

Research Article

Weighing Theories: on Chaitin's Heuristic Principle

Saeed Salehi¹

1. Plaksha University, India

It would be a heavenly reward if there were a method of weighing theories and sentences in such a way that a theory could never prove a heavier sentence (Chaitin's Heuristic Principle). Alas, no satisfactory measure has been found so far, and this dream seemed too good to ever come true. Here, we attempt to revive Chaitin's lost paradise of heuristic principle as much as logic allows.

Corresponding author: Saeed Salehi, root@SaeedSalehi.ir

1. Introduction and Preliminaries

The history of the heuristic principle goes back to at least 1974, when Chaitin wrote in [1] that “there are circumstances in which [...] it is possible to reason in the following manner. If a set of theorems constitutes t bits of information, and a set of axioms contains less than t bits of information, then it is impossible to deduce these theorems from these axioms” (p. 404). Let us put this formally as

Definition 1.1 (Heuristic Principle, HP). *A real-valued mapping $\mathcal{W}$ on theories and sentences (over a fixed language) is said to satisfy Chaitin's heuristic principle, HP, when for every theory T and every sentence ψ , if we have $\mathcal{W}(\psi) > \mathcal{W}(T)$ then T cannot prove ψ ; in symbols*

$$(HP) \quad \mathcal{W}(\psi) > \mathcal{W}(T) \implies T \not\vdash \psi. \diamond$$

An equivalent formulation is “the theorems weigh no more than the theory”: $T \vdash \psi \implies \mathcal{W}(T) \geq \mathcal{W}(\psi)$. The principle was suggested after Chaitin's (proof of Gödel's first) incompleteness theorem, which says that for every sufficiently strong and consistent theory T , there exists a constant c such that for no w , T can prove that “the Kolmogorov complexity of w is greater than c ” (below, we will see a formal definition of Kolmogorov complexity). It was criticized by several authors, probably the first time in 1989 by van Lambalgen, who wrote in [2] that “Chaitin's views enjoy a certain popularity” (p. 1389) and that

“Chaitin’s mathematics does not support his philosophical conclusions” (p. 1390). He concluded “that the *complexity* of the axioms is not a good measure of information” (p. 1395; original italics).

Chaitin retreated, taking just one step back and not completely, in one of his 1992 papers ^[3]: “In fact, any set of axioms that yields an infinite set of theorems *must* yield theorems with arbitrarily high complexity! [...] So what is to become of our heuristic principle that ‘A set of axioms of complexity N cannot yield a theorem of complexity substantially greater than N ’??? An improved version of this heuristic principle, which is not really any less powerful than the original one, is this: ‘One cannot prove a theorem from a set of axioms that is of greater complexity than the axioms *and know* that one has done this. I.e., one *cannot realize* that a theorem is of substantially greater complexity than the axioms from which it has been deduced, if this should happen to be the case” (p. 115, original italics). He then tries “to avoid all these problems and discussions by rephrasing [his] fundamental principle in the following totally unobjectionable form: ‘A set of axioms of complexity N cannot yield a theorem that asserts that a specific object is of complexity substantially greater than N .’ It was removing the words ‘asserts that a specific object’ that yielded the slightly overly-simplified version of the principle” (p. 116).

The fact of the matter is that Chaitin’s last statement, that “a set of axioms of complexity N cannot yield a theorem that asserts that a specific object is of complexity substantially greater than N ”, is nothing but a reformulation of his incompleteness theorem, and the mildly diluted statement, that “one cannot prove a theorem from a set of axioms that is of greater complexity than the axioms *and know* that [that] theorem is of substantially greater complexity than the axioms from which it has been deduced”, follows from his incompleteness theorem and is indeed far from the original HP.

The criticism continued, for good reasons. In 1996, Fallis ^[4] noted that “for any sound formal system FS , there are infinitely many formulas which have greater complexity than FS and are provable in FS . [...] Since only finitely many of the (infinitely many) formulas provable in FS have a complexity less than the complexity of FS , there are infinitely many formulas provable in FS with greater complexity than FS ” (p. 265). As a result, Chaitin’s claim (that “if a theorem contains more information than a given set of axioms, then it is impossible for the theorem to be derived from the axioms” ^[4]) is false.

In 1998, Raatikainen, maybe unaware of ^[4] as he did not cite it, wrote in ^[5] that “Chaitin’s metaphor that ‘if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms’, if referring to Chaitin’s theorem, seems to commit [the] confusion [that] it compares

the complexity of axioms as *mentioned* and the complexity asserted by a theorem when *used*” (p. 581; emphasis added).

Anyhow, HP was too beautiful a dream to let go easily. In 2004, Sjögren designed, in his licentiate thesis [6], “a measure of the power” of theories and sentences that satisfy HP, where theories extended Peano’s Arithmetic and the sentences belonged to a rather narrow class of arithmetical translations of “the letterless modal sentences” of Gödel-Löb logic. A report of the results was presented later in a 2008 paper [7].

In 2005, Calude and Jürgensen claimed in [8] that they “prove that the ‘heuristic principle’ proposed by Chaitin [...] is correct if we measure the complexity of a string by the difference between the program-size complexity and the length of the string, [their] δ -complexity” (pp. 3–4). The δ -complexity was defined as $\delta(x) = \mathcal{K}(x) - |x|$, where $|x|$ denotes the length of x and $\mathcal{K}(x)$, the Kolmogorov complexity of x , is the length of the shortest input-free program (in a fixed programming language) that outputs only x (and then halts). The main result of [8] reads: “Consider a finitely-specified, arithmetically sound (i.e. each arithmetical proven sentence is true), consistent theory strong enough to formalize arithmetic, and denote by T its set of theorems [...] Let g be a Gödel numbering for T . Then, there exists a constant N [...] such that T contains no x with $\delta_g(x) > N$ ” (p. 9, Theorem 4.6). This claim was praised by, e.g., Grenet (in 2010), who wrote in [9] that “In [2] Chaitin’s heuristic principle is proved to be valid for an appropriate measure [that] gives us some indication about the reasons certain statements are unprovable” (p. 404), and that his study led him “to modify the definition of δ_g in order to correct some of the proofs” (p. 423).

