b) From Kraft's inequality (Theorem 126) we have $\Omega_U \leq 1$. Since U does not converge on all strings, actually we have $\Omega_U < 1$. Since on the other hand converges on *some* string, also we have $0 < \Omega_U$; hence, the Ω_U These are reals irrationals strictly between 0 and 1.

(c) Each Ω_U is computable from $\emptyset'$; hence it is Δ_2^0 and by the "Limit lemma" on p.60 is computable in the limit: $\Omega_U(x) = \lim_s f(x, s)$ for some computable functions f with values 0-1. In other words, it is generated by a procedure of type "trial and error" (Putnam).

Let us fix now a universal machine U, namely the U defined as $U(0^e1\sigma) \simeq \phi_e(\sigma)$, for which we have $K_U(\sigma) \leq K_M(\sigma) + c_M$, for all other machines M and let $\Omega = \Omega_U$. Before proving some important properties of Ω, recall that for the binary expansions of real numbers $\alpha = \sum_i \varepsilon_i 2^{-(i+1)}$ where $0 < \alpha \leq 1$ and in $\varepsilon_0\varepsilon_1\varepsilon_2...$, for all i, $\varepsilon_i \in \{0, 1\}$ the following holds:

$$\underbrace{\sum_{i<n} \varepsilon_i 2^{-(i+1)}}_{\alpha\restriction n} < \alpha \leq \underbrace{\sum_{i<n} \varepsilon_i 2^{-(i+1)}}_{\alpha\restriction n} + \underbrace{\sum_{i\geq n} \varepsilon_i 2^{-(i+1)}}_{2^{-n}}$$

where $\alpha \restriction n = \alpha(0), \alpha(1), ..., \alpha(n-1)$.

We remark that the real numbers in $[0, 1]$ that have in common with α the first n elements of their dyadic expansion (the "cylinder" of $\alpha \restriction n$) are in this interval. Hence these additional inequalities hold:

(a) $\Omega \restriction n < \Omega$

(b) $\Omega \leq \Omega \restriction n + 2^{-n}$

If $\Omega_s = \{\sum 2^{-|\sigma|} | U(\sigma) \downarrow \ \textit{in at most } s - \textit{steps}\}$, we have that if $\Omega_s > \Omega \restriction n$, then for all σ of length shorter than or equal to n, if it did not appear in programs that contribute to the determination of Ω_s, then $U(\sigma) \uparrow$. Suppose on the contrary that this is false; hence σ would add a contribution of the amount $2^{-|\sigma|} \geq 2^{-n}$, from which:

$$\Omega \geq \Omega_s + 2^{-|\sigma|} > \Omega \restriction n + 2^{-n}$$

(against as determined at 2.). These observations allow us to highlight the link between *Chaitin's Ω and Turing's Halting Problem.* Knowledge of Ω permits indeed to *solve* the "Halting Problem": later we will prove that in fact $\Omega =_T K$. Actually the knowledge of the first $n - bits$ of Ω_U would solve the halting problems coded with at most n bits, or decide the sentences that can be codes with at most n-bits. It should first be noted that:

(a) if the length of a program is n, its contribution to determine Ω_U is 2^{-n};

(b) also applies the general condition for which $\Omega_U \leq \Omega \restriction n + 2^{-n}$.

So we begin a systematic check of all programs of any length, until we have found enough programs that halts and that allow us to go beyond $\Omega \restriction n$ (suppose to have reached a approximation $\Omega' > \Omega \restriction n$). Hence, if a string w of length less than or equal to n is not among these, $U(w)$ will never halts: if the machine stopped, accordingly we would that $\Omega_U \geq \Omega' + 2^{-|w|} > \Omega \restriction n + 2^{-n}$, against the above conditions. We are therefore able to decide the halting problem for w and Chaitin spoke of "enormous wisdom concentrated in a small space".

