On the Independence Postulate, A Finitary Church-Turing Thesis

Samuel Epstein*

July 12, 2022

Abstract

The Independence Postulate (IP), is an unprovable inequality on the information content shared between two sequences. Among other applications, IP is a finitary Church Turing Thesis, postulating that certain infinite and finite sequences cannot be found in nature, a.k.a. have high "physical addresses". In this paper we show that IP explains why outliers are found in the physical world.

1 Introduction

In this section, we revisit the celebrated Church-Turing thesis (CT) and define the Independence Postulate (IP), introduced in [Lev84, Lev13]. CT relates mechanical methods to functions computed from Turing machines. For more information about Turing machines, we refer readers to [Sip13]. A method, M, for achieving some desired result is "effective" or "mechanical" if it can be carried out by a human with a pencil and paper. More formally,

1. M is set out in terms of a finite number of exact instructions (each instruction being expressed by means of a finite number of symbols).
2. M will, if carried out without error, produce the desired result in a finite number of steps.
3. M can (in practice or in principle) be carried out by a human being unaided by any machinery except paper and pencil.
4. M demands no insight, intuition, or ingenuity, on the part of the human being carrying out the method.

The Church-Turing thesis states

CT: *A method is effective if and only if it can be computed by a Turing machine.*

One well known variant of CT is the physical Church-Turing thesis, which states *all physically computable functions are Turing-computable*. However there are several drawbacks associated with CT. The notion of an "effective method" is vague, admitting multiple different interpretations. On such early assessment of this fact can be found in [Kle52],

> *Since our original notion of effective calculability of a function ... is a somewhat vague intuitive one, the thesis cannot be proved. ... While we cannot prove Church's thesis, since its role is to delimit precisely an hitherto vaguely conceived totality, we require evidence.*

Turing himself had reservations about his thesis, [Tur36]

*JP Theory Group. samepst@jptheorygroup.org

...fundamentally, appeals to intuition, and for this reason rather unsatisfactory mathematically.

IP is an unprovable inequality on the information measure of two sequences. Among other applications, IP is a finitary Church Turing Thesis, postulating that certain infinite and *finite* sequences cannot be found in nature, a.k.a. have high "physical addresses". IP provides a solution to the concerns of the somewhat vague formulation of CT. The statement of the IP is as follows [Lev13].

IP: *Let α be a sequence defined with an n-bit mathematical statement (e.g., in PA or set theory), and a sequence β can be located in the physical world with a k-bit instruction set (e.g., ip-address). Then $\mathbf{I}(\alpha : \beta) < k + n + c$ for some small absolute constant c.*

We take $\mathbf{I}$ to be the information term of Definition 2 in Section 2. Whereas IP is simpler, CT is more abstract. IP is supported by the so-called Independence Conservation Inequalities (Section 6). IP was succinctly described in a single page. In this paper, we expand upon the arguments in [Lev13] to make them accessible for a general audience. Section 3 details how IP is applied to finite sequences. Section 4 describes the applications of IP to logic. Section 5 details the surprising result that IP implies the existence of outliers in the physical world. IP is a statement in the field of algorithmic information theory (AIT), but no prior knowledge of AIT is required by the readers.

Contents

2 Complexity and Information

IP is a statement that uses complexity and information terms in Algorithmic Information Theory (AIT). Whereas classical information theory deals with in part the entropy of random variables, AIT deals with the entropy of individual sequences of 1's and 0's. This entropy measure, known as Kolmogorov complexity, is equal to the level of compressibility of finite and infinite sequences with respect to a reference Turing machine. Another notion of AIT is the amount of algorithmic mutual information of sequences, which represents their shared algorithmic information content. This information term is used in the statement of IP.