Unfortunately, δ does not satisfy HP, as can be seen by the following argument. Let $\perp$ denote a contradiction, like $p \wedge \neg p$ or $\exists x(x \neq x)$. There are two fixed, and rather small, natural numbers m and n , such that for every formula φ we have $|\perp \rightarrow \varphi| = |\varphi| + m$ and $\mathcal{K}(\varphi) \leq \mathcal{K}(\perp \rightarrow \varphi) + n$. For the latter, notice that one can make some small changes to the shortest input-free program that outputs only $\perp \rightarrow \varphi$, to get an input-free program, not necessarily with the shortest length, that outputs only φ ; and those changes are uniform and do not depend on φ .¹ Now, fix an arbitrary theory T and assume that $\delta(T) = t$; one could take t to be the constant N in the above quoted Theorem 4.6 of [8]. Also, fix a sentence $\mathfrak{S}$ with $\delta(\mathfrak{S}) > t + m + n$ (which should exist by [8]). We have

$\delta(\perp \rightarrow \mathfrak{S})$	=	$\mathcal{K}(\perp \rightarrow \mathfrak{S})$ - $ \perp \rightarrow \mathfrak{S} $		by the definition of δ ,
	$\geq$	$\mathcal{K}(\mathfrak{S})$ - n - $(\mathfrak{S})$ + m		by the choice of m, n ,
	=	$\delta(\mathfrak{S})$ - (m) + n		by the definition of δ ,
	>	t		by the choice of $\mathfrak{S}$,
	=	$\delta(T)$		by the definition of t .

So, by [8], quoted above, $T \not\vdash (\perp \rightarrow \mathfrak{S})$ since $\delta(\perp \rightarrow \mathfrak{S}) > \delta(T)$; but $\perp \rightarrow \mathfrak{S}$, for every sentence $\mathfrak{S}$, is a tautology, and so should be provable in every theory. One can give a similar argument by using the formula $\mathfrak{S} \rightarrow \top$, where $\top$ denotes a tautology, like $p \vee \neg p$ or $\forall x(x = x)$; or by using the tautologies $p \rightarrow (\mathfrak{S} \rightarrow p)$ or $\neg p \rightarrow (p \rightarrow \mathfrak{S})$ or $[(p \rightarrow \mathfrak{S}) \rightarrow p] \rightarrow p$, for a fixed, short, and uncomplicated formula p .² Summing up, the following holds according to the cited results and our argument above:

Proposition 1.2 (HP does not hold so far). *Neither the Kolmogorov-complexity (the length of the shortest input-free program that outputs only the axioms, or the theorems, of the theory) nor the δ -complexity (the difference between the Kolmogorov-complexity and the length) satisfies HP. $\square$*

This is not the end of the story. In 2021, Porter [10] hoped to show “a possible vindication of Chaitin’s interpretation [HP], drawing upon the [...] recent work of Bienvenu et al. [11] that extends and refines [Chaitin’s incompleteness theorem]” (p. 149). He was content with a weak version of HP, as his “statement is significantly weaker than being a full-fledged instance of Chaitin’s heuristic principle” (p. 160).

In the rest of this paper, we suggest some ways of weighing theories and sentences that satisfy HP.

2. Weighing Theories

We can work with *theories* only and dismiss the *sentences*; it suffices to take the weight of a sentence ψ to be the weight of the theory $\{\psi\}$. For us, a *theory* is an arbitrary *set of sentences* (we will consider *re* theories later in §2.4.1). The most trivial mapping that satisfies HP is the constant weight (all the

theories have a fixed weight): since $\mathcal{W}(U) > \mathcal{W}(T)$ can never hold for any theories T and U , then “ $\mathcal{W}(U) > \mathcal{W}(T) \Rightarrow T \not\models U$ ” is vacuously true. In the sequel, we note that HP can be satisfied by some other, less trivial weights.

2.1. Finitely Many Weights

Let ν be a propositional evaluation from formulas to $\{0, 1\}$, where 0 indicates the falsum and 1 the truth. If $P \rightarrow Q$ holds, then we have $\nu(P) \leq \nu(Q)$. This suggests the following:

Definition 2.1 ($\mathcal{W}_V, \mathcal{W}_{\mathfrak{M}}$). Let ν be a mapping from propositional atoms to $\{\text{false}, \text{true}\}$, and let V be its truth-table extension to all the propositional formulas. For a formula φ , let $V \models \varphi$ mean that $V(\varphi) = \text{true}$; and for a theory T let $V \models T$ mean that $V \models \tau$ holds for each and every element (axiom) τ of T . Let $\mathcal{W}_V$ be the following mapping, where T is a propositional theory.

$$\mathcal{W}_V(T) = \begin{cases} 0 & \text{if } V \models T, \\ 1 & \text{if } V \not\models T. \end{cases}$$

Likewise, for a fixed first-order structure $\mathfrak{M}$, let $\mathcal{W}_{\mathfrak{M}}$ be the following mapping:

$$\mathcal{W}_{\mathfrak{M}}(T) = \begin{cases} 0 & \text{if } \mathfrak{M} \models T, \\ 1 & \text{if } \mathfrak{M} \not\models T; \end{cases}$$

where T is a first-order theory. $\diamond$

Theorem 2.2 ($\mathcal{W}_V, \mathcal{W}_{\mathfrak{M}}$ satisfy HP). For every evaluation V and every structure $\mathfrak{M}$, $\mathcal{W}_V$ and $\mathcal{W}_{\mathfrak{M}}$ satisfy HP.

Proof. If $\mathcal{W}_{\mathfrak{M}}(U) > \mathcal{W}_{\mathfrak{M}}(T)$, then $\mathcal{W}_{\mathfrak{M}}(U) = 1$ and $\mathcal{W}_{\mathfrak{M}}(T) = 0$, so $\mathfrak{M} \models T$ but $\mathfrak{M} \not\models U$, therefore $T \not\models U$. $\square$

Let us recall that no structure can satisfy an *inconsistent* theory T (for which we write $T \vdash \perp$); and every structure satisfies a *tautological* theory T (written as $\top \vdash T$). If we replace “ $\mathfrak{M} \models$ ” with “ $\top \vdash$ ” in Definition 2.1, then we get the following weighing $\mathcal{W}_{\top}$; the other weighing $\mathcal{W}^{\perp}$ is its dual.