Theorem 131. *Ω_U is a random real.*

Proof. Let $\sigma_0, \sigma_1\, \sigma_2....$ an enumeration of the domain of U. Add $2^{-|\sigma_i|}$ to Ω as σ_i is enumerated and let:

$$\Omega_s = \sum_{U(\sigma)\downarrow \ in \leq s \ steps} 2^{-|\sigma|}$$

Observe that there are 2^n strings of length n and therefore at most a finite number of possible changes in passing from the initial segment $\Omega_s \restriction n$ to the segment $\Omega_{s+1} \restriction n$ (if there were more than 2^n possible changes, would mean that the number decreases or exceeds 1, but both things are impossible, since $\Omega_s \leq \Omega_{s+1}$).

Hence there will be a stage k at which it will become stable $\Omega_k \upharpoonright n = \Omega \upharpoonright n$. If you knew what $\Omega \upharpoonright n$ is, we would be able to determine $k = \psi(\Omega \upharpoonright n)$; namely, we could express $\Omega \upharpoonright n$ in this way:

$$\Omega \upharpoonright n = (\sum_{i \leq \psi(\Omega \upharpoonright n)} 2^{-|\sigma_i|}) \upharpoonright n$$

for some partial recursive function ψ, where $\sigma_0, \sigma_1, \sigma_2 ...$ is an enumeration of $Dom(U)$. Observe that at stage $\psi(\Omega \upharpoonright n)$ will be, however, already appeared all strings σ of length n such that $U(\sigma)\downarrow$. Suppose that this is not true: then if another appeared after, we would have $\Omega \upharpoonright n + 2^{-n} < \Omega$, being an initial segment of Ω (against the inequality at point 2. above mentioned). But by definition for all $\tau \in 2^{<\omega}$ such that $K(\tau) \leq n$, there is a σ of length less or equal to n such that $U(\sigma) \simeq \tau$. Hence $\sigma = \sigma_i$, for some $i \leq \psi(\Omega \upharpoonright n)$. In other words, if $\tau \notin \{U(\sigma_i) | i \leq \psi(\Omega \upharpoonright n)\}$, then $K(\tau) > n$. If now F is a function that on input $(\Omega \upharpoonright n)$ select a $\tau \notin \{U(\sigma_i) | i \leq \psi(\Omega \upharpoonright n)\}$, note that $K(F(\Omega \upharpoonright n)) > n$. However we have mentioned that for all $\tau \in 2^{<\omega}$, $K(F(\tau)) \leq K(\tau) + c$, from which $K(\Omega \upharpoonright n) > n + c$ for all n, namely, Ω is incompressible. QED

Theorem 132. $\emptyset' \equiv_T \Omega$.

Proof. First show that $\emptyset' \leq_T \Omega$. Let us take a machine $\mathcal{M}(0^n) = s$ if and only if $\phi_{n,s}(n)\downarrow$ and let e the code of this machine. We decide whether $n \in K$ by means of an oracle in Ω. We remark preliminarily that $U(0^e 10^n) \simeq \phi_e(0^n)$ and that $\phi_n(n)\downarrow$ if and only if for some s, $U(0^e 10^n) \simeq s$. If we have an oracle in Ω to establish its first n bits, then we are able to find s^* big enough to have $\Omega_{\upharpoonright n+e+1} = \Omega_{s^* \upharpoonright n+e+1}$. If $\phi_n(n)\downarrow$, this must have happened before the stadium s^*, otherwise we would have that for some $t \geq s^*$, $U(0^e 10^n)\downarrow = t$ and $0^e 10^n$ would contribute to the determination of Ω with $2^{-|0^e 10^n|}$ and therefore $\Omega_{\upharpoonright n+e+1} + 2^{-|0^e 10^n|} < \Omega$ (contradiction, against what we have seen above). Hence $n \in \emptyset'$ if and only if it was added before the stage s^*.

