Journées Automates Cellulaires 2008 (Uzès), pp. 74-89

GÖDEL INCOMPLETENESS REVISITED

GREGORY LAFITTE [1]

[1] Laboratoire d'Informatique Fondamentale de Marseille (LIF), CNRS – Aix-Marseille Université,
39 rue Joliot-Curie, 13453 Marseille Cedex 13, France
E-mail address: Gregory.Lafitte@lif.univ-mrs.fr
URL: http://www.lif.univ-mrs.fr/~lafitte/

Abstract. We investigate the frontline of Gödel's incompleteness theorems' proofs and the links with computability.

The Gödel incompleteness phenomenon

Gödel's incompleteness theorems [Göd31, SFKM$^+$86] are milestones in the subject of mathematical logic.

Apart from Gödel's original syntactical proof, many other proofs have been presented. Kreisel's proof [Kre68] was the first with a model-theoretical flavor. Most of these proofs are attempts to get rid of any form of self-referential reasoning, even if there remains diagonalization arguments in each of these proofs. The reason for this quest holds in the fact that the diagonalization lemma, when used as a method of constructing an independent statement, is intuitively unclear. Boolos' proof [Boo89b] was the first attempt in this direction and gave rise to many other attempts. Sometimes, it unfortunately sounds a bit like finding a way to sweep self-reference under the mathematical rug.

One of these attempts has been to prove the incompleteness theorems using another paradox than the Richard and the Liar paradoxes. It is interesting to note that, in his famous paper announcing the incompleteness theorem, Gödel remarked that, though his argument is analogous to the Liar paradox, "Any epistemological antinomy could be used for a similar proof of the existence of undecidable propositions". G. Boolos has proved quite recently (1989) a form of the first incompleteness theorem using Berry's paradox consisting in the fact that "the least integer not nameable in fewer than seventy characters" has just now been named in sixty-three characters. G. Boolos thought the interest of such proofs is that they provide a *different sort of reason* for incompleteness. It is true that each of these new arguments gives us a better understanding of the incompleteness phenomenon.

When studying proofs and provability, there are two different points of view: the proof-theoretical one (axioms and inference rules) and the model-theoretical one (axioms, models, consequences). The former one tends to be quite syntactical and the latter one more semantical. We have tried to present both points of view and linger over the model-theoretical side because, at least from the author's point of view, model-theoretic arguments are intuitively clearer than proof-theoretic ones.

Gödel's argumentation was heavily based not only on the arithmetization of syntax, but on the arithmetization of all mathematical objects (sentences, proofs, theories) and the fact that all this arithmetization is primitive recursive. In fact, it has opened the way for the notions of computation and computability to arise.

The goal of this paper is twofold: a survey of incompleteness proofs and to precise links with computability.

Computability and incompleteness are inherently linked. For instance, one can obtain a first form of the first incompleteness theorem by considering propositions of the form $n \notin X$, where X is a non-recursive but recursively enumerable set, *e.g.*, the *diagonal halting* set $\mathcal{K}$. Even if the language of the considered theory does not contain $\in$, there is a simple algorithm that generates given n the proposition "$n \notin X$". Given a sound (every provable statement is true) recursively enumerable theory T, there is a number n_0 such that $n_0 \notin X$ but T does not prove it. The proof is direct: Suppose that there is no such n_0, then we would have that T proves "$n \notin X$" if and only $n \notin X$, and X would be recursive (generate the theorems of T and at the same time enumerate X; if $n \in X$ then n will eventually show up in the enumeration; otherwise, "$n \notin X$" will eventually show up in the theorems of T and be true by the soundness assumption). We thus have a true sentence, "$n_0 \notin X$", which is not provable in T.

Incompleteness is also famously linked to computability via Chaitin's incompleteness theorem. Chaitin's result, showing that there are unprovable statements on Kolmogorov-Chaitin complexity[1], is a form of Gödel's first incompleteness theorem. Actually, Kolmogorov showed in the sixties that the set of non-random (or incompressible) numbers, *i.e.*, $\{x : K(x) \geqslant x\}$, is recursively enumerable but not recursive, and, by the above argument, this is already a version of Gödel's first incompleteness theorem. Moreover, Kolmogorov's proof can be seen as an application of Berry's paradox. Following Boolos, it is thus no wonder that we can get proofs using this Kolmogorov complexity function (or other *similar* computability-related functions) of both incompleteness theorems.

One of the reason of the existence of the quest of better understanding the incompleteness phenomenon holds in the peculiarity of Gödel's unprovable statements. They are not natural mathematical statements: no mathematician has ever stumbled on them (or should we say *over them*?). And thus, it seems to many that normal mathematical practice is not concerned with the incompleteness phenomenon. More and more results show however the contrary. In particular, Harvey Friedman's Π^0_1 statements, that are unprovable in Zermelo-Fraenkel (ZF) set theory and need the 1-consistency of strong set-theoretical unprovable statements, going way beyond ZF, to be proved, are examples of such results.

Nevertheless, incompleteness theorems only provide unprovable statements like the *consistency* of a theory, that are of an unclear nature. What combinatorial properties does the consistency statement bring to a theory? Feferman [Fef62, Her88] has shown that a certain reflection principle, an unprovable statement, has to be *added* $\omega^{\omega^{\omega+1}}$ times to Peano arithmetic in order to *cover* all true arithmetical statements. Adding the 1-consistency, a soundness assumption, of strong set-theoretical unprovable statements, *e.g.,* large cardinal axioms, to a given arithmetical theory amounts to asserting that "every Π^0_2 consequence of these strong statements is true". In this case, the combinatorial properties that are *added* are the combinatorial Π^0_2 consequences of these strong statements. Having a link between consistency (or soundness) and computability, in particular Kolmogorov complexity, would

[1] Loosely speaking, the Kolmogorov-Chaitin complexity of a natural number n, denoted by $K(n)$, is the smallest size of a program which generates n.

make possible an understanding of what properties consistency adds to a theory. Adding consistency as an axiom would then yield new combinatorial properties because of the existing links between combinatorics and Kolmogorov complexity. This could be one reason behind the stir surrounding Chaitin's incompleteness result.

This paper is organized as follows. We start by recalling the basic notions behind formulæ, proofs, theories and arithmetization. Then we present Gödel's original proofs. We continue by presenting a survey of existing incompleteness proofs, of both first and second incompleteness theorems. We finish with incompleteness results and proofs that are computability-related and discuss the interpretation of Chaitin's incompleteness theorem.

1. Objects to play with

1.1. What are the basic objects?

On top of the usual logical connectives ($\wedge$, $\vee$ and $\neg$), we will respectively denote the logical connectives of *implication* and *equivalence* by $\rhd$ and $\equiv$.

$\mathcal{L}_{\mathrm{PA}}$ will designate the first order language on the signature of arithmetic $\{\mathbf{S}, +, \times, \leqslant, \mathbf{0}\}$. $\mathbf{S}$, $+$, $\times$ designate respectively the *successor*, *addition* and *multiplication* functions. $\leqslant$ designates the *lower-or-equal* relation and $\mathbf{0}$ designates the constant *zero*.