Definition 2.3 ($\mathcal{W}_{\top}, \mathcal{W}^{\perp}$). Let $\mathcal{W}_{\top}$ and $\mathcal{W}^{\perp}$ be the following mappings,

$$\mathcal{W}_{\top}(T) = \begin{cases} 0 & \text{if } \top \vdash T, \\ 1 & \text{if } \top \not\vdash T; \end{cases} \quad \text{and} \quad \mathcal{W}^{\perp}(T) = \begin{cases} 0 & \text{if } T \not\vdash \perp, \\ 1 & \text{if } T \vdash \perp; \end{cases}$$

for a theory T . $\diamond$

The weighing $\mathcal{W}^{\perp}$ is the so-called “drastic inconsistency measure”, introduced by Hunter and Konieczny in 2008; see e.g. [12]. It is easy to see that both the mappings $\mathcal{W}_{\top}$ and $\mathcal{W}^{\perp}$ satisfy HP (see Theorem 2.5 below). In fact, $\top$ and $\perp$ play no special roles in $\mathcal{W}_{\top}$ or $\mathcal{W}^{\perp}$.

Definition 2.4 ($\mathcal{W}_{\nabla}, \mathcal{W}^{\nabla}$). For a fixed theory ∇ , let $\mathcal{W}_{\nabla}$ and $\mathcal{W}^{\nabla}$ be the following mappings,

$$\mathscr{W}_{\mathbb{V}}(T) = \begin{cases} 0 & \text{if } \mathbb{V} \vdash T, \\ 1 & \text{if } \mathbb{V} \not\vdash T; \end{cases} \quad \text{and} \quad \mathscr{W}^{\mathbb{V}}(T) = \begin{cases} 0 & \text{if } T \not\vdash \mathbb{V}, \\ 1 & \text{if } T \vdash \mathbb{V}; \end{cases}$$

where T is a theory. $\diamond$

Below, we will show that both $\mathscr{W}_{\mathbb{V}}$ and $\mathscr{W}^{\mathbb{V}}$ satisfy HP. Let us note that

- (a) If $\mathbb{V}$ is tautological, then $\mathscr{W}_{\mathbb{V}}$ is $\mathscr{W}_{\top}$ in Definition 2.3, and $\mathscr{W}^{\mathbb{V}}$ is the constant weighing 1; and
- (b) If $\mathbb{V}$ is inconsistent, then $\mathscr{W}_{\mathbb{V}}$ is the constant weighing 0, and $\mathscr{W}^{\mathbb{V}}$ is $\mathscr{W}^{\perp}$ in Definition 2.3.

Theorem 2.5 ($\mathscr{W}_{\mathbb{V}}, \mathscr{W}^{\mathbb{V}}$ satisfy HP).

For a fixed theory $\mathbb{V}$, both $\mathscr{W}_{\mathbb{V}}$ and $\mathscr{W}^{\mathbb{V}}$ satisfy HP.

Proof. If $\mathscr{W}^{\mathbb{V}}(U) > \mathscr{W}^{\mathbb{V}}(T)$, then $\mathscr{W}^{\mathbb{V}}(U) = 1$ and $\mathscr{W}^{\mathbb{V}}(T) = 0$, so $U \vdash \mathbb{V}$ but $T \not\vdash \mathbb{V}$, therefore $T \not\vdash U$.

The case of $\mathscr{W}_{\mathbb{V}}$ is very similar to the proof of Theorem 2.2 (just replace “ $\mathfrak{M} \models$ ” with “ $\mathbb{V} \vdash$ ”). $\square$

One main tool in the proofs was the transitivity of the deduction relation: if $S \vdash T \vdash U$, then $S \vdash U$.

There are some other HP-satisfying mappings that have more than two values. Let us skip the proof of the following proposition, which could be an interesting exercise in elementary logic.

Proposition 2.6 (Some HP-satisfying weightings with more than two values). *The following mapping, for a theory T , satisfies HP.*

$$T \mapsto \begin{cases} 0 & \text{if } \top \vdash T \text{ (i.e., if } T \text{ is tautological);} \\ 1 & \text{if } \top \not\vdash T \not\vdash \perp \text{ (i.e., if } T \text{ is non-tautological and consistent);} \\ 2 & \text{if } T \vdash \perp \text{ (i.e., if } T \text{ is inconsistent).} \end{cases}$$

Fix a consistent and non-tautological theory $\mathbb{V}$ (that is, $\top \not\vdash \mathbb{V} \not\vdash \perp$). The following mappings, for a theory T , satisfy HP.

$$T \mapsto \begin{cases} 0 & \text{if } T \not\vdash \mathbb{V} \vdash T; \\ 1 & \text{if } T \vdash \mathbb{V} \vdash T \text{ or } T \not\vdash \mathbb{V} \not\vdash T; \\ 2 & \text{if } T \vdash \mathbb{V} \not\vdash T \text{ and } T \not\vdash \perp; \\ 3 & \text{if } T \vdash \perp. \end{cases} \quad \text{and} \quad T \mapsto \begin{cases} 0 & \text{if } \top \vdash T; \\ 1 & \text{if } \top \not\vdash T \text{ and } T \not\vdash \mathbb{V} \vdash T; \\ 2 & \text{if } T \vdash \mathbb{V} \vdash T \text{ or } T \not\vdash \mathbb{V} \not\vdash T; \\ 3 & \text{if } T \vdash \mathbb{V} \not\vdash T \text{ and } T \not\vdash \perp; \\ 4 & \text{if } T \vdash \perp. \end{cases}$$

$\square$

2.2. HP-Satisfying Weights

HP forces the existence of a minimum and a maximum.