For the other way round, recall that α is left computably enumerable iff α is the limit of a non decreasing sequence of rationals. Since Ω fulfils the second property, it is therefore left-computably enumerable, and therefore is computable from the halting set K. QED

The author's *interpretation* of this undeniably interesting result has also been the object of criticism in Raatikainen (2000). Chaitin says:

> This is an impenetrable stone wall, it's a worst case. From Gödel we knew that we couldn't get a formal axiomatic system to be complete. We knew we were in trouble, and Turing showed us how basic it was, but Ω is an extreme case where reasoning fails completely (Chaitin (2007), p. 93).

But in what sense, then Ω would be an extreme example of unsolvability, if it is at the level of $\mathsf{0}'$ namely at the level of the halting set K in the upper semilattice of Turing degrees? Actually there are sets rather simple as $X = \{x | W_x \; infinite\}$ and $Y = \{x | W_x \; finite\}$ that are respectively $\Pi^0_2 - complete$ and $\Sigma^0_2 - complete$ (recall that Z is $\Sigma^0_n - complete$, if it is Σ^0_n and for any other A that is Σ^0_n, $A \leq_m Z$). Sets X and Y are therefore more difficult to calculate than Ω. Moreover, it is well known that $\mathsf{PA} + Th_{\Pi_1}(\mathbb{N})$ decides the halting problem, and then also Ω. But it is not able to decide the Paris-Harrington sentence of the previous chapter. So in what sense - Raatikainen asks - is the undecidability of Chaitin's constant *extreme*?

In Chaitin (2007) the Argentine mathematician obtains a further result of incompleteness establishing essentially that if T is a theory capable of interpreting PA, then we can explicitly compute a bound on the number of bits of Ω that can be determined by T. Of the variations on this result obtained subsequently, we believe the clearest and most elegant is the following one, due to Solovay (2000).

Theorem 133. *Let* T *be a theory sufficiently strong and* Σ_1^0 *– sound. It is possible to effectively build a universal machine* U *(provably in* PA*) such that* T *cannot even prove a single statement of the form:"the* n*-th bit of* Ω_U *is* k*".*

Proof. Preliminarly it should be noted that if the theory is Σ_1^0 sound, it is true also each Π_2^0-statement provable in it. Indeed, suppose by contradiction that $\forall x \exists y \phi(x, y)$ was provable but false. Hence for some n it will be true $\neg \exists y \phi(n, y)$. At the same time, however, will be provable also $\exists y \phi(n, y)$, against the Σ_1^0-soundness. A careful formalization shows that the statement: "the $n - th$ bit in Ω_U is k" has complexity Π_2^0; so if it can be proved in a theory Σ_1^0-sound, then is also true. We agree to start counting from zero. Therefore, the first element of a sequence will be the 0^{th} bits etc. Hence the proof consists of three steps:

(Step 1) fix a universal Turing machine V (provably in PA) and define $U(\sigma)$ by cases:

(a) if $\sigma = \emptyset$, then $U(\sigma) \uparrow$;
(b) if $\sigma = 0\tau$, let $U(\sigma) = V(\tau)$;
(c) if $\sigma = 1\tau$, then go to the second step:

(Step 2) Define first an algorithm $\psi(e, \sigma)$ as follows:

(i) preliminarly determine a pair $\langle n, k \rangle$ by means of the following computation: list the theorems of T until a theorem of the form "the n-th bit of Ω_{ϕ_e} is k" appears. If it appears, fix n, k (if the computation does not converge, let $U(\sigma) \uparrow$, for all σ that falls under the case (c)).
(ii) Fixed n, k and recalling that $\sigma = 1\tau$, if $|\tau| \neq n$, let $U(1\tau) \uparrow$; if instead $|\tau| = n$, then $U(\sigma)$ will be defined as follows: let r the diadic rational[2] whose binary expansion is $\tau k 000000....$ and let $r' = r + 2^{-(n+1)}$; look for the minimum s such that $\Omega_{\phi_{e,s}} \in (r, r')$ (clearly this search may fail). Then verify if $\phi_{e,s}(\sigma) \downarrow$. If yes, let $U(\sigma) \uparrow$, otherwise $U(\sigma) = \emptyset$.