The Turing machines indexed by their codes, for any appropriate coding which we will later on make to coincide with the Gödel numbering, are denoted by $\{T_i\}_{i\in\mathbb{N}}$. A computation (of a Turing machine, or any equivalent computation model) either *diverges*, denoted by $\uparrow$, or *converges*, denoted by $\downarrow$. The partial recursive functions computed by Turing machines, following a fixed convention, are denoted by $\{\varphi_i\}_{i\in\mathbb{N}}$ (agreeing with the Turing machines' coding). The sets $\{W_i\}_{i\in\mathbb{N}}$ denote the recursively enumerable sets, *i.e.*, the domains of partial recursive functions. A central set in computability theory is the *diagonal halting set* $\mathcal{K} = \{x : \varphi_x(x) \downarrow\} = \{x : T_x(x) \downarrow\} = \{x : x \in W_x\}$. The set $\mathcal{K}$ is recursively enumerable but not recursive; it is the archetypal *creative* set.

Concerning computability, the reader is referred to [Odi89, Odi99, Rog67, Rog58, Smu93, VS03].

1.2. What is a proof?

A formal theory T is determined by a first order[2] language $\mathcal{L}_T$ and a set of axioms $\mathcal{A}_T$, which are formulæ in that language. The set Thms_T of *theorems* consists of those formulæ ϕ for which there is a proof in T.

There are two different points of view concerning proofs: the proof-theoretical one (axioms and inference rules) and the model-theoretical one (axioms, models, consequences).

[2] All our reasoning also works for theories on second order languages. For simplicity and brevity, we will only consider first order theories in this article.

1.2.1. *Proof theoretical.* Proofs are most commonly seen as a deduction sequence from a set of axioms.

In proof theory, we can for example take the following deduction rules:

$\Gamma \Rightarrow \phi$ if $\phi \in \Gamma$	$\Gamma \Rightarrow \phi \wedge \psi$ iff $\Gamma \Rightarrow$ and $\Gamma \Rightarrow \psi$
If $\Gamma \Rightarrow \phi$ or $\Gamma \Rightarrow \psi$, then $\Gamma \Rightarrow \phi \vee \psi$	If $\Gamma \cup \{\phi\} \Rightarrow \psi$, then $\Gamma \Rightarrow \phi \rhd \psi$
If $\Gamma \Rightarrow \phi$ and $\Gamma \Rightarrow \phi \rhd \psi$, then $\Gamma \Rightarrow \psi$	If $\Gamma \Rightarrow (s = t)$ and $\Gamma \Rightarrow \phi(s)_x$, then $\Gamma \Rightarrow \phi(t)_x$
If $\Gamma \Rightarrow \psi$ and $\Gamma \Rightarrow \neg\psi$, then $\Gamma \Rightarrow \phi$	$\Gamma \Rightarrow \forall x\ (x = x)$
If $\Gamma \cup \{\neg\phi\} \Rightarrow \psi$ and $\Gamma \cup \{\neg\phi\} \Rightarrow \neg\psi$, then $\Gamma \Rightarrow \phi$	
If $\Gamma \cup \{\phi\} \Rightarrow \psi$ and $\Gamma \cup \{\phi\} \Rightarrow \neg\psi$, then $\Gamma \Rightarrow \neg\phi$	
If $\Gamma \cup \{\phi\} \Rightarrow \theta$ and $\Gamma \cup \{\psi\} \Rightarrow \theta$, then $\Gamma \cup \{\phi \vee \psi\} \Rightarrow \theta$	
If $\Gamma \Rightarrow \phi$ and x does not occur free in Γ, then $\Gamma \cup \Delta \Rightarrow \forall x\ \phi$	
If $\Gamma \Rightarrow \forall x\ \phi$, then $\Gamma \Rightarrow \phi(s)_x$ for any term s free for x in ϕ	
If $\Gamma \Rightarrow \phi(s)_x$, then $\Gamma \Rightarrow \exists x\ \phi$, for any term s free for x in ϕ	
If $\Gamma \cup \{\phi(y)_x\} \Rightarrow \psi$ and y is not free in Γ or ψ, then $\Gamma \cup \Delta \cup \{\exists x\ \phi\} \Rightarrow \psi$	
If $\Gamma \Rightarrow \forall x\ (x \in X \equiv x \in Y)$, then $\Gamma \Rightarrow X = Y$	

$\Delta \Rightarrow \phi$ holds if and only if there is a derivation showing this in the form of a finite sequence $\langle \Gamma_1, \phi_1 \rangle$, ..., $\langle \Gamma_n, \phi_n \rangle$, where $\langle \Gamma_n, \phi_n \rangle$ is $\langle \Delta, \phi \rangle$ and each $\langle \Gamma_i, \phi_i \rangle$ follows by one of the above rules from previous pairs in the sequence. A *derivation* is a sequence number $\langle s_0, \ldots, s_n \rangle$ where each s_i is a pair $\langle t_i, \phi_i \rangle$ with t_i a sequence number of formulæ and s_i is related as indicated in the rules to zero, one or two previous pairs in the sequence. s is a derivation *of* ϕ *from* Γ if s_n is $\langle t_n, \phi \rangle$ where every formula in the sequence t_n is a member of Γ. A *proof* in a theory T is a derivation from $\mathcal{A}_T$.

1.2.2. *Model theoretical.* Another way to consider provability is through models. A sentence ϕ is provable in an axiomatic theory T if all models of T satisfy ϕ.

Leon Henkin gave in 1949 a non-constructive but easier (than Gödel's original) proof of Gödel's completeness theorem. It consists in reducing the consistency of a set of sentences in a language L to that of a set of quantifier-free sentences in an extended language. This process can be arithmetized to build a partial order, called the *Henkin tree.* It gives an arithmetical Δ_{n+1} model for any consistent Σ_n or Δ_n theory. For a complete description, the reader is referred to [Kay91].

Through Henkin's method, we can obtain a more model-theoretical notion of proof. If ϕ and ψ are sentences, to say that ψ is a consequence of ϕ is to say that the set $\{\phi, \neg\psi\}$ is inconsistent. The consistency of $\{\phi, \neg\psi\}$ can be determined by Henkin's method. We get a proof that ψ is a consequence of ϕ as soon as we reach a natural number p at which the branches of Henkin's tree all end at a contradiction. This natural number p can take the place of a proof.

If a theory T is a Σ_n fragment of arithmetic, then consistency can be expressed by a Π_n sentence: it is enough to express that $\mathbf{0} = \mathbf{S0}$ is not a consequence of the axioms of T or else to express that the Henkin tree associated with T is infinite.

For more on model theory, see [Hod93].

1.3. What is an arithmetical-able theory?

Throughout this paper, T will be some fixed, but unspecified, consistent formal theory.

The properties that a theory T should meet to satisfy the clauses of incompleteness theorems are for it to *contain arithmetic.* There are several ways to precise these properties. These properties are encodability conditions and, as Gödel showed, one can do a great deal of encoding on natural numbers.

To follow classical expositions of the incompleteness theorems, we assume that the encoding is done in some fixed formal theory S and that T contains S. S is usually not specified but it is commonly taken to be a formal system of arithmetic, although a weak set theory[3] is often more convenient. If S is a formal system of arithmetic, *e.g.*, PA (Peano Arithmetic), and T is ZF (Zermelo-Fraenkel set theory), then T contains S in the sense that there is a well-known embedding of S in T.

S needs to be able to represent primitive recursive functions. It should be *primitive recursive*-able. A more model-theoretical way to require these properties is to require of the theory to have Σ_1-induction.