Theorem 2.7 ($\mathscr{W}$ has a min and a max if it satisfies HP). *Every weighting $\mathscr{W}$ that satisfies HP has a minimum and a maximum.*

Proof. For a theory T , we have $\perp \vdash T \vdash \top$, where we recall that $\perp$ is a contradiction and $\top$ is a tautology. Now, by HP, we have $\mathcal{W}(\perp) \geq \mathcal{W}(T) \geq \mathcal{W}(\top)$. Thus, $\min_{\mathcal{W}} = \mathcal{W}(\top)$ and $\max_{\mathcal{W}} = \mathcal{W}(\perp)$. $\square$

Corollary 2.8 (integer-valued HP-satisfying weights are finitely many valued). *If an integer-valued weighing, like the Kolmogorov complexity $\mathcal{K}$ or its difference with the length (the δ complexity), satisfies HP, then it can take finitely many values only. $\square$*

We will see in Theorem 2.20 below that there are plenty of finitely many valued weightings that satisfy HP. We already noted in Proposition 1.2 that neither $\mathcal{K}$ nor δ satisfies HP. Now, we notice a couple of properties of arithmetical theories that extend Robinson's Arithmetic Q , which imply that finitely many weights are not suitable for these theories, even if they satisfy HP.

Theorem 2.9 (Arithmetical theories form an infinite and a pseudo-dense hierarchy).

1. *There is an infinite hierarchy of finitely axiomatizable arithmetical theories containing Q with strictly increasing proof power.*
2. *Between every two finite extensions of Q , one of which is a strict sub-theory of the other, there exists another finite theory that sits strictly between the two. In other words, if $Q \subseteq S \subsetneq T$ for two finite theories S and T , then there is a finite theory U such that $S \subsetneq U \subsetneq T$.*

Proof.

(1): For a consistent theory T that contains Q , by Gödel-Rosser's Incompleteness Theorem, there exists a sentence ρ , called a Rosser Sentence of T , such that T can neither prove nor disprove it. Therefore, both theories $T_0 = T + \rho$ and $T_1 = T + \neg\rho$ are consistent. Now, T_0 and T_1 strictly contain T (and Q), and they are both finite if T is so. Continuing this way, one can show the existence of an infinite (strictly increasing in power) hierarchy of finite arithmetical theories.

(2): Let us identify a finitely axiomatizable theory with a single sentence that axiomatizes it (which can be the conjunction of the finitely many axioms of the theory). Let S and T be two sentences such that they both contain Q (i.e., $S, T \vdash Q$) and S is a strict subtheory of T (i.e., $T \vdash S$ and $S \not\vdash T$). So, the theory $W = S + \neg T$ is consistent; it also contains Q . By Gödel-Rosser's Incompleteness Theorem, there exists a sentence ρ independent from W . So, $W \not\vdash \rho$ and $W \not\vdash \neg\rho$. Let $U = T \vee (S \wedge \rho)$. This sentence contains Q , as $U \vdash Q$ follows from $T \vdash Q$ and $S \vdash Q$ (with $S \wedge \rho \vdash S$). Obviously, U lies between S and T , since $T \vdash U$ is clear from the tautology $T \rightarrow T \vee X$, and $U \vdash S$ follows from the assumption $T \vdash S$. We show that U strictly lies between S and T , i.e., $U \not\vdash T$ and $S \not\vdash U$. The former follows from $W \not\vdash \neg\rho$, which implies $S \wedge \neg T \not\vdash \neg\rho$, or equivalently $S \wedge \rho \not\vdash T$, thus $U \not\vdash T$. The latter follows from $W \not\vdash \rho$, which

implies $S \wedge \neg T \not\vdash \rho$, thus $S \not\vdash T \vee \rho$, so $S \not\vdash U$. Essentially the same argument shows that $U' = T \vee (S \wedge \neg \rho)$ too strictly lies between S and T . It can also be shown that U and U' are incomparable with each other, i.e., $U \not\vdash U'$ and $U' \not\vdash U$. $\square$

Consider the finite theory Q , and let ρ be a Rosser sentence of it. So, $Q \subsetneq Q + \rho$. By Theorem 2.9, there are finite and consistent theories $\{T_m\}_m$, where m ranges over integer numbers, such that

$$Q \subsetneq \cdots \subsetneq T_{-2} \subsetneq T_{-1} \subsetneq T_0 \subsetneq T_1 \subsetneq T_2 \subsetneq \cdots \subsetneq Q + \rho.$$

So, an HP-satisfying weight should contain a decreasing and an increasing sequence of values. Thus, as a corollary of Corollary 2.8, we get the following.

Corollary 2.10 (integer-valued HP-satisfying weights for arithmetical theories). *If an integer-valued weighing, like the Kolmogorov complexity $\mathcal{K}$ or its difference with the length (the δ complexity), satisfies HP, then it will have a fixed value for infinitely many distinct arithmetical theories. $\square$*

2.3. The Equivalence Principle

The converse of HP, that is,

$$\text{HP}^{-1} : \quad T \not\vdash U \Rightarrow \mathcal{W}(U) > \mathcal{W}(T), \text{ for theories } T \text{ and } U,$$

cannot hold for real-valued weightings (see also [2]). The reason is that, firstly, HP^{-1} is equivalent to $\mathcal{W}(U) \leq \mathcal{W}(T) \Rightarrow T \vdash U$, and, secondly, there are *incomparable* theories (neither of which can prove the other). In fact, every non-provable and non-refutable sentence is incomparable to its negation; take, for example, any atom in propositional logic or $\forall x \forall y (x = y)$ in predicate logic with equality. For incomparable theories T and T' and a real-valued weighing $\mathcal{W}$, either $\mathcal{W}(T) \leq \mathcal{W}(T')$ or $\mathcal{W}(T') \leq \mathcal{W}(T)$ holds, but neither $T' \vdash T$ holds nor $T \vdash T'$. So, HP^{-1} is out of the question as long as our weighing mappings are real-valued.

Let us now consider a couple of non-real-valued mappings that satisfy both HP and HP^{-1} . For the first example, consider the deductive closure T^+ of a theory T , which consists of all the T -provable sentences. Now, for all theories T and U , we have $T \vdash U \Leftrightarrow U \subseteq T \Leftrightarrow T^+ \supseteq U^+$. Thus, deductively closed sets can weigh theories, and they satisfy $\text{HP} + \text{HP}^{-1}$ with the inclusion order ($\supseteq$), which is transitive but not linear. Our second example will give rise to a real-valued weighing.

Definition 2.11 ($\langle \psi_n \rangle_{n>0}$). *Fix $\psi_1, \psi_2, \psi_3, \dots$ to be a list of all the sentences (in a fixed countable language and computing framework). The list can be taken to be effective in the sense that for a given $n > 0$ it is possible*

to find, in a computable way, the sentence ψ_n . $\diamond$

We consider the infinite binary $\{0, 1\}$ -sequences for our second example.