The following applies:

i. The universality (provable) of U, follows from the definition of U on strings that begin with 0, from that of V.
ii. The domain of U is prefix-free: Suppose fact that $\sigma_1, \sigma_2 \in Dom(U)$ and $\sigma_1 \subseteq \sigma_2$; as by definition U is not defined on the empty string, then σ_1, σ_2 will be of length bigger than 0. Hence they will be of the form $\sigma_i = r * \tau_i$, where $\tau_1 \subseteq \tau_2$:
   A. if $r = 0$, then $\tau_1, \tau_2 \in Dom(V)$, that is prefix-free and therefore $\sigma_1 = \sigma_2$.
   B. if $r = 1$, than by definition $U(1 * \tau_i)$ is defined if and only if $|\tau_i| = n$. But therefore $|\sigma_1| = |\sigma_2|$ and then $\sigma_1 = \sigma_2$.

(Step 3) We can think of this algorithm through which we define U, starting from ϕ_e and σ, as a function $\psi(e, \sigma)$ and then apply to it the fixed point theorem. This, together with the parameterization theorem allow to deduce that there is an e such that $\psi(e, \sigma) \simeq \phi_e(\sigma)$. Let therefore this $\phi_e = U$ just that involved in the previous step.

[2] Recall that a rational of the form $r \cdot 2^{-n}$ (where r is an integer and n a natural number) is called "dyadic"; each dyadic rational has two binary expansions, and one of two has a tail of the form 0000.... Recall also that a dyadic interval is of the form:

$$[\frac{m}{2^s}, \frac{m+1}{2^s}]$$

There is a correspondence between cylinders $[\sigma]$ and dyadic intervals: in the above, if $|\sigma| = s$, then $m = \sum_{i=1}^{s} 2^{s-1} \cdot \sigma(i)$. Notice that if σ is a finite binary string then the leftmost element (thinking to the binary tree) in the cylinder generated by it is $\sigma 0000000...$, and the rightmost is $\sigma 111111.....$ Indeed, $m + 1 \cdot 2^{-s} = m \cdot 2^{-s} + 2^{-s} = m \cdot 2^{-s} + \sum_{i>s} 2^{-i}$, namely σ followed by an infinite tail of only digits 1 (see e.g. Nies (2009) 12-3.

We check that the hypothesis that the theory can determine some bits of

$$\Omega_U = \Omega_{\phi_e}$$

leads to a contradiction: suppose in fact that we have obtained n, k according to the described procedure; consider (r, r') where $r = h/2^{n+1}$ and $r' = r + 2^{-(n+1)} = h + 1/2^{n+1}$, such that:

$$\Omega_{\phi_e} \in (\frac{h}{2^{n+1}}, \frac{h+1}{2^{n+1}}) = (r, r')$$

Since the theory is *sound*, if it says that the $n - th$ bit of Ω_{ϕ_e} is k, then this must be true. So the binary expansion of r must be of the form $\tau k0000.....$, where $|\tau| = n$. For m big enough we will have that $\Omega_{\phi_e,m} \in (r, r')$; hence consider that (r, r') is the pair required by the computation at point 2.ii and the search for an m such that $\Omega_{\phi_e,m} \in (r, r')$ is successful. Let $\sigma = 1\tau$. We wonder therefore, if $\phi_{e,m}(\sigma)$ converges and the answer is no, because otherwise by definition $U(\sigma) \uparrow$; but since $W_{e,m} \subseteq W_e$, if $\phi_{e,m}(\sigma)$ converged, we would have $\sigma \in W_e$. On the other hand $W_e = Dom(U)$ and therefore $U(\sigma) \downarrow$, contradiction. Hence $\phi_{e,m}(\sigma)$ does not converge and again by the definition in 2.ii $U(\sigma) \downarrow = \emptyset$; hence there is a σ in the domain of U such that $\sigma \notin Dom(\phi_{e,m})$ and $|\sigma| = n + 1$, from which $\Omega_U \geq r + 2^{-(n+1)}$, against the definition of r and the fact that $\Omega_U \in (r, r')$ (contradiction).