2.1 Self Delimiting Codes

When it is clear from the context, we will use whole numbers interchangeably with their binary representations. For example, each whole number $n \in \mathbb{W}$ can be associated with the $(n+1)$th sequence of a length increasing lexicographical ordering $\{\xi_n\}_{n=1}^{\infty}$. Each ξ_n is a finite sequence, with

$$(0,0),(1,1),(2,00),(3,01),(4,10),(5,11),(6,000)\dots$$

Thus $\xi_6 = 000$. A prefix free set of sequences S is a set of finite sequences such that there does not exist two distinct sequences x, y in S where one sequence is a prefix of the other. We say such S is a *self-delimiting* code because there exists a method to determine where each code word $x \in S$ ends without reading past its last symbol. Let $\|x\|$ be the length of the sequence x. One such self-delimiting code is $\langle x\rangle = 1^{\|x\|}0x$, where the decoding algorithm would first count the number of $1's$ before the first 0 to determine the length of x and then output the $\|x\|$ remaining bits in the input, (corresponding to x). Thus $\|\langle x\rangle\| = 2\|x\| + 1$.

2.2 Algorithms

The set of finite sequences is $\{0,1\}^*$. The set of infinite sequences is $\{0,1\}^{\infty}$. The set of finite and infinite sequences is $\{0,1\}^{*\infty}$. Our paper uses Turing machines M which have four tapes: a main input tape, an auxiliary input tape, a work tape, and an output tape. The alphabet for all tapes is $\{0,1,\$\}$. We give M a (partial) functional representation $M : \{0,1\}^* \times \{0,1\}^{*\infty} \to \{0,1\}^*$, defined by $y = M_\alpha(x)$ when

1. M starts with all its heads in the leftmost square. The main input tape starts with $x\$^{\infty}$. The auxiliary tape is set to α if it is an infinite sequence, otherwise it starts with $\alpha\$^{\infty}$. The work and output tape start with $\$^{\infty}$.

2. During its operation, M reads exactly $\|x\|$ bits from the main input tape.

3. The output tape is $y\$^{\infty}$ when M halts.

when this does not occur for inputs x and α then $M_\alpha(x) = \perp$ is undefined. The domain of such M is prefix free, where for all $x, y \in \{0,1\}^*$, $\alpha \in \{0,1\}^{*\infty}$, with $y \neq \emptyset$, it must be that $T_\alpha(x) = \perp$ or $T_\alpha(xy) = \perp$.

We use a fixed universal Turing machine U, where for each Turing machine T, there exists $t \in \{0,1\}^*$, where for all $x \in \{0,1\}^*$ and $\alpha \in \{0,1\}^{*\infty}$, $U_\alpha(tx) = T_\alpha(x)$. One example is for such t to be equal to $\langle i\rangle$, where i is the first index of T in an enumeration of Turing machines.

2.3 Big Oh Notation

As is typical of the field of algorithmic information theory, the inequalities in this paper are oftentimes relative to the universal machine U, and therefore their statements are only relative up to additive and logarithmic precision. For positive real functions f the terms in the follow table represent such inequalities. The Big Oh terms are only dependent on the choice of the universal machine U.

Inequality	Representation
$<^+ f$ and $>^+ f$	$<f+O(1)$ and $>f-O(1)$
$\overset{*}{<} f$ and $\overset{*}{>} f$	$< f/O(1)$ and $> f/O(1)$
$<^{\log} f$ and $>^{\log} f$	$<f+O(\log(f+1))$ and $>f-O(\log(f+1))$

2.4 Complexity and Information

The Kolmogorov complexity of sequence $x\in\{0,1\}^*$ relative to sequence $\alpha\in\{0,1\}^{*\infty}$ is defined to be the shortest U-program which produces x.

Definition 1 (Kolmogorov Complexity) $\mathbf{K}(x|\alpha) = \min\{\|p\| : U_\alpha(p) = x\}$.