To each formula ϕ of the language of T is assigned a closed term, $\ulcorner\phi\urcorner$, called the *code* of ϕ. For any natural number n, $\ulcorner n\urcorner$ designates a closed term, the *numeral* for n, in the language S that represents n, *e.g.*, $\sigma(\sigma(\dots(0)\dots))$. n is called the *value* of this numeral $\ulcorner n\urcorner$.

To avoid any ambiguity, we define a function called var. An ambiguity arises in the following example. There are two possible meanings for $\ulcorner x\urcorner$: the code for the value for the variable x *or* the code for the variable x. $\mathrm{var}(x)$ designates the latter case, *i.e.*, the code for the variable x.

S will have certain function symbols corresponding to the logical connectives and quantifiers : neg, implies, *etc.*, such that, for all formulæ ϕ, ψ, $S \vdash \mathrm{neg}(\ulcorner\phi\urcorner) = \ulcorner\neg\phi\urcorner$, $S \vdash \mathrm{implies}(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) = \ulcorner\phi \rhd \psi\urcorner$, *etc.*

The *substitution* operator, represented in S by the function symbol sub, is of particular importance. For any codes c_1 and c_2 for terms t_1 and t_2 and a variable x, $\mathrm{sub}_x(c_1, c_2)$ is the code of the term that results from substituting t_1 for every occurrence of x in t_2 : $S \vdash \mathrm{sub}_x(\ulcorner t\urcorner, \ulcorner\phi(x)\urcorner) = \ulcorner\phi(t)_x\urcorner$.

For readability, we will use the same names for functions and predicates in formulæ and in the running text. All the functions previously defined are actually primitive recursive.

From the previous discussion on proofs, we have a binary relation, whose symbol in S is Proof, such that for closed t_1 and t_2: $S \vdash \mathrm{Proof}_T(t_1, t_2)$ iff t_1 is the code of a *proof* in T of the formula with code t_2. It follows that $T \vdash \phi$ if and only if $S \vdash \mathrm{Proof}_T(t, \ulcorner\phi\urcorner)$ for some closed term t.

We then define a predicate, whose symbol in S is Prov, asserting provability :

$$\mathrm{Prov}_T(y) \equiv \exists x \;\; \mathrm{Proof}_T(x, y)$$

One must be careful and understand that we do not always have : $T \vdash \phi$ if and only if $S \vdash \mathrm{Prov}_T(\ulcorner\varphi\urcorner)$. It depends on the *soundness* properties of our theory. Soundness is linked to consistency as summarized in section 1.4.

We will use a special notation for formalizations of provability statements. If ϕ is a sentence, we write $\Box\phi$ for the sentence $\mathrm{Prov}_T(\ulcorner\phi\urcorner)$, where the theory T is implicit. Accordingly, $\Box\,\Box\,\phi$ is the formula $\mathrm{Prov}_T(\ulcorner\psi\urcorner)$ where ψ is $\mathrm{Prov}_T(\ulcorner\phi\urcorner)$.

[3] See [Dev84, Jec78].

This encoding (of S in T) can be carried out in such a way that the following important conditions, the *deducibility* (or *derivability*) *conditions*, are met for all sentences ϕ:

$$T \vdash \phi \text{ implies } S \vdash \mathrm{Prov}_T(\ulcorner\phi\urcorner), \text{ for every sentence } \phi. \tag{1.1}$$

$$S \vdash \mathrm{Prov}_T(\ulcorner\phi\urcorner) \rhd \mathrm{Prov}_T(\ulcorner\mathrm{Prov}_T(\ulcorner\phi\urcorner)\urcorner), \text{ for every sentence } \phi. \tag{1.2}$$

$$S \vdash \mathrm{Prov}_T(\ulcorner\phi\urcorner) \wedge \mathrm{Prov}_T(\ulcorner\phi \rhd \psi\urcorner) \rhd \mathrm{Prov}_T(\ulcorner\psi\urcorner), \text{ for all sentences } \phi,\, \psi. \tag{1.3}$$

Much of the intricacy of Gödel's incompleteness theorems' proofs lies in the scarcely illuminating details of setting up and checking the properties of a coding system representing the syntax of $\mathcal{L}_{\mathrm{PA}}$ within that same language. For this reason a number of efforts have been made to present the essentials of the proofs of Gödel's theorems without getting entangled in syntactic details. One of the most important of these efforts was made by Löb [Löb55] and Hilbert and Bernays [HB39]. They formulated these three conditions on the provability predicate in a formal system which are jointly sufficient to yield Gödel's second incompleteness theorem.

Given that the axioms of T are defined using a Σ-formula, these deducibility conditions all hold: the first two conditions are corollaries of the Σ-completeness theorem and the third condition is a formalization of an obvious argument.

1.4. What are *consistency* statements?

A theory T is *inconsistent* if there exists ϕ such that ϕ **and** $\neg\phi$ are theorems of T, and otherwise *consistent.*

A theory T is *complete* if for every sentence ϕ in the language of T, ϕ or $\neg\phi$ is a theorem of T, and otherwise *incomplete.*

Gödel introduced a stronger form of consistency, coined ω-consistency. In a w-consistent theory T, we cannot have at the same time $T \vdash \exists x\ \phi(x)$ and $T \vdash \neg\phi(\ulcorner 0\urcorner), T \vdash \neg\phi(\ulcorner 1\urcorner), \ldots$ (having for all natural number i, $T \vdash \neg\phi(\ulcorner i\urcorner)$).

More formally: T is ω-*consistent* if for any formula ϕ

$$\mathrm{Prov}_T(\ulcorner\exists x\ \phi(x)\urcorner) \text{ implies } \exists x\ \neg\,\mathrm{Prov}_T(\ulcorner\neg\phi(x)\urcorner) \tag{1.4}$$

ω-consistency is a restriction of another property, *reflection*:

$$\mathrm{Refl}_T : \quad \mathrm{Prov}_T(\ulcorner\phi\urcorner) \text{ implies } \phi \text{ for closed } \phi$$

For $\phi \in \Delta_0$, (1.4) is called 1-*consistency.* It can be shown that 1-consistency means that all Σ_1 provable statements are true. It is actually Refl_T for Σ_1 statements, denoted by $\mathrm{Refl}_T^{\Sigma_1}$. Reflection is also called *soundness.* Σ-soundness is 1-consistency

Nevertheless, every arithmetical-able theory T has the following property.

Theorem 1.1 (Σ_1-completeness)**.** *If ϕ is a Σ_1 statement, then $S \vdash \phi \rhd \mathrm{Prov}_T(\ulcorner\phi\urcorner)$.*

Hence, in T, Cons_T, the statement expressing that there is no proof in T of $\mathbf{0} = \mathbf{S0}$, is equivalent to reflection of Π_1 statements, denoted by $\mathrm{Refl}_T^{\Pi_1}$.

Consistency statements play a major role in the incompleteness theorems. Each incompleteness result necessitates a consistency statement assumption on the considered theory. Plain consistency is the weakest of these statements. Gödel introduced ω-consistency to be able to obtain an independent statement. The weaker assumption, 1-consistency, generally suffices. In all our incompleteness theorems and proofs, the strongest assumption is 1-consistency. For more details on consistency and reflection statements, the reader is referred to [Smo77].

2. Original and model-theoretical proofs

For various descriptions of mathematical logic in general and Gödel's incompleteness theorems in particular, see [Smo77, Kle52, Fef60, Kre50, Kot94, Kot96, Kot98, Kot04, Ros36, Boo95, End72, Hen57].