Definition 2.12 ($\sigma, \sqsubseteq$). For a theory T , let $\sigma(T) = \langle \mathcal{W}^{\{\psi_n\}}(T) \rangle_{n>0}$ (see Definition 2.4). For binary sequences $\varsigma = \langle \varsigma_n \rangle_{n>0}$ and $\tau = \langle \tau_n \rangle_{n>0}$, let $\varsigma \sqsubseteq \tau$ mean that $\varsigma_n \leq \tau_n$ holds for every $n > 0$. $\diamond$

Let us note that the binary relation $\sqsubseteq$ is transitive but non-linear (for example, the sequence $\langle 0, 1, 1, 1, \dots \rangle$ is $\sqsubseteq$ -incomparable with $\langle 1, 0, 1, 1, \dots \rangle$). However, σ satisfies HP and HP^{-1} with respect to $\sqsubseteq$.

Proposition 2.13 ($\text{HP} + \text{HP}^{-1}$ for σ with $\sqsubseteq$). For all theories T and U , we have $T \vdash U \Leftrightarrow \sigma(T) \sqsubseteq \sigma(U)$.

Proof.

(1) If $T \vdash U$, then for every sentence ψ_n , $U \vdash \psi_n$ implies $T \vdash \psi_n$; thus, $\mathcal{W}^{\{\psi_n\}}(T) \geq \mathcal{W}^{\{\psi_n\}}(U)$ for every $n > 0$; therefore, $\sigma(T) \sqsubseteq \sigma(U)$. (2) If $T \not\vdash U$, then for some sentence ψ_m , we have $U \vdash \psi_m$ but $T \not\vdash \psi_m$ (one can take ψ_m to be one of the T -unprovable axioms of U); thus, by $\mathcal{W}^{\{\psi_m\}}(T) = 0$ and $\mathcal{W}^{\{\psi_m\}}(U) = 1$, we obtain $\sigma(T) \not\sqsubseteq \sigma(U)$. $\square$

Before going back to real-valued weights, let us notice another property of the binary sequence $\sigma(T)$, in Definition 2.12, which will be needed later.

Lemma 2.14 (When $\sigma(T)$ is eventually constant). For a theory T , the sequence $\sigma(T)$ is eventually constant if and only if it is all 1 if and only if T is inconsistent.

Proof. Since $\langle \psi_n \rangle_{n>0}$ contains infinitely many tautologies, $\sigma(T)$ cannot be eventually 0. Clearly, $\sigma(T)$ is all 1 for an inconsistent theory T . Conversely, if $\sigma(T)$ is eventually 1, then T must be inconsistent, since a consistent theory cannot derive infinitely many contradictions that exist in the list $\langle \psi_n \rangle_{n>0}$. $\square$

We saw that HP^{-1} does not hold for real-valued mappings. So, let us consider a rather weak consequence of HP^{-1} , whose fulfillment will save the weights from being trivial or finitely-many-valued.

Definition 2.15 (Equivalence Principle, EP). A real-valued mapping $\mathcal{W}$ on theories is said to satisfy the Equivalence Principle, EP, when for all theories T and U , if we have $\mathcal{W}(T) = \mathcal{W}(U)$ then T is equivalent to U (i.e., $T \vdash U$ and $U \vdash T$); in symbols

(EP) $\mathcal{W}(T) = \mathcal{W}(U) \implies T \equiv U$. $\diamond$

Thus, under EP, non-equivalent theories should have different weights; this was not the case for any of the real-valued mappings that we have considered so far. Let us notice that the converse of EP, that is,

$$\text{EP}^{-1} : T \equiv U \implies \mathcal{W}(T) = \mathcal{W}(U),$$

is a consequence of HP (and as we noted above, EP follows from HP^{-1}).

Remark 2.16 (Chaitin's Characteristic Constant). *The minimum natural number c such that for every w , $T \not\vdash \mathcal{K}(x) > c$, is called Chaitin's characteristic constant of the theory T , denoted $\mathbb{C}_T$. Its existence follows from Chaitin's proof of the incompleteness theorem for sufficiently strong consistent theories. This constant satisfies HP, since if $\mathbb{C}_U > \mathbb{C}_T$, then for some w we should have $U \vdash \mathcal{K}(w) > \mathbb{C}_T$, thus $T \not\vdash U$. But neither it nor any integer-valued weighing can satisfy EP, since there could be infinitely many distinct theories between two consistent theories; see Corollary 2.10. $\diamond$*

Finally, we now introduce a real-valued weighing that satisfies both HP and EP.

Definition 2.17 ($\mathcal{V}$). *For a theory T , let $\mathcal{V}(T) = \sum_{n>0} 2^{-n} \mathcal{W}^{\{\psi_n\}}(T)$. $\diamond$*

Thus, $\mathcal{V}(T) = 0 \bullet \sigma(T)$ in base 2 (see Definition 2.12); recall from Definition 2.4 that

$$\mathcal{W}^{\{\psi_n\}}(T) = \begin{cases} 0 & \text{if } T \not\vdash \psi_n, \\ 1 & \text{if } T \vdash \psi_n. \end{cases}$$

Theorem 2.18 ($\mathcal{V}$ satisfies HP+EP). *The mapping $\mathcal{V}$ satisfies both HP and EP for all theories.*

Proof. Let T and U be two theories. (HP): If $T \vdash U$, then Proposition 2.13 implies $\mathcal{V}(T) \geq \mathcal{V}(U)$. (EP): If $\mathcal{V}(T) = \mathcal{V}(U)$, then since by Lemma 2.14 neither $\mathcal{V}(T)$ nor $\mathcal{V}(U)$ can be eventually 0, $\mathcal{W}^{\{\psi_n\}}(T) = \mathcal{W}^{\{\psi_n\}}(U)$ holds for every $n > 0$. Again, Proposition 2.13 implies $T \equiv U$. $\square$

2.4. Computability and Probability

What is the use of a weighing if it cannot be computed from (a finite specification of) the theory? It is easy to see that our mapping $\mathcal{V}$ in Definition 2.17 is computable when the underlying logic is decidable (like propositional logic or monadic first-order logic) and our theories are finite. So, over a decidable logic, we do have some computable weightings that satisfy both HP and EP for finite theories. But the story changes dramatically when the underlying logic is not decidable.