$\mathbf{K}$ is a measure of the information content of a sequence. If a string x is random, $\mathbf{K}(x) \approx \|x\|$, in that there is no algorithmic means to compress it. Otherwise, if $\mathbf{K}(x) \ll \|x\|$, then x does not have much information content. For example, a string x consisting of 2^{100} consecutive pairs 01, can be compressed to a program of size

$$\mathbf{K}(x) <^+ \mathbf{K}(100) \ll \|x\| = 2^{100}.$$

Thus x is very much not random. For numbers $n \in \mathbb{N}$, their Kolmogorov complexity is of logarithmic magnitude, with $\mathbf{K}(n) = \mathbf{K}(\xi_n) = O(\log n)$. For sequences $x \in \{0,1\}^*$, their Kolmogorov complexity is bounded by $\mathbf{K}(x) < \|x\| + \mathbf{K}(\|x\|) + O(1)$, as there is a Turing machine that when given a program for a number n, computes this number n and then reads n bits on the input tape, and copies the contents to the output. The Kolmogorov complexity of a pair of strings $x, y \in \{0,1\}^*$, is $\mathbf{K}(x,y) = \mathbf{K}(\langle x\rangle y)$. We define information with respect to two finite or infinite sequences $\alpha, \beta \in \{0,1\}^{*\infty}$. This definition was first introduced in [Lev74], and is used in the statement of IP.

Definition 2 (Information) $\mathbf{I}(\alpha : \beta) = \log \sum_{x,y\in\{0,1\}^*} 2^{\mathbf{K}(x)+\mathbf{K}(y)-\mathbf{K}(x,y)-\mathbf{K}(x|\alpha)-\mathbf{K}(y|\beta)}$.

We derive the following inequality for use in Section 3. Let $z \in \{0,1\}^*$ be a finite string. Then

$$\begin{aligned}\mathbf{I}(z : z) &= \log \sum_{x,y\in\{0,1\}^*} 2^{\mathbf{K}(x)+\mathbf{K}(y)-\mathbf{K}(x,y)-\mathbf{K}(x|z)-\mathbf{K}(y|z)} \\ &> 2\mathbf{K}(z) - \mathbf{K}(z,z) - 2\mathbf{K}(z|z) \\ &>^+ \mathbf{K}(z),\end{aligned}$$

where the last inequality is up to a small constant dependent solely on the universal Turing machine U. If $\mathbf{I}(\alpha : \beta)$ is high, then the two sequences α and β share a lot of mutual algorithmic information. One example is an infinite sequence $\alpha \in \{0,1\}^\infty$ that is random, where there is some constant $c \in \mathbb{N}$

where for all prefixes $\alpha_n \in \{0,1\}^n$ of α of size n, we have that $n - c < \mathbf{K}(a_n)$. Then

$$
\begin{aligned}
2^{\mathbf{I}(\alpha:\alpha)} &> \sum_{n\in\mathbb{N}} 2^{2\mathbf{K}(\alpha_n)-\mathbf{K}(\alpha_n,\alpha_n)-2\mathbf{K}(\alpha_n|\alpha)} \\
&>^+ \sum_{n\in\mathbb{N}} 2^{2\mathbf{K}(\alpha_n)-\mathbf{K}(\alpha_n)-2\mathbf{K}(\alpha_n|\alpha)} \\
&>^+ \sum_{n\in\mathbb{N}} 2^{2\mathbf{K}(\alpha_n)-\mathbf{K}(\alpha_n)-2\mathbf{K}(n)} \\
&>^+ \sum_{n\in\mathbb{N}} 2^{\mathbf{K}(\alpha_n)-2\mathbf{K}(n)} \\
&>^+ \sum_{n\in\mathbb{N}} 2^{n-O(\log n)} \\
&= \infty.
\end{aligned}
$$

3 Non-Recursive Finite Sequences

One consequence of IP is a finite version of the Church-Turing Thesis (CT). This advantage was mentioned by L. A. Levin in [Lev13],

> *IP is simpler, CT more abstract. All sequences we ever see are computable just by being finite: CT is useless for them! IP works equally well for finite and infinite sequences.*