2.1. Original (syntactical) proof

The original proof of Gödel's incompleteness theorems goes necessarily through proving the *diagonalization lemma.*

Lemma 2.1 (Diagonalization lemma)**.** *For every formula ψ with a single free variable x there is a sentence ϕ such that $S \vdash \phi \equiv \psi(\ulcorner \phi \urcorner)_x$.*

Proof. Given ψ, let θ_x be $\psi(\mathrm{sub}_x(\mathrm{var}(x), x))_x$, the diagonalization of ψ. Let $m = \ulcorner \theta_x \urcorner$ and $\phi = \theta(m)_x$. Then we have

$$S \vdash \phi \equiv \psi(\ulcorner \phi \urcorner)_x$$

In S, we have that

$$\phi \equiv \theta(m)_x \equiv \psi(\mathrm{sub}_x(\mathrm{var}(m), m))_x \equiv \psi(\mathrm{sub}_x(\mathrm{var}(m), \ulcorner \theta_x \urcorner))_x \equiv \psi(\ulcorner \theta(m)_x \urcorner)_x \equiv \psi(\ulcorner \phi \urcorner)_x$$

■

The[4] Gödel sentence G_T for T consists in diagonalizing $\neg\,\mathrm{Proof}_T(\cdot)$. By the diagonalization lemma, we have a sentence G_T such that $G_T \equiv \neg\,\Box\, G_T$ is provable in S.

Theorem 2.2 (Gödel's first incompleteness theorem)**.** *If T is consistent, G_T is not provable in T, and if T is Σ-sound, then G_T is independent of T.*

Proof. If G_T is provable in T, then $\Box G_T$ is also provable by the first deducibility condition. By definition of G_T, we thus have that $\neg G_T$ is provable in T, so T is inconsistent.

If $\neg G_T$ is provable in T, either T is inconsistent and thus not Σ-sound, or if T is consistent, $\neg G_T$ is false (since G_T is true: we have just proved that "G_T is not provable", which is equivalent in S to G_T) and again T is not Σ-sound, since $\neg G_T$ is equivalent in S (and thus in T) to $\Box G_T$, which is equivalent in S to a Σ-formula. ■

[4] There is no uniquely defined Gödel sentence for a theory T, because the sentences depend on the Σ-formula used to define the axioms of T. It is an abuse of language.

This proof is formalizable in T: $\Box(G_T \rhd \neg \Box G_T)$ is provable in T by definition of G_T and 1.1, so $\Box G_T \rhd \Box \neg \Box G_T$ is provable in T by 1.3, and $\Box G_T \rhd \Box \Box G_T$ is provable in T by 1.2, so $\Box G_T \rhd (\Box \Box G_T \wedge \Box \neg \Box G_T)$ is provable in T. Hence, $\Box G_T \rhd \neg \mathrm{Cons}_T$ is provable in T.

Thus, Cons_T implies $\neg \Box G_T$ (and also G_T) in T.

This yields the second incompleteness theorem:

Theorem 2.3 (Gödel's second incompleteness theorem)**.** *If T is consistent,* Cons_T *is not provable in T.*

Proof. The formalization of theorem 2.2 shows that Cons_T implies G_T in T and by the same theorem 2.2, G_T cannot be provable in T, and thus in T neither can Cons_T.

The implication $G_T \rhd \mathrm{Cons}_T$ is also provable in T, since "if T is inconsistent, every formula is provable in T" is provable in T. Thus, G_T and Cons_T are in fact equivalent in T. ■

The second incompleteness theorem can also be strengthened:

Theorem 2.4 (Löb's theorem)**.** *If ϕ is a sentence for which $\Box\phi \rhd \phi$ is provable in T, then ϕ is provable in T.*

Kreisel's proof. If $\Box\phi \rhd \phi$ is provable in T, then $T + \neg\phi \vdash \neg \Box \phi$, which is equivalent in T to $\mathrm{Cons}_{T+\neg\phi}$. Thereby, $T + \neg\phi$ proves its own consistency, and so by theorem 2.3 is inconsistent. Thus ϕ is a theorem of T. ■

Löb's original proof. Suppose that $\Box\phi \rhd \phi$ is provable in T and let ψ be the diagonalization of $\Box x \rhd \phi$: the diagonal lemma gives ψ such that $\psi \equiv (\Box\psi \rhd \phi)$ is provable in T. Thus we have by 1.1 and two applications of 1.3 that

$$\Box\psi \rhd (\Box \Box \psi \rhd \Box\phi) \text{ is provable in } T.$$

By 1.2, we obtain that $\Box\psi \rhd \Box\phi$ is provable in T and also by the assumption on ϕ,

$$\Box\psi \rhd \phi \text{ is provable in } T. \tag{2.1}$$

By definition of ψ, it follows that ψ is provable in T. All of this has been proven in T, thus $\Box\psi$ is provable in T, and by 2.1, ϕ is provable in T. ■

Rosser's theorem is a variant of Gödel first incompleteness theorem dropping the soundness condition on T to obtain an independent statement. It uses a modification of Proof_T.

$$\mathrm{Proof}^R_T(x, \ulcorner y \urcorner) \text{ iff } \mathrm{Proof}_T(x, \ulcorner y \urcorner) \wedge \forall z, \ulcorner w \urcorner \leqslant x,\ (\mathrm{Prov}_T(z, \ulcorner w \urcorner) \rhd y \neq \neg w)$$

From Proof^R_T, one defines Prov^R_T and Cons^R_T.

Theorem 2.5 (Rosser's theorem)**.** *Let ϕ be a sentence by the diagonalization lemma such that $S \vdash \phi \equiv \neg \mathrm{Prov}^R_T(\ulcorner \phi \urcorner)$. Then*

(1) *$T \nvdash \phi$;*
(2) *$T \nvdash \neg\phi$;*
(3) *$T \vdash \mathrm{Cons}^R_T$.*

2.2. Semantical proofs

By semantical proofs, we mean "more model-theoretical" proofs.

2.2.1. *$G_T \equiv \mathrm{Cons}_T$.* From the first syntactical incompleteness theorem, one can obtain a model theoretical proof of the second incompleteness theorem: a model theoretical way to prove the equivalence of G_T and Cons_T. It is a forerunner of the ideas underlying the subsequent model-theoretical proofs of both incompleteness theorems.

Model-theoretical proof of $G_T \equiv \mathrm{Cons}_T$. Suppose that we have a model $\mathcal{M}$ of $T + \mathrm{Cons}_T$ such that $\mathcal{M} \models \neg G_T$.

Since $\mathcal{M} \models \mathrm{Cons}_T$, Henkin's completeness theorem gives a Δ_2 model $\mathcal{M}'$ such that $\mathcal{M}' \models T$.

$\mathcal{M} \models \neg G_T$, thus $\mathcal{M} \models \Box G_T$ and by Henkin's construction, $\mathcal{M}' \models G_T$.

In $\mathcal{M}$, we define a function which to $x \in \mathcal{M}$ gives $x_{\mathcal{M}'}$, the x-th successor of $\mathbf{0}$ in the sense of $\mathcal{M}'$. Let $\mathcal{M}''$ be the image of $\mathcal{M}$ by this function; it is an initial segment of $\mathcal{M}'$.