QED

The author summarises the relevance he attaches to his findings as follows:

> My work is a fundamental extension of the work of Gödel and Turing on undecidability in pure mathematics. I show that not only does undecidability occur, but in fact sometimes there is complete randomness, and mathematical truth becomes a perfect coin toss (Chaitin (2003), pp. 109-110).

However also in this case the received view is not free from criticism. Recall that Ω is Δ_2 definable, hence computable by a *trial-and-error* machine. For this reason Raatikainen (2000) pp. 221-222 points out that Ω can still be generated by a completely deterministic procedure. Actually the Finnish logician seems to question the plausibility of the theory of algorithmic randomness itself, although, as we have seen, these types of deterministic machines are not equivalent to Turing machines. That is, we are in the domain of hypercomputation, i.e. of hypothetical and unrealistic models of computation that transcend the Church-Turing thesis.

8.5. Farewell: randomness, incompleteness and physical theories

Although the definition of "randomness" as *avoidance of null sets* proposed by Martin-Löf and its equivalents to some extent constitute the standard notion, nevertheless other definitions of randomness have also been proposed. For instance in Schnorr (1971), it is criticised that Martin-Löf randomness was too strong, and two weaker randomness notions are introduced. Martin-Löf-randomness is conversely considered by some to be too weak a notion of randomness. The notion proposed in Demuth (1982), an expression of the Russian school of constructive mathematic, is for instance stronger than Martin-Löf-randomness. But sticking to the standard definition, are all random sequences, so to speak, "equally random" and equally powerful? It may be perplexing that a random sequence is strong enough to compute the Halting Set $\emptyset'$, as is the case with Ω (it is said that Ω can be seen as a highly compressed version of $\emptyset'$ and this seems to be unintuitive). The Gács and Kǔcera theorem 120 also returns computability-theoretically powerful random sequence. However, in Stephan (2006) it is proved that there are (only) two types of random sequences: those that compute $\emptyset'$ and those that are computationally much weaker, which fail to compute a complete extension of PA, that are less powerful than $\emptyset'$ and for somebody these are more intuitively "random". It has

been proved that almost all random sequences are indeed in this second class (see Downey and Hirschfeldt (2019) for an *excursus*).

The theory of randomness we have seen is founded on computability theory. We also pointed out that doubts have been raised as to whether algorithmic randomness is the best account of the notion of randomness. It may indeed be surprising that a deterministic algorithm is involved in the definition: are there alternative proposals? Interestingly, the concept of "random sequence" or "random real" has also been studied without resorting to the notion of an algorithm. In particular, Solovay (1970), by using sophisticated methods and results from *Descriptive Set Theory*, introduced a notion of *real random number* based on set theory, with a peculiar extension of the technique of forcing (see Jech (1978) pp. 493-563 for a detailed account and Durand, et al. (2003) and Ferbus-Zanda and Grigorieff (2014) for a discussion and for some development of non-algorithmically-based randomness approaches and Lafitte (2002) for links with strong axioms of infinity). Remaining within the sphere of *Set Theory*, Kreisel (1969) proposed the programme of expanding the theory with new primitive concepts and new predicates in the language, in addition to set membership, in order to seek solutions to classical problems that were undecidable in standard axiomatisations: among these new predicates to be added, the primitive predicate of being a random sequence. The idea therefore is to abandon the project of giving an explicit characterisation of the concept of 'random sequence', and instead move towards an axiomatisation of randomness. A first attempt at axiomatisation goes back to Myhill in 1963. The axiomatic approach has been systematically pursued in Van Lambalgen (1990) and Van Lambalgen (1992) by adopting Kreisel and Myhill's observations about the substantially intensional character of the notion of 'random sequence', or at least not fully extensional, starting from the consideration that none of the current formalizations captures its meaning exhaustively.