IP says that the only finite sequences that can be found in nature (i.e. have short physical addresses) will have non-recursive descriptions that are equal in length to their recursive descriptions. This can be seen when IP is applied to the case when $\alpha = \beta \in \{0,1\}^*$ is a finite sequence which has a non-recursive description of length $\mathbf{NR}(\alpha)$ that is much shorter than its recursive description $\mathbf{K}(\alpha)$, with $\mathbf{NR}(\alpha) \ll \mathbf{K}(\alpha)$. Let k be the shortest physical address of α. Then by IP, with $\beta = \alpha$,

$$
\begin{aligned}
\mathbf{K}(\alpha) <^+ \mathbf{I}(\alpha:\alpha) &<^+ k + \mathbf{NR}(\alpha) + c \\
\mathbf{K}(\alpha) - \mathbf{NR}(\alpha) - c &<^+ k. \qquad (1)
\end{aligned}
$$

Thus k is large and α cannot be easily located in the physical world. The only sequences α with short physical addresses must have $\mathbf{NR}(\alpha) \approx \mathbf{K}(\alpha)$. Thus if a sequence $x \in \{0,1\}^*$ is mathematical, with $\mathbf{NR}(x) \ll \|x\|$, then it must be algorithmic to be physically obtainable, that is, produced from a simple program, with $\mathbf{K}(x) \approx \mathbf{NR}(x) \ll \|x\|$. In general, not all sequences generated are algorithmic, take any typical outcome of the rolling of random dice.

3.1 Prefixes of the Halting Sequence

A canonical example of the inequality in Equation 1 is prefixes of the halting sequence. The halting sequence $\mathcal{H} \in \{0,1\}^\infty$ is the unique sequence defined by $\mathcal{H}[i] = 1$ iff $U(i)$ halts. Let $\mathcal{H}_n \in \{0,1\}^*$ be the finite sequence that is the prefix of size 2^n of $\mathcal{H}$. It is well known that

$$\mathbf{K}(\mathcal{H}_n) \in (n - O(1), n + \mathbf{K}(n) + O(1)).$$

The entire halting sequence $\mathcal{H}$ can be described in a mathematical statement of size equal to some small constant c_{HM}. Each n can be described using a program of size $\mathbf{K}(n)$. Therefore each

$\mathcal{H}_n$ can be defined by a mathematical statement of size $<^+ c_{\mathrm{HM}} + \mathbf{K}(n)$. So by IP applied to $\alpha = \beta = \mathcal{H}_n$ where k_n is the smallest physical address of $\mathcal{H}_n$,

$$n <^+ \mathbf{K}(\mathcal{H}_n) <^+ \mathbf{I}(\mathcal{H}_n : \mathcal{H}_n) <^+ c_{\mathrm{HM}} + \mathbf{K}(n) + c + k_n$$
$$n - \mathbf{K}(n) - c_{\mathrm{HM}} - c <^+ k_n.$$

Therefore the prefixes of $\mathcal{H}$ of size 2^n have physical address of size at least $n - O(\log n)$, and thus are not physically obtainable.

4 Logic

IP can also be used in instances where $\alpha \neq \beta$, and one canonical example is to logic, and in particular Peano Arithmetic (PA). PA is a logic system that encodes statements of arithmetic through a set of initial axioms and a deduction system. Gödel proved that PA is incomplete, in that there are well formed formulas in the language of PA which are true but are unprovable in PA. Suppose we order every well formed formula of PA and let the infinite sequence L be defined such that its ith bit is 1 iff the ith formula of PA is true. Then L is undecidable, in that there is no algorithm that can compute it. However Gödel himself thought that there can be other means to produce true axioms of mathematics [Gö61]:

> *Namely, it turns out that in the systematic establishment of the axioms of mathematics, new axioms, which do not follow by formal logic from those previously established, again and again become evident. It is not at all excluded by the negative results mentioned earlier that nevertheless every clearly posed mathematical yes-or-no question is solvable in this way. For it is just this becoming evident of more and more new axioms on the basis of the meaning of the primitive notions that a machine cannot imitate.*

However, as detailed in [Lev13], IP forbids such information leaks. The sequence L can be defined by a small mathematical formula of size n. Let β be any source of information with a reasonably small physical address of size k, such as the contents of an entire mathematical library. Then by IP, with $\alpha = L$, this information source will have negligible shared information with L (which encodes PA):

$$\mathbf{I}(\beta : L) < k + n + c.$$

More generally, in [Lev13], it was shown that every consistent completion β of PA has $\mathbf{I}(\beta : \mathcal{H}) = \infty$. Then by IP, since $\mathcal{H}$ is represented by a formula of size c_{HM}, no consistent completion β has a finite physical address. A generalization of this result can be found in [BP16].