G_T is Π_1 and $\mathcal{M}' \models G_T$, thus so does $\mathcal{M}''$ and $\mathcal{M}$. *Contradiction* with $\mathcal{M} \models \neg G_T$. ■

2.2.2. *Chaitin's proofs of theorem 2.2.* Let $\{\varphi_i\}_{i\in\mathbb{N}}$ designate a recursive enumeration of all partial recursive functions. We work with an *acceptable* enumeration φ, in which all classical computability results hold (enumeration, *s-m-n*, fixed point, *etc.*).

For the purpose of proving Chaitin's incompleteness theorem, the following simple definition of *Kolmogorov complexity* is sufficient.

$$K_\varphi(x|y) = \text{ smallest } e \text{ such that } \varphi_e(y) = x, \text{ and } K_\varphi(x) = K_\varphi(x|0)$$

A more classical definition of Kolmogorov complexity goes as follows. A complexity is defined according to a *decompressor*, giving the length of a smallest input to the decompressor yielding the sought string. The Kolmogorov complexity is then the complexity according to an *optimal* decompressor; *optimal* in the sense that it differs only by an additive constant from other decompressors. For more on classical Kolmogorov complexity, see [LV90]. Our definition is merely a change of scale from the classical one. Both share the same basic computability properties: a computable function which is a lower bound for them is necessarily bounded and their graphs are Turing-complete. We call these functions the Kolmogorov functions.

Theorem 2.6 (Chaitin's theorem)**.** *Let T be a arithmetical-able sound theory. There is a constant $\mathfrak{c}_T$ such that T does not prove "$K_\varphi(x) > \mathfrak{c}_T$" for any x.*

Proof. Let f be the recursive function assigning to c the code m of a the Turing machine M, such that M enumerates the theorems of T, searches for a theorem of the form "$K_\varphi(x) > c$" and in case of success, outputs x.

By Kleene's recursion theorem there is an e such that $\varphi_e = \varphi_{f(e)}$. Suppose that T_e halts when started with input 0. T_e outputs x such that $K_\varphi(x) > e$ because of the soundness assumption. On the other hand, if T_e outputs x with input 0, then $K_\varphi(x) \leqslant e$ by the definition of K_φ. *Contradiction.*

Hence, T_e does not halt and thus there is no proof of "$K_\varphi(x) > e$" for any x. Thereby $\mathfrak{c}_T = e$ works. ■

Chaitin has given many other variants of this proof. Another proof goes by the observation that if no such $\mathfrak{c}_T$ existed, then there would exist an unbounded lower bound function of the Kolmogorov complexity function.

2.2.3. *Other proofs.* Many other *model-theoretical* proofs of the incompleteness theorems have appeared.

In [Vop66], Vopěnka proved theorem 2.3 for Bernays-Gödel axiomatic set theory using Richard's paradox: "the least number not definable in 1000 words". More recently, in [Jec94], Jech gave a short proof of theorem 2.3 for set theory.

In [Kre68], Kreisel gave the first proof of theorem 2.3 using Henkin's arithmetized completeness theorem.

In [Boo89b, Boo89a], Boolos proved both theorems 2.2 and 2.3 using both model-theoretical techniques and Berry's paradox.

3. Incompleteness revisited

From Henkin's proof of the completeness theorem, one can derive the *arithmetized completeness theorem.* It is an important result that is essential for constructing arithmetical models and thus for proving Gödel's second incompleteness theorem.

The arithmetized completeness theorem asserts that any recursively axiomatizable consistent theory has an arithmetically definable model. We say that a formula ϕ in $\mathcal{L}_{\mathrm{PA}}$ *defines a model of T in a theory S in* $\mathcal{L}_{\mathrm{PA}}$ if we can prove within S that the set $\{\sigma : \sigma \text{ is a sentence in } \mathcal{L}_T \cup C \text{ that satisfies } \phi(\ulcorner\sigma\urcorner)\}$, where C is a set of new constants, forms an elementary diagram of a model of T with a universe from C.

Theorem 3.1 (Hilbert-Bernays arithmetized completeness theorem)**.** *There exists a Δ_2 formula Tr_T in $\mathcal{L}_{\mathrm{PA}}$ that defines a model of T in* $\mathrm{PA}+\mathrm{Cons}_T$*.*

The following is a corollary of this theorem: if $\mathcal{M}_0$ is a model of $\mathrm{PA}+\mathrm{Cons}_T$, then there exists a model $\mathcal{M}_1$ of T such that

(1) for any sentence ϕ in $\mathcal{L}_{\mathrm{PA}}$, $\mathcal{M}_1 \models \phi$ if and only if $\mathcal{M}_0 \models \mathrm{Tr}_T(\ulcorner\phi\urcorner)$,
(2) for any Σ_1 sentence ϕ in $\mathcal{L}_{\mathrm{PA}}$, if $\mathcal{M}_0 \models \phi$, then $\mathcal{M}_1 \models \phi$.

We then say that $\mathcal{M}_1$ is a model of T *definable* in a model of $\mathcal{M}_0$ of $\mathrm{PA}+\mathrm{Cons}_T$ and write $\mathcal{M}_1 \prec_d \mathcal{M}_0$.

3.1. Incompleteness in computability

3.1.1. *From $\mathcal{K} = \{x : \varphi_x(x)\downarrow\}$ or similar.* Using the same basic arguments we have used in the introduction, if we consider any non-recursive recursively enumerable set L, then given a Π_1-sound ($\equiv$ consistent) theory T, there is an $\mathfrak{n}_T^L \notin L$ such that T does not prove it. This is a form of the first incompleteness theorem.

Using the arithmetized completeness theorem 3.1, the second incompleteness theorem ($T \vdash 1\text{-}\mathrm{Cons}_T \rhd\neg \mathrm{Prov}_T(\mathrm{Cons}_T)$) can be proved as follows:

$\mathcal{K}$-related proof of a variant of theorem 2.3. We assume that Cons_T is derivable from T. Then by the completeness theorem, there exists a model $\mathcal{M}_0$ of T.

If $\mathcal{M}_0 \models \mathrm{Prov}_T(\ulcorner 0 \in L \urcorner)$, then let $\mathcal{M}_1 = \mathcal{M}_0$. Otherwise $\mathcal{M}_0 \models \neg \mathrm{Prov}_T(\ulcorner 0 \in L \urcorner)$ and thus $\mathcal{M}_0 \models \mathrm{Cons}_{T+0\notin L}$. Hence, by the Hilbert-Bernays arithmetized completeness theorem, there exists $\mathcal{M}_1 \prec_d \mathcal{M}_0$ such that

$$\text{either } \mathcal{M}_1 \models \mathrm{Prov}_T(\ulcorner 0 \in L \urcorner) \text{ (in case } \mathcal{M}_1 = \mathcal{M}_0\text{) or } \mathcal{M}_1 \models \mathrm{Prov}_T(\ulcorner 0 \notin L \urcorner).$$

We iterate this construction.