Theorem 2.19 (Undecidability implies incomputability).

Over an undecidable classical logic, no weighing can be computable if it satisfies both HP and EP.

Proof. Assume that a computable weighing $\mathcal{W}$ satisfies HP and EP. Let $\mathfrak{C} = \mathcal{W}(\{\perp\})$. Then, for every sentence ψ we have

$\vdash \psi$	$\Leftrightarrow$	$\{\neg\psi\}$ $\equiv \{\perp\}$		by classical logic,
	$\Leftrightarrow$	$\mathscr{W}(\{\neg\psi\})$ $= \mathscr{W}(\{\perp\})$		by HP+EP,
	$\Leftrightarrow$	$\mathscr{W}(\{\neg\psi\})$ $= \mathfrak{C}$		by the definition of $\mathfrak{C}$.

Thus, the logic should be decidable by the computability of $\mathscr{W}$ (and of $\mathfrak{C}$, $\neg$). $\square$

So, we cannot have any computable weighing that satisfies both HP and EP over first-order logic with a binary relation symbol, even if our theories are all finite. One may wonder if Theorem 2.19 still holds if we loosen the requirements to not require EP. We noted at the beginning of §2 that each constant weighing satisfies HP; those weightings are obviously computable. For a decidable structure $\mathfrak{M}$, the weight $\mathscr{W}_{\mathfrak{M}}$ in Definition 2.1 is computable and satisfies HP by Theorem 2.2. Let us recall that a structure $\mathfrak{M}$ is decidable when there exists an algorithm that decides (outputs Yes or No) if a given sentence ψ holds in $\mathfrak{M}$ or not (whether $\mathfrak{M} \models \psi$ or $\mathfrak{M} \not\models \psi$). Every finite structure over a finite language is decidable, such as the ordered set $\{1, 2, \dots, n\}$ over the language $\{<\}$, or the ring $\mathbb{Z}_n (= \mathbb{Z}/n\mathbb{Z})$ over the language $\{+, \times\}$.

Theorem 2.20 (Computable weightings satisfying HP but not EP). *Over a finite language, for every natural number n , there is some $(n + 1)$ -valued weighing of finite theories that is computable and satisfies HP.*

Proof. Let the set $\mathcal{C}$ consist of n finite, pairwise non-equinumerous structures. For every sentence ψ , let $\mathscr{W}(\psi)$ be the number of structures in $\mathcal{C}$ that do not satisfy ψ . For a finite theory T , let $\mathscr{W}(T)$ be $\mathscr{W}(\tau)$, where τ is the conjunction of the finitely many axioms of T . By the decidability of the structures in the finite set $\mathcal{C}$, the weighing $\mathscr{W}$ is computable. We show that it satisfies HP. Suppose $\mathscr{W}(\psi) > \mathscr{W}(\varphi)$ for sentences φ and ψ . Then the number of structures in $\mathcal{C}$ that do not satisfy ψ is strictly greater than the number of structures in $\mathcal{C}$ that do not satisfy φ . By the pigeonhole principle, one of the structures in $\mathcal{C}$ that does not satisfy ψ , say $\mathfrak{M}$, should satisfy φ . Thus, $\mathfrak{M} \not\models \psi$ but $\mathfrak{M} \models \varphi$; therefore, $\varphi \not\vdash \psi$. $\square$

It is tempting to interpret $\mathscr{V}(T)$ as a *proving measure* of the theory T , for the coefficient of 2^{-n} in the binary expansion of $\mathscr{V}(T)$ is 1 if $T \vdash \psi_n$ and is 0 otherwise. The following definition and theorem show that one should strongly resist this temptation.

Definition 2.21 ($\mathcal{V}_\alpha^{(a,b)}$). Let a, b be two real numbers such that $b > a \geq 0$. Let $\langle \alpha_n \rangle_{n>0}$ be a sequence of positive real numbers such that the series $\sum_{n>0} \alpha_n$ is fast converging with respect to a/b , in the sense that for every $n > 0$ we have $\sum_{i>n} \alpha_i < \alpha_n(1 - a/b)$. One can take, for example, $\alpha_n = c^{-n}$ for a real number c with $c > 1 + \frac{b}{b-a}$. Let

$$\sigma_n^{(a,b)}(T) = \begin{cases} a & \text{if } T \not\vdash \psi_n, \\ b & \text{if } T \vdash \psi_n; \end{cases}$$

for a theory T . Finally, put $\mathcal{V}_\alpha^{(a,b)}(T) = \sum_{n>0} \alpha_n \sigma_n^{(a,b)}(T)$. $\spadesuit$

Theorem 2.22 ($\mathcal{V}_\alpha^{(a,b)}$ satisfies HP+EP). If a, b , and $\langle \alpha_n \rangle_{n>0}$ are as in Definition 2.21 above, then the mapping $\mathcal{V}_\alpha^{(a,b)}$ satisfies both HP and EP for all theories.

Proof. The analogue of Proposition 2.13 holds for $\sigma_n^{(a,b)}$: for all theories T and U , $T \vdash U \Leftrightarrow \forall n > 0 [\sigma_n^{(a,b)}(T) \geq \sigma_n^{(a,b)}(U)]$.