5 Outliers

An outlier is an observation whose value lies outside the set of values considered likely according to some hypothesis (usually one based on other observations); an isolated point. It is an observation that lies an abnormal distance from other values in a random sample from a population. Usually, the presence of an outlier indicates some sort of problem. In this section, we detail the surprising relationship between IP and outliers:

Assuming the Independence Postulate, outliers are guaranteed to occur.

5.1 Deficiency of Randomness

In the realm of algorithmic information theory, outliers are modeled using the deficiency of randomness. For a probability measure over numbers $p : \mathbb{N} \to \mathbb{R}_{\geq 0}$, the deficiency of randomness of $x \in \{0,1\}^*$ with respect to p is defined as the difference between the p-code of x and the length of x's shortest description $\mathbf{d}(x|p) = -\log p(x) - \mathbf{K}(x)$. It was introduced in [Lev74, Lev84]. the deficiency of randomness of a string is large then there is a description of x that is much shorter than the p-code of size $-\log p(x)$. The deficiency of randomness measures the extent of the refutation against the hypothesis p given the result x. The more regularities discovered in the string x, the greater the deficiency of randomness $\mathbf{d}$. Strings that are optimally encoded by using Shannon-Fano p codes, will have low $\mathbf{d}$-scores, and thus be p-typical.

5.1.1 Universal Test

The deficiency of randomness has the remarkable property of being a universal test. A discrete test t over a probability p over numbers $\mathbb{N}$, has the property that for all $a \in \mathbb{N}$, $t(a) \geq 0$, and $\sum_{a\in\mathbb{N}} p(a)t(a) \leq 1$. For a set S of tests, we say that a test $t_0 \in S$ is universal for S, iff for each test $t \in S$, there exists a constant $c_t > 0$ such that $t_0(a) > c_t t(a)$, for all $a \in \mathbb{N}$. A universal test, captures in a manner, the contribution of each test in the set S.

A function $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ is lower computable if and only if there exists a computable function $f' : \mathbb{N} \times \mathbb{N} \to \mathbb{Q}$, such that for all $a, n \in \mathbb{N}$, $f'(a,n) \leq f'(a, n+1)$, and $\lim_{n\to\infty} f'(a,n) = f(a)$. Lower computable functions are halfway between being a function that can't be computed by an algorithm, and a function that can be. They represent a very useful definition in the field of Algorithmic Information Theory, because there are many classes of lower-computable functions which have universal elements.

The deficiency of randomness $2^{\mathbf{d}(a|p)}$ is a universal test over the set of all lower computable p tests. First, this statement implies $2^{\mathbf{d}(\cdot|p)}$ is a p-test. Indeed, it can be proven that $\sum_{a\in\mathbb{N}} 2^{\mathbf{d}(a|p)} p(a) \leq 1$. $\mathbf{d}(a|p) = -\log p(a) - \mathbf{K}(a|p)$ is lower computable, because $\mathbf{K}$ is upper computable[1]. Furthermore for any lower computable p-test t, we have the inequality, for all $a \in \mathbb{N}$,

$$t(a) \overset{*}{<} 2^{\mathbf{d}(a|p)+\mathbf{K}(t)+\mathbf{K}(p)},$$

where $\mathbf{K}(t)$ is the size of the smallest program which lower computes t. The proof for this statement can be found in [G21]. Thus $\mathbf{d}$ dominates all lower computable tests, and the simplier the test, the more that $\mathbf{d}$ dominates it. Thus contribution of simple tests is valued more over complex tests. This is aligned with the general principle of Occam's razor, where simple explanations are valued over complex explanations. Thus the simple function $\mathbf{d}(\cdot|p)$ is a weighted contribution of all lower computable p-tests, each weighted by their complexity.