The ZFR theory, obtained from the set theory ZF (i.e. Zermelo-Fraenkel set theory without the *Axiom of Choice*) by adding the Van Lambalgen axioms for the primitive predicate $R(x, y)$= "x is independent of y", i.e. "y has no information for x" , or "x is random with respect to y" (where $R(x) = R(x, \emptyset)$), turned out to be for example incompatible with the *Axiom of Choice*. The Dutch logician also proposed another axiomatization, in terms of generalized quantifiers of the type $Qx\phi(x)$= "if x is generated randomly, then it is practically certain that $\phi(x)$". The theory ZFQ , obtained from ZFC (i.e. ZF plus the *Axiom of Choice*) by adding these axioms is conservative on ZFC. Actually the theories ZFQ and ZFR turn out to be relatively interpretable one into the other. A particular fragment of the axiomatization of the theory ZFQ significantly allows to decide the continuum hypothesis $\aleph_1 = 2^{\aleph_0}$, proving the so-called "Freeling's Axiom", equivalent in ZFC to its denial. It must be noted, however, that the program of introducing new axioms for new primitive concepts, such as that of randomness, rather than axioms (such as those on large cardinals or on forcing) formulated in the pure language of set theory with the concepts of set and membership, has not found full approval and has even found fierce opponents.

There is furthermore the unavoidable issue of the relationship between algorithmic randomness and randomness in the sense of the physical sciences, either as *unpredictability* (deterministic chaos) or as ontic randomness, according to the standard interpretation of Quantum Mechanics. Gödel's aversion to generalisations of the significate of his results beyond the specifically logico-mathematical terrain is well known. There is in this respect the famous episode of John Wheeler who, having gone to Gödel's office to ask whether there was a connection between Heisenberg's uncertainty principle and the incompleteness theorem, was rather rudely kicked out (Wheeler comments with irony that this was mainly due to the bad influx exerted by Einstein on Gödel). However, the suggestive analogy between mathematical incompleteness and certain natural phenomena has been an irresistible attraction for many physicists. Incompleteness in classical, as well as quantum, physics became a popular topic for some years (see e.g. Da Costa and Doria (1991) or Moore (1990)). On the side of deterministic theories, dynamical systems are called *chaotic* if they are sensitive to initial conditions. In scientific thought between the 18th and 19th centuries, the deterministic

character of theories was revealed in their predictive capacity. Since Poincaré, however, a clearer distinction has begun to be made between the concepts of *determination* and of *prediction*: indeed, in a deterministic context, Poincaré proved that the system of equations describing the the interaction of three celestial bodies is provably incapable of predicting their evolution. In Bailly and Longo (2008), the authors consider Poincaré's three-body problem, which is at the origin of the theories of deterministic chaos, as a forerunner from an epistemological point of view of Gödel's limiting theorems: if the Laplacian conception can be compared in the logical-mathematical sphere, to the Hilbertian idea of a complete formal system, Poincaré's results can then be equated by analogy with Gödel's. A parallel is therefore instinctively established between the paradigmatic positions of Hilbert and Laplace, on the one hand, and Gödel and Poincaré on the other, that is, on the one hand the demand for decidability of every statement concerning the future (Laplacean predictability), while on the other, Poincaré, with his theorem on three-body problem, showed the "unpredictability of interesting non-linear systems, which we may understand as undecidability of future states".