So, HP holds too: if $T \vdash U$ then $\mathcal{V}_\alpha^{(a,b)}(T) \geq \mathcal{V}_\alpha^{(a,b)}(U)$. For showing EP, suppose $\mathcal{V}_\alpha^{(a,b)}(T) = \mathcal{V}_\alpha^{(a,b)}(U)$. By the $\sigma_n^{(a,b)}$ -analogue of Proposition 2.13, for proving $T \equiv U$, it suffices to show $\sigma_n^{(a,b)}(T) = \sigma_n^{(a,b)}(U)$ for each $n > 0$. If this is not the case, then let m be the minimum $i > 0$ such that $\sigma_i^{(a,b)}(T) \neq \sigma_i^{(a,b)}(U)$. Without loss of generality, we can assume that $\sigma_m^{(a,b)}(T) = a$ and $\sigma_m^{(a,b)}(U) = b$. Then

$\mathcal{V}_\alpha^{(a,b)}(T)$	=	$\sum_{0 < i < m} \alpha_i \sigma_i^{(a,b)}(T)$ + $\alpha_m a$ + $\sum_{j > m} \alpha_j \sigma_j^{(a,b)}(T)$	by Definition 2.21,
	=	$\sum_{0 < i < m} \alpha_i \sigma_i^{(a,b)}(U)$ + $\alpha_m a$ + $\sum_{j > m} \alpha_j \sigma_j^{(a,b)}(T)$	by the choice of m ,
	$\leq$	$\sum_{0 < i < m} \alpha_i \sigma_i^{(a,b)}(U)$ + $\alpha_m a$ + $\sum_{j > m} \alpha_j b$	by $\sigma_j^{(a,b)}(T) \leq b$,
	<	$\sum_{0 < i < m} \alpha_i \sigma_i^{(a,b)}(U)$ + $\alpha_m b$	by the α_n 's property,
	$\leq$	$\sum_{0 < i < m} \alpha_i \sigma_i^{(a,b)}(U)$ + $\alpha_m b$ + $\sum_{j > m} \alpha_j \sigma_j^{(a,b)}(U)$	by $0 \leq \sigma_j^{(a,b)}(U)$,
	=	$\mathcal{V}_\alpha^{(a,b)}(U)$	by Definition 2.21.

This contradicts the assumption $\mathcal{V}_\alpha^{(a,b)}(T) = \mathcal{V}_\alpha^{(a,b)}(U)$ above. $\square$

The mapping $\mathcal{V}_\alpha^{(a,b)}$ can take values in the interval $(0, 1)$, for example, when we put $\alpha_n = c^{-n}$ for some real numbers a, b, c that satisfy $0 \leq a \leq b - 1 < c - 2$ (such as $a = 2, b = 4, c = 7$). But there is no reason to see $\mathcal{V}_\alpha^{(a,b)}$ as the probability of anything.³

2.4.1. Recursively Enumerable Theories

We now consider *recursively enumerable* (re) theories; an re set is the (possibly infinite) set of the outputs of a fixed input-free program. Let $\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3, \dots$ be an effective list of all the re theories (in a fixed language and computing framework). We notice how arbitrary an (HP+EP)-satisfying weighing of re theories can be:

Remark 2.23 (Arbitrary weightings that satisfy HP and EP for all re theories). Let us define $\mathcal{U}(\mathbb{T}_n)$ by induction on $n > 0$, in such a way that both HP and EP are satisfied. Take $\mathcal{U}(\mathbb{T}_1)$ to be an arbitrary real number. Suppose that $\{\mathcal{U}(\mathbb{T}_1), \dots, \mathcal{U}(\mathbb{T}_n)\}$ are defined and that HP and EP hold for $\mathbb{T}_1, \dots, \mathbb{T}_n$. We now define $\mathcal{U}(\mathbb{T}_{n+1})$. If $\mathbb{T}_{n+1} \equiv \mathbb{T}_m$ for some $m \leq n$, then let $\mathcal{U}(\mathbb{T}_{n+1}) = \mathcal{U}(\mathbb{T}_m)$. Now, assume that $\mathbb{T}_{n+1}$ is not equivalent to any of $\mathbb{T}_1, \dots, \mathbb{T}_n$. Let $\mathbb{T}_{i_1}, \dots, \mathbb{T}_{i_k}$ be all the sub-theories of $\mathbb{T}_{n+1}$ among $\mathbb{T}_1, \dots, \mathbb{T}_n$ if any; if there is no such sub-theory, then $k = 0$. Also, let $\mathbb{T}_{j_1}, \dots, \mathbb{T}_{j_\ell}$ be all the super-theories of $\mathbb{T}_{n+1}$ among $\mathbb{T}_1, \dots, \mathbb{T}_n$ if any; if there is no such super-theory, then $\ell = 0$. Since both HP and EP hold for $\mathbb{T}_1, \dots, \mathbb{T}_n$ by the induction assumption, then for every $u \leq k$ and $v \leq \ell$ we have $\mathcal{U}(\mathbb{T}_{i_u}) < \mathcal{U}(\mathbb{T}_{j_v})$; note that all $\mathbb{T}_{i_u}$'s are *strict* sub-theories of $\mathbb{T}_{n+1}$ and all $\mathbb{T}_{j_v}$'s are *strict* super-theories of $\mathbb{T}_{n+1}$. Let $\text{II} = \max\{\mathcal{U}(\mathbb{T}_{i_1}), \dots, \mathcal{U}(\mathbb{T}_{i_k})\}$ and $\text{JJ} = \min\{\mathcal{U}(\mathbb{T}_{j_1}), \dots, \mathcal{U}(\mathbb{T}_{j_\ell})\}$; notice that $\max \emptyset = -\infty$ and $\min \emptyset = +\infty$ by convention. Finally, take $\mathcal{U}(\mathbb{T}_{n+1})$ to be any real number between II and JJ (note that we have $\text{II} < \text{JJ}$ by what was said above; thus, $\text{II} < \mathcal{U}(\mathbb{T}_{n+1}) < \text{JJ}$ is quite possible). $\diamond$

The above construction can be computable if our logic is decidable and our theories are all finite. If the underlying logic is not decidable or we do not want to be restricted to finite theories, then the above construction is $\emptyset''$ -computable (i.e., is computable if one has access to the oracle $\emptyset''$). All we have to do is check $\mathbb{T}_n \vdash \mathbb{T}_m$, which amounts to the Π_2 -statement $\forall \tau \in \mathbb{T}_m \exists \pi (\pi \text{ is a } \sim \mathbb{T}_n \text{-proof of } \sim \tau)$; here we need the effectivity of the list $\langle \mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3, \dots \rangle$. Let us also note that the construction of Remark 2.23 highly depends on how the re theories are ordered; cf. also [5].