5.2 Outliers and IP

This section shows that IP implies observable sequences in the physical world will have emergent outliers. For $\tau \in \mathbb{N}^{\mathbb{N}}$, lets its encoding into a sequence be $\{0,1\}^\infty \ni \langle\tau\rangle = \langle\tau[1]\rangle\langle\tau[2]\rangle\langle\tau[3]\rangle\ldots$ Let $\tau(n)$ be the first 2^n unique numbers found in τ. The sequence τ is assumed to have an infinite amount of unique numbers, and represents a series of observations. Each observation $\tau[i] \in \mathbb{N}$ is encoded by a number.

[1] Kolmogorov complexity $\mathbf{K}(a)$ is upper computable because there is a program that can try running all programs simultaneously and see which programs output a, and keep outputting the length of the smallest program. This output is decreasing and its limit will be $\mathbf{K}(a)$, because eventually the smallest program will be found.

Definition 3 *Let* $\mathrm{score}(\tau, p) \in \mathbb{N} \cup \infty$ *be the smallest number such that for all* $n \in \mathbb{N}$*, we have* $\mathrm{score}(\tau, p) >^{+} n - \max_{a \in \tau(n)} \mathbf{d}(a|p) - 4\mathbf{K}(n)$*.*

The number $\mathrm{score}(\tau, p)$ represents the level of outliers found in the observations of τ. If $\mathbf{d}(\tau[i]|p) < c$ for all i, then the observations of τ are bounded in their outlier scores and $\mathrm{score}(\tau, p) = \infty$. If τ contains ever increasing outliers at a rapid rate, then $\mathrm{score}(\tau, p)$ is very small. By [Eps22], we get the following relationship between $\mathrm{score}(\tau, p)$ and $\mathcal{H}$.

Theorem 1 ([Eps22]) $\mathrm{score}(\tau, p) <^{\log} \mathbf{I}(\langle \tau \rangle : \mathcal{H}) + \mathbf{K}(p)$.

The term $\langle \tau \rangle$ is an encoding of τ into an infinite sequence. Let k be the physical address of an infinite sequence of numbers τ. As defined above, the halting sequence $\mathcal{H}$ can be described in a mathematical statement of size equal to some small constant c_{HM}. Then by Theorem 1, for some sequence $\tau \in \mathbb{N}^{\mathbb{N}}$ and computable probability p over $\mathbb{N}$, the following statement is derived from IP, with

Statement. $\mathrm{score}(\tau, p) <^{\log} \mathbf{I}(\langle \tau \rangle : \mathcal{H}) + \mathbf{K}(p) <^{\log} k + c_{\mathrm{HM}} + c + \mathbf{K}(p)$.

Thus sequences τ with large $\mathrm{score}(\tau, p)$, as defined in Definition 3, will have a large physical address. Thus it is hard to find physical sequences which do not have large outliers, and completely impossible to find sequences with no outliers. As the complexity of the probability p in the randomness deficiency term increases, the bounds loosen. One can also use IP to characterize sequences of reals.

6 Information Conservation Inequalities

IP is an upper bound on the information between two sequences. The set of all true statements in arithmetic has no information about the stock market. Or the halting problem has nothing to say about any easily accessible series of physical of measurements.

This leaves open the possibility of deterministic or randomized processing to increase the amount of information that the sequences have. For example, one such method is to select statements of arithmetic with probability .5, in hopes of gleaning information about the next stock market crash. However the door to such circumventions is closed due to Independence Conservatism Inequalities (ICI) [Lev74, Lev84], which complements IP.