Consider the final model $\mathcal{M}_{\mathfrak{n}_T^L}$, it satisfies $\neg\,\mathrm{Cons}_T$ by the previous form of the first incompleteness theorem: If we have $\mathcal{M}_{\mathfrak{n}_T^L} \models \mathrm{Prov}_T(\ulcorner \mathfrak{n}_T^L \in L \urcorner)$, then the 1-consistency of T is contradicted by the fact that $\mathfrak{n}_T^L \notin L$, since it is a Σ_1 statement. Hence, $\mathcal{M}_{\mathfrak{n}_T^L} \models \mathrm{Prov}_T(\ulcorner \mathfrak{n}_T^L \notin L \urcorner)$ which implies the non-consistency of T, since $\mathfrak{n}_T^L$ is such that $\mathfrak{n}_T^L \notin L$ is not provable in T. ∎

3.1.2. *From computability functions.* In the sixties, Tibor Radó, a professor at the Ohio State University, thought of a simple non-computable function besides the standard halting problem for Turing machines. Given a fixed finite number of symbols and states, select those Turing machines which eventually halt when run with a blank tape. Among these programs, find the maximum number of non-blank symbols left on the tape when they halt. Alternatively, find the maximum number of time steps before halting. These functions are well-defined but uncomputable. Tibor Radó called them the Busy Beaver functions. For more on the Busy Beaver problem, read [Rad62, Lin63, LR65, Bra66, Bra83, Dew84, Dew85, Her88, MS90, MB90, LP07].

Alternative functions can be defined that are close in nature to these Busy Beaver functions. Let σ^{steps} be the function which to i gives the maximum number of steps for which a Turing machine with code $\leqslant i$ will keep running before halting starting with a blank tape. For a Turing machine M, t_M denotes the time complexity function of M: $t_M(x) = s$ if $M(x)$ halts after s steps. Following the Busy Beaver functions' definitions, we define σ^{value} to be the function which to i gives the maximum number which a Turing machine with code $\leqslant i$ will output, following a fixed convention, after halting starting with an input $\leqslant i$. These functions are in a sense inverses of the K_φ function.

Other functions can be defined following classical Kolmogorov complexity, *e.g.*, the function which to n gives the biggest number with Kolmogorov complexity lower than n.

We call these functions the σ functions. For each variant, we can define a function focusing on maximizing the *number of steps*, *e.g.*, σ^{steps}, or the *outputted values*, *e.g.*, σ^{value}. The value of one of the functions on a certain x is computable from x and the value of the other function on input $x + c$ for a certain constant c.

A result similar to Chaitin's result (see section 2.2.2) can be obtained concerning the σ functions:

Theorem 3.2 (Chaitin-like incompleteness theorem for σ functions)**.** *Let σ be one of the σ functions. Let T be an arithmetical-able consistent theory. There is a constant $\mathfrak{n}_T^\sigma$ such that*

$$T \vdash \mathrm{Cons}_T \rhd \forall s \neg\,\mathrm{Prov}_T(\ulcorner \sigma(\mathfrak{n}_T^\sigma) < s \urcorner). \tag{3.1}$$

Proof. Consider a Π_1 formula ϕ_σ in the language of T such that $\phi_\sigma(x, s)$ expresses that $\sigma(x) < s$.

Working in T, for a given x, take the smallest s such that $\mathrm{Prov}_T(\ulcorner \phi_\sigma(x,s) \urcorner)$ holds. T being consistent and ϕ_σ Π_1, $\phi_\sigma(x,s)$ also holds.

$\mathrm{Prov}_T(\ulcorner \phi_\sigma(x,s) \urcorner)$ is a Σ_1 formula and thus can be seen as $\exists y \psi(x,s,y)$ or equivalently $\exists \langle s, y \rangle \psi(x,s,y)$ where ψ is Δ_0.

Thus there is a Turing machine computing ψ. Consider its code i_ψ (or its number of states or transitions, depending on the choice of σ). For large enough x, *i.e.*, $x > i_\psi + c$, knowing that $\phi_\sigma(x,s)$ holds (using the computation through shifting, *i.e.*, the constant c, between both types of σ functions), we know that $\sigma(x) < s$ and thus there is an $s' = \langle s'_1, s'_2 \rangle < s$ such that $\psi(x, s'_1, s'_2)$ holds. But for each $s' = \langle s'_1, s'_2 \rangle$ smaller than s, the statement $\neg\psi(x, s'_1, s'_2)$ is true by the minimality of s, and provable (being Δ_0). Thus we have $\neg \mathrm{Cons}_T$. ∎

It is also possible to go through the same proof but using Kolmogorov functions (classical Kolmogorov complexity function C, K_φ, $\ldots$) and a variant of Chaitin's theorem 2.6, following closely the proof of theorem 3.2: Let K be a Kolmogorov function. If T is consistent, then there exists $\mathfrak{n}_T^K$ such that for all x, $T \nvdash K(x) > \mathfrak{n}_T^K$. Moreover, if T is ω-consistent, then for all x, if $K(x) > \mathfrak{n}_T^K$, then $T \nvdash K(x) \leqslant \mathfrak{n}_T^K$.

From there, we can show Gödel's second incompleteness theorem in a model-theoretical way.

Model-theoretical proof of theorem 2.3 using σ functions. We have supposed that T is consistent. So, let $\mathcal{M}_0$ be a model of T.

If $\mathcal{M}_0 \models \mathrm{Prov}_T(\ulcorner \forall x\ \neg t_{T_0}(0) = x \urcorner)$, then let $\mathcal{M}_1 = \mathcal{M}_0$. Otherwise,

$$\mathcal{M}_0 \models \neg \mathrm{Prov}_T(\ulcorner \neg\exists x\ t_{T_0}(0) = x \urcorner)$$

Thereby, $\mathcal{M}_0 \models \mathrm{Cons}_{T+\exists x\ t_{T_0}(0)=x}$. Hence there exists, by theorem 3.1, $\mathcal{M}_1 \prec_d \mathcal{M}_0$ such that either $\mathcal{M}_1 \models \mathrm{Prov}_T(\ulcorner \neg\exists x\ t_{T_0}(0) = x \urcorner)$ (when $\mathcal{M}_1 = \mathcal{M}_0$), or $\mathcal{M}_1 \models \exists x\, \mathrm{Prov}_T(\ulcorner t_{T_0}(0) = x \urcorner)$ (because $\{(i,x) : t_{T_i}(0) = x\}$ is Δ_0, in other words primitive recursive, and thus its truth in $\mathcal{M}_1$ implies its provability).

We iterate this construction (consider now $\mathcal{M}_1$ and "$t_{T_1}(0) = x$", instead of $\mathcal{M}_0$ and "$t_{T_0}(0) = x$"; in i-th iteration, consider $\mathcal{M}_i$ and "$t_{T_i}(0) = x$").

Consider the last model $\mathcal{M}_{\mathfrak{n}_T^\sigma}$, the model constructed after the $\mathfrak{n}_T^\sigma$-th iteration. This model satisfies $\forall i \leqslant \mathfrak{n}_T^\sigma\ \mathrm{Prov}_T(\ulcorner \neg\exists x\ t_{T_i}(i) = x \urcorner) \vee \exists x\, \mathrm{Prov}_T(\ulcorner t_{T_i}(i) = x \urcorner)$ by theorem 3.1. In this model, for each $i \leqslant \mathfrak{n}_T^\sigma$ such that the second case holds ($\exists x\, \mathrm{Prov}_T(\ulcorner t_{T_i}(i) = x \urcorner)$), we take the smallest appropriate x and choose s to be greater than all these x's. Thus, this model satisfies the provability in T of $\sigma(\mathfrak{n}_T^\sigma) < s$ and thus satisfies $\neg \mathrm{Cons}_T$ by (3.1). ∎

This argument can be carried out for other functions than σ. In particular for the variants of the Busy Beaver functions.