3. Conclusions

The Kolmogorov–Chaitin complexity is not a good way of weighing theories or sentences since it does not satisfy Chaitin’s heuristic principle (HP), which says that theories can prove lighter sentences only (Definition 1.1). Neither does the δ -complexity, which is the Kolmogorov-complexity minus the length (Proposition 1.2). Due to the existence of incomparable pairs of theories (none of which derives from the other), the converse of HP (that all the lighter sentences are provable from the theory) cannot hold for real-valued weightings. HP can be satisfied insipidly by every constant weighing (all the theories and sentences weigh the same); this constant weighing is trivially computable. To make HP more meaningful, we added the Equivalence Principle (EP), which says that only equivalent theories can have equal weights (Definition 2.15). Note that a consequence of HP is that equivalent theories must weigh equally. EP, equivalently saying that equally weighted theories are logically equivalent, is the converse of this statement. If the underlying logic is undecidable, then every weight that satisfies HP and EP should be uncomputable (Theorem 2.19). Here, we demonstrated some weightings (Definitions 2.17 and 2.21) that satisfy HP and EP, and moreover, they are (i) computable if the underlying logic is decidable and the considered theories are all finitely axiomatizable, and (ii) uncomputable if the underlying logic is undecidable. So, regarding the satisfaction of HP and EP and the computability of the weighing, this seems to be the best that can be done. Our weightings should not be regarded as the probability of any event, even though they resemble the Ω number that is presumed to be the *probability of halting*. All this probably shows that it takes a genius’ idea (Chaitin’s constant number) to solve a genius’ problem (Chaitin’s heuristic principle).

A question that now comes to mind is: Are there any uses for HP-satisfying weights of theories? In fact, constructing counter-examples is the most natural way to show unprovability (for example, matrix multiplication shows that the commutativity axiom does not follow from the axioms of group theory). In other words, the most well-known unprovability technique has been our Definition 2.1 (and Theorem 2.2) above. Lobachevsky’s geometry is a historic example of a model for the unprovability of the parallel postulate from the rest of the axioms of Euclidean geometry. Gödel’s incompleteness theorems provide monumental examples of the other kinds: the unprovability of Gödelian sentences (in the first incompleteness theorem) and the unprovability of the consistency statement (in the second incompleteness theorem), for which Gödel did not construct any model (that satisfies the theory but does not satisfy a Gödelian sentence or the consistency statement); his proof was totally syntactic. So far, we

do not know if there ever was an unprovability result in the history of mathematics that used an (HP+EP)-satisfying weighing of theories.

Notes

2020 Math Subject Classification: 03F40 · 68Q30.

Footnotes

¹ Actually, for most of the standard formalisms and frameworks, it suffices to take $m = 2$ and $n = 0$.

² It would be a good exercise to go through the arguments of [8] and [9], pinpoint the possible errors, and see what went wrong.

³ On the contrary, for $4 \leq a = b - 8$ we have $1 + b/(b - a) = 2 + a/8 < a$, so for every c between $2 + a/8$ and a , such as $c = a - 1$ (for example, $a = 8, b = 16, c = 7$), we will have $\mathcal{V}_\alpha^{(a,b)}(T) > 1$, for every theory T (with $\alpha_n = c^{-n}$ for each $n > 0$). Let us also note that for every weight $\mathcal{W}$, the mapping $(1 + 2^\mathcal{W})^{-1}$ ranges over $(0, 1)$ but cannot be interpreted as a probability.

References

1. [△]Chaitin G (1974). "Information-Theoretic Limitations of Formal Systems." *J Assoc Comput Mach.* 21(3):403–424. doi:[10.1145/321832.321839](https://doi.org/10.1145/321832.321839).
2. [△]van Lambalgen M (1989). "Algorithmic Information Theory." *J Symb Log.* 54(4):1389–1400. doi:[10.2307/2274821](https://doi.org/10.2307/2274821).
3. [△]Chaitin G (1992). "LISP Program-Size Complexity II." *Appl Math Comput.* 52(1):103–126. doi:[10.1016/0096-3003\(92\)90100-F](https://doi.org/10.1016/0096-3003(92)90100-F).
4. [△]_a [△]_b [△]_cFallis D (1996). "The Source of Chaitin's Incorrectness." *Philos Math.* 4(3):261–269. doi:[10.1093/philmat/4.3.261](https://doi.org/10.1093/philmat/4.3.261).
5. [△]_a [△]_bRaatikainen P (1998). "On Interpreting Chaitin's Incompleteness Theorem." *J Philos Log.* 27(6):569–586. doi:[10.1023/A:1004305315546](https://doi.org/10.1023/A:1004305315546).
6. [△]Sjögren J (2004). "Measuring the Power of Arithmetical Theories." University of Gothenburg, Department of Philosophy. <https://gupea.ub.gu.se/handle/2077/19472>.

7. ^a ^bSjögren J (2008). "On Explicating the Concept "the Power of an Arithmetical Theory"" *J Philos Log*. 37 (2):183–202. doi:[10.1007/s10992-007-9077-8](https://doi.org/10.1007/s10992-007-9077-8).
8. ^a ^b ^c ^d ^eCalude CS, Jürgensen H (2005). "Is Complexity a Source of Incompleteness?" *Adv Appl Math*. 35 (1):1–15. doi:[10.1016/j.aam.2004.10.003](https://doi.org/10.1016/j.aam.2004.10.003).
9. ^a ^bGrenet B (2010). "Acceptable Complexity Measures of Theorems." *Complex Syst*. 18(4):403–425. doi:[10.25088/ComplexSystems.18.4.403](https://doi.org/10.25088/ComplexSystems.18.4.403).
10. ^ΔPorter CP (2021). "Revisiting Chaitin's Incompleteness Theorem." *Notre Dame J Formal Log*. 62(1):147–171. doi:[10.1215/00294527-2021-0006](https://doi.org/10.1215/00294527-2021-0006).
11. ^ΔBienvenu L, Romashchenko A, Shen A, Tavenaux A, Vermeeren S (2014). "The Axiomatic Power of Kolmogorov Complexity." *Ann Pure Appl Logic*. 165(9):1380–1402. doi:[10.1016/j.apal.2014.04.009](https://doi.org/10.1016/j.apal.2014.04.009).
12. ^ΔHunter A, Konieczny S (2010). "On the Measure of Conflicts: Shapley Inconsistency Values." *Artif Intell*. 174 (14):1007–1026. doi:[10.1016/j.artint.2010.06.001](https://doi.org/10.1016/j.artint.2010.06.001).

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.