In Bailly and Longo (2007) the authors wonder whether there is a randomness specific to the various fields of Physics. Typically in non-linear systems, randomness is of an "epistemic"nature, unpredictability is the joint result of sensitive dependency of the boundary conditions and the theoretical properties of classical measurement. Is it possible to describe the deterministic chaos by means of Chaitin complexity? This problem was raised in the 1990s of the last century and was part of a list of open problems. For a survey of progress in the applications of the theory of algorithmic randomness to the *Ergodic Theory* see for instance V'yugin (2022) and Towsner (2020). On the other hand, another branch, *Quantum Physics* proposes an intrinsic notion of randomness, as it is associated with *any* measurement operation. Challenging this "objective", "ontic", not epistemic concept of randomness, as early as 1935 Einstein, Podolsky and Rosen spoke of the "incompleteness" of *Quantum Mechanics*, formulating the famous EPR paradox. "Completeness" means in this framework that every element of the physical reality must have a counterpart in the physical theory. The paradox forces to conclude that the quantum-mechanical description of physical reality given by wave functions is not complete in this sense. The wave function does not provide in other words an exhaustive description of the objective properties of the system; this as is known led to the formulation of various hypotheses around hidden variables in order to produce a deterministic completion of the theory which would classically lead back to our ignorance its apparent probabilistic character. However, the remarkable results obtained by John Bell, as is well known, establish the impossibility of a local, deterministic theory predictively equivalent to *Quantum Mechanics* QM, i.e. no local theory admitting hidden variables is capable of reproducing its statistical predictions.

Starting from the Einstein, Podolsky and Rosen definition of "reality", the paradox stated in Peres (1985) highlights a case in which we know the (negative) answer to the question of whether or not the spin of a given electron is $3\hbar$, although there is no direct way of posing the question in the QM formalism and this too is reminiscent of the phenomenon of incompleteness. In Peres, Zurek (1982) it is emphasized that the theory of *Quantum Mechanics* cannot be considered "closed", in the sense that it can describe in principle anything, although in every situation something remains unanalyzed:

> This may not be just a flaw of quantum theory: it is likely to emerge as a logical necessity in any theory which is self- referential, as it attempts to describe its own means of verification. In this sense it is analogous to Gödel's undecidability theorem of formal number theory.

However, the reformulation of the incompleteness theorem *as a result about algorithmic randomness* has made it possible to investigate out of metaphor its meaning in fields other than Logic and has allowed a connection with disciplines such as Physics or Information Theory. There has been much speculation also on the relationship between indeterminacy in Heisenberg's sense and incompleteness in Gödel's sense (despite the latter's hostility). Beyond the strong suggestiveness of the topic, this issue has been thoroughly investigated on a

rigorously mathematical level in Calude and Stay (2007), that used for this purpose Chaitin's reformulation of the incompleteness theorem as a *trait-d'union* with Physics, presenting the Gödel-Chaitin theorem as a true uncertainty principle. Algorithmic randomness, as these two mathematicians show, is equivalent to what they call a "formal uncertainty principle", which in turn implies incompleteness.

The research programme that investigates the relationship between algorithmic randomness and quantum randomness is actually, in our opinion, one of the most exciting developments. Yurtsever (2000) showed that a sequence of bits produced by tossing a quantum coin is, almost certainly, algorithmically random (although the proof has been considered by some to be lacking in some points). More recently Nies and Scholz (2017) develop an algorithmic theory of randomness for infinite sequences of quantum bits and introduce a quantum analog of Martin-Löf tests, showing the existence of a universal such test in the framework of the C^*-algebra formulation of Quantum Mechanics. Actually at present there are different and non-equivalent approaches to this problem and we are still in an experimental phase of what appears to be a promising research program. The bibliography we have mentioned around algorithmic randomness and in particular its ralation with the notions of randomness that emerge from the various branches of Physics, is far from being exhaustive. The breadth and depth of the technical literature is enormous and we cannot account for it, nor is it the purpose of this book.