The second incompleteness theorem 2.3 can also be proved in this manner from the above Kolmogorov function variant of theorem 3.2.

It is an open question to carry out this type of argument (for proving both incompleteness theorems) for bizarre functions derived from the Busy Beaver functions, defined in [LP07], *e.g.,* consider the function giving the parity of one of the Busy Beaver functions. One of the obvious missing properties of these functions is unboundedness.

3.1.3. *Giving its own relative consistency.* We say that a statement ϕ is a *revelation for* T if ϕ is unprovable in T and its consistency relative to T (if T is consistent, so is $T + \phi$) is provable from itself in T:

$$T \vdash \phi \rhd \mathrm{Cons}_T(\phi)$$

The links between incompleteness and computability functions described above in section 3.1.2 have yielded the following serendipitous result.

Theorem 3.3 (Serendipitous incompleteness theorem for σ functions)**.** *Let σ be one of the σ functions. If T is consistent, then there exists a natural number $\mathfrak{r}_T^\sigma$ such that for all x, $\sigma(\mathfrak{r}_T^\sigma) < x$ is a revelation for T.*

Proof. Consider the Π_1 statement $\forall x\ \psi(x)_x$ equivalent to $\mathrm{Cons}_{T+\phi}$.

$\psi \in \Delta_0$ and thus there is a machine M_ψ with code i_ψ such that M_ψ decides $\{x : \psi(x)_x\}$: M_ψ on input x eventually enters an acceptance state if $\psi(x)_x$, or a rejection state otherwise.

Consider another Turing machine M'_ψ which runs M_ψ successively on each natural number starting from 0 and stops and writes the counter example of ψ if the simulation of M_ψ enters a rejection state.

Let i'_ψ be the code of Turing machine M'_ψ. $\sigma(i'_\psi)$ makes the verification of $\forall x\ \psi(x)_x$ a Δ_0 property.

By using Kleene's recursion theorem on this previous construction, we find $\mathfrak{r}_T^\sigma$ such that knowing (or bounding) the value of $\sigma(\mathfrak{r}_T^\sigma)$ makes the verification of $\mathrm{Cons}_{T+\sigma(\mathfrak{r}_T^\sigma)\leqslant x}$ a Δ_0 property. Knowing that T is consistent and assuming $\sigma(\mathfrak{r}_T^\sigma) \leqslant x$, T thus proves $\mathrm{Cons}_T(\sigma(\mathfrak{r}_T^\sigma) \leqslant x)$.

By Gödel second incompleteness theorem, $\sigma(\mathfrak{r}_T^\sigma) \leqslant x$ is an unprovable statement in T. ∎

3.2. Interpretations of Chaitin's theorem

Chaitin's famous version of Gödel's first incompleteness theorem (see section 2.2.2) is compelling for various obvious reasons. Firstly a statement of the type "the Kolmogorov complexity of this integer is greater than that integer" looks more mathematically natural than a consistency statement and secondly it gives a bound on the provable complexity of objects in a given theory. The question that arises forthrightly is the relevance of this bound to measure the *complexity*, the *power*, or *information content* of a theory.

We will now discuss the validity of the common way of interpreting Chaitin's theorem. Many people have addressed criticisms towards this interpretation. In particular see [Fal96, Raa98]. We try to sum up these criticisms here.

Chaitin's result, theorem 2.6, has been interpreted to show that in a formalized theory one cannot prove an object to be more complex than the complexity of the theory itself. This received interpretation claims that the limiting constant $\mathfrak{c}_T$ is determined by the complexity of the theory T itself and is a good measure of the strength of the theory.

As Chaitin puts it in [Cha82]: "I would like to measure the power of a set of axioms and rules of inference. I would like to be able to say that if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms."

It is assumed here that the algorithmic complexity of the axioms gives a good measure of the *power*, or *information content*, of the theory. The constant $\mathfrak{c}_T$ is assumed to depend on the complexity of the axioms of T. The finite bound given by the constant $\mathfrak{c}_T$ is hence thought to reflect the *power*, or *information content*, of the theory.

By playing with Kleene's fixed point theorem, for any suitable theory T, one can construct acceptable enumerations of partial recursive functions yielding constants $\mathfrak{c}_T$ equal to 0 or arbitrarily large, whatever the theory T.

A closer inspection shows that the value of $\mathfrak{c}_T$ 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. It is really hard to see why the code of such a Turing machine would reveal anything interesting about the *power* or *information content* of T.

Considering a strong theory like ZFC, *Zermelo-Fraenkel set theory with the axiom of choice*, we could compare its constant $\mathfrak{c}_{\mathrm{ZFC}}$ to the constant of a weak theory, say PA, *Peano Arithmetic.* The constants depend on our acceptable enumeration of partial recursive functions. Thus, suppose we have $\mathfrak{c}_{\mathrm{ZFC}} > \mathfrak{c}_{\mathrm{PA}}$. We can then add to *PA* all true sentences of the form $\neg\exists x\ \varphi_e(0) = x$ which are provable in ZFC, for all $e < \mathfrak{c}_{\mathrm{ZFC}}$. It follows from a result of Kreisel and Levy [KL68] that this new theory cannot possibly come even close to the power of ZFC. But the constants of this new theory and ZFC are now equal, and hence, they should, according to the received interpretation, have the same *power.* Furthermore, we may still add to our new theory one more true sentence $\neg\exists x\ \varphi_{\mathfrak{c}_{\mathrm{ZFC}}}(0) = x$. Now the constant of this theory is bigger than the one of ZFC. This whole argumentation shows that one has to be careful on the interpretation given to the constants $\mathfrak{c}_T$'s.

As mentioned in [Fal96], the only thing that these constants could at most tell of a theory is what propositions of the form "$K(\cdot) > \cdot$" it can prove. Therefore, withstanding all the above arguments, one could wonder whether adding as an axiom a sentence of the form "$K(x) > c$" could not be equivalent to the relative 1-consistency of a strong (consistency-wise) unprovable statement or even to the 1-consistency of a theory. For example, the 1-consistency of a large cardinal axiom[5] would also only add "information" about some of the propositions and be incredibly weaker than the large cardinal axiom itself. This would give credit to Chaitin's interpretation of his theorem and present the constant $\mathfrak{c}_T$ as a *partial* measure of the *power* of a theory T.

Having a link between consistency (or soundness) and computability ($\mathcal{K}$, Busy beaver functions or Kolmogorov complexity) would make possible an understanding of what properties consistency adds to a theory. Adding consistency as an axiom would then yield new combinatorial properties. Until now, consistency has been seen as a strange statement, only considered because of Gödel's second incompleteness theorem. It is true that even if one can construct stronger theories by adding as a new axiom its consistency, it is not clear in what way the obtained theory is stronger. It could well be that the only additional *information* this new theory has is this consistency statement and that nothing else is added because of this additional axiom. In fact, as mentioned in the introduction of this paper, one would need to *add*[6] $\omega^{\omega^{\omega+1}}$ times a reflection principle to our theory to cover all true arithmetic statements.

Taking into account the above arguments, we see that the only *total* measure one could get of a theory through Chaitin's theorem 2.6 would not be a constant $\mathfrak{c}_T$ but a set $\mathcal{C}_T$ of constants for which Chaitin's proposition is unprovable in T. If we want the above objections not to apply (in particular modifying by Kleene's fixed point theorem the

[5] See [KM78, Kan94].