Whereas IP is an unprovable postulate, ICI are provable statements in the field of algorithmic information theory that says target information cannot be increased. The origins of ICI are in data processing inequalities in classical information theory, detailed in [CT91]. For two random variables $\mathcal{X}$ and $\mathcal{Y}$,

$$\mathcal{I}(\mathcal{X} : T(\mathcal{Y})) \leq \mathcal{I}(\mathcal{X} : \mathcal{Y}).$$

The term $\mathcal{I}$ is the classical information measure between two variables and T is any local processing done on the random variable $\mathcal{Y}$. Theorems in classical information theory often have equivalents in algorithmic information theory, and this is the case for the data processing inequality. The ICI for deterministic processing is

$$\text{For sequences } \alpha \text{ and } \beta, \text{ computable function } f,\ \mathbf{I}(f(\alpha) : \beta) <^{+} \mathbf{I}(\alpha : \beta) + \mathbf{K}(f).$$

A function can add some mutual information between two sequences, but no more than the complexity of the function. There is also a randomized ICI. There are several forms, with the following inequality being one such instance [Lev74, Gei12, Ver21]. Let f be a function that transforms a sequence β using a random seed ω. Let p be a computable probability over infinite sequences.

$$\mathbf{E}_{\omega\sim p}[\mathbf{I}(f(\beta,\omega):\alpha)] <^+ \mathbf{I}(\beta:\alpha)+\mathbf{K}(f,p).$$

Thus ICI prevents the processing of data to gain more target. As L. A. Levin states,

torturing an uniformed witness cannot give information about the crime.

Another application of ICI discussed in [Lev13] is to processes, represented by infinite sequences. A process ω_1 is "explained" by a simpler process ω_2 if there is some computable function f_1 such that $f_1(\omega_2)=\omega_1$. In CS, we say ω_1 reduces to ω_2. Say $\mathbf{I}(\omega_1:\beta)=\infty$, where β is specified by a finite mathematical sequence, and ω_1 is a process. Assume ω_1 is a complicated process, and can be explained by a series of reductions to simpler process

$$\omega_1 \leftarrow_{f_1} \omega_2 \leftarrow_{f_2} \omega_3 \cdots \leftarrow_{f_{n-1}} \omega_{f_n}.$$

Assume ω_n is a "simple" process, admitting no further meaningful reduction. This notion of a simple, unexplainable process is a subjective one, as process can always be reduced to another one. Thus by ICI, $\mathbf{I}(\omega_n:\beta)=\infty$, and by IP, ω_n cannot be found in nature, as all its physical addresses have infinite length. Thus to recap, using ICI, complicated processes with unlimited target information can be reduced to unexplainable simple processes and by using IP can be shown to not exist in nature.

References

[BP16] L. Bienvenu and C. Porter. Deep π_1^0 classes. *The Bulletin of Symbolic Logic*, 22(2):249–286, 2016.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York, NY, USA, 1991.

[Eps22] Samuel Epstein. Outliers, dynamics, and the independence postulate. *Corr*, abs/2207.03955, 2022.

[G̈61] Kurt Gödel. The modern development of the foundations of mathematics in the light of philosophy. *In: Kurt Gödel. Collected Works. Volume III. Oxford University Press.*, 1961.

[Ǵ21] Peter Gács. Lecture notes on descriptional complexity and randomness. *CoRR*, abs/2105.04704, 2021.

[Gei12] Philipp Geiger. *Mutual information and Gödel incompleteness.* PhD thesis, Heidelberg University, 10 2012.

[Kle52] S. C. Kleene. *Introduction to Metamathematics.*, page 318. North-Holland Publishing., Amsterdam, 1952.

[Lev74] L. A. Levin. Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory. *Problemy Peredachi Informatsii*, 10(3):206–210, 1974.

[Lev84] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Lev13] L. A. Levin. Forbidden information. *J. ACM*, 60(2), 2013.

[Sip13] M. Sipser. *Introduction to the Theory of Computation*. Course Technology, Boston, MA, third edition, 2013.

[Tur36] M. Turing, A. On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42):230–265, 1936.

[Ver21] N. Vereshchagin. Proofs of conservation inequalities for levin's notion of mutual information of 1974. *Theoretical Computer Science*, 856, 2021.