[6] For any uniform reflection progression $\{T_a\}_{a\in\mathbf{O}}$, there is a branch B in an ordinal notation system $\mathbf{O}$, such that there is, for any true arithmetical sentence ϕ, an a in B with $|a| < \omega^{\omega^{\omega+1}}$ for which ϕ is provable in T_a. For details, see [Fef62, FS62].

acceptable enumeration with which we work), $\mathcal{C}_T$ should necessarily be infinite and non-recursive. This set could not then be used to form an additional axiom if we want our theory to stay recursively axiomatizable.

References

[Boo89a] Boolos (G.), « A letter from George Boolos », *Notices of the American Mathematical Society*, vol. 36, 1989, p. 676.

[Boo89b] Boolos (G.), « A new proof of the Gödel incompleteness theorem », *Notices of the American Mathematical Society*, vol. 36, 1989, p. 383–391.

[Boo95] Boolos (G.), *The Logic of Provability.* Cambridge University Press, Cambridge, 1995.

[Bra66] Brady (A. H.), « The conjectured highest scoring machines for Rado's $\sigma(k)$ for the value $k = 4$ », *IEEE Transactions on Elec. Comput.*, vol. EC-15, 1966, p. 802–803.

[Bra83] Brady (A. H.), « The determination of the value of Rado's noncomputable function $\sigma(k)$ for four-state Turing machines », *Mathematics of Computation*, vol. 40, 1983, p. 647–665.

[Cha82] Chaitin (G.), « Gödel's theorem and information », *International Journal of Theoretical Physics*, vol. 22, 1982, p. 941–954.

[Dev84] Devlin (K. J.), *Constructibility.* Springer, 1984.

[Dew84] Dewdney (A. K.), « Computer recreations: A computer trap for the busy beaver, the hardest-working Turing machine », *Scientific American*, vol. 251, n° 2, August 1984, p. 19–23.

[Dew85] Dewdney (A. K.), « Computer recreations », *Scientific American*, vol. 252, n° 4, April 1985, p. 12–16.

[End72] Enderton (H. B.), *A Mathematical Introduction to Logic.* Academic Press, New York, 1972.

[Fal96] Fallis (D.), « The source of Chaitin's incorrectness », *Philosophia Mathematica*, vol. 3, n° 4, 1996, p. 261–269.

[Fef60] Feferman (S.), « Arithmetization of metamathematics in a general setting », *Fundamenta Mathematicae*, vol. 49, 1960, p. 35–92.

[Fef62] Feferman (S.), « Transfinite recursive progressions of axiomatic theories », *Journal of Symbolic Logic*, vol. 27, n° 3, 1962, p. 259–316.

[FS62] Feferman (S.) et Spector (C.), « Incompleteness along paths in progressions of theories », *Journal of Symbolic Logic*, vol. 27, n° 4, December 1962, p. 383–390.

[Göd31] Gödel (K.), « Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I », dans *Monatshefte für Mathematik und Physik* [SFKM$^+$86], p. 173–198.

[HB39] Hilbert (D.) et Bernays (P.), *Grundlagen der Mathematik*, vol. 2. Springer, 1939.

[Hen57] Henkin (L.), « A generalization of the concept of ω-completeness », *Journal of Symbolic Logic*, vol. 22, n° 1, March 1957, p. 1–14.

[Her88] Herken (R.), *The Universal Turing Machine: A Half-Century Survey.* Oxford University Press, Oxford, England, 1988.

[Hod93] Hodges (W.), *Model theory*, vol. 42 (coll. *Encyclopedia of Mathematics and its Applications*). Cambridge University Press, Cambridge, 1993.

[Jec78] Jech (T.), *Set Theory.* Academic Press, New York, 1978.

[Jec94] Jech (T.), « On Gödel's second incompleteness theorem », *Proceedings of the American Mathematical Society*, vol. 121, 1994, p. 311–313.

[Kan94] Kanamori (A.), *The Higher Infinite.* Springer Verlag, 1994.

[Kay91] Kaye (R.), *Models of Peano arithmetic*, vol. 15 (coll. *Oxford Logic Guides*). Oxford University Press, Oxford, 1991.

[KL68] Kreisel (G.) et Lévy (A.), « Reflection principles and their use for establishing the complexity of axiomatic systems », *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, vol. 14, 1968, p. 97–142.

[Kle52] Kleene (S.), *Introduction to Metamathematics.* North-Holland Publishing, New York, 11th edition édition, 1952.

[KM78] Kanamori (A.) et Magidor (M.), « The evolution of large cardinal axioms in set theory », dans Muller (G. H.) et Scott (D. S.), éditeurs, *Higher Set Theory*, vol. 669 (coll. *Lecture Notes in Mathematics*), p. 99–275. Springer Verlag, Berlin, 1978.

[Kot94] Kotlarski (H.), « On the incompleteness theorems », *Journal of Symbolic Logic*, vol. 59, n° 4, 1994, p. 1414–1419.

[Kot96] Kotlarski (H.), « An addition to Rosser's theorem », *Journal of Symbolic Logic*, vol. 61, n° 1, 1996, p. 285–292.

[Kot98] Kotlarski (H.), « Other proofs of old results », *Mathematical Logic Quarterly*, vol. 44, 1998, p. 474–480.

[Kot04] Kotlarski (H.), « The incompleteness theorems after 70 years », *Annals of Pure and Applied Logic*, vol. 126, 2004, p. 125–138.

[Kre50] Kreisel (G.), « Notes on arithmetical models for consistent formulae of the predicate calculus », *Fundamenta Mathematicae*, vol. 37, 1950, p. 265–285.

[Kre68] Kreisel (G.), « A survey of proof theory », *Journal of Symbolic Logic*, vol. 33, 1968, p. 321–288.

[Laf02] Lafitte (G.), *Calculs et Infinis*. PhD thesis, École Normale Supérieure de Lyon, 2002.

[Lin63] Lin (S.), *Computer Studies of Turing Machine Problems*. PhD thesis, The Ohio State University, Colombus (Ohio), 1963.

[Löb55] Löb (M. H.), « Solution of a problem of Leon Henkin », *Journal of Symbolic Logic*, vol. 20, 1955, p. 115–118.

[LP07] Lafitte (G.) et Papazian (C.), « The fabric of small Turing machines », dans Cooper (S. B.), Kent (T. F.) et Benedikt Löwe (A. S.), éditeurs, *Computation and Logic in the Real World, Third Conference of Computability in Europe, CiE 2007*. 2007.

[LR65] Lin (S.) et Radó (T.), « Computer studies of Turing machine problems », *Journal of the Association for Computing Machinery*, vol. 12, n° 2, April 1965, p. 196–212.

[LV90] Li (M.) et Vitányi (P.), *Handbook of Theoretical Computer Science*, chap. Kolmogorov complexity and its applications, p. 187–254. Amsterdam, Elsevier, 1990.

[MB90] Marxen (H.) et Buntrock (J.), « Attacking the busy beaver 5 », *Bulletin of the EATCS*, vol. 40, 1990, p. 247–251.

[MS90] Machlin (R.) et Stout (Q. F.), « The complex behavior of simple machines », *Physica*, vol. 42D, 1990, p. 85–98.

[Odi89] Odifreddi (P.), *Classical Recursion Theory*. North Holland Publishing, 1989.

[Odi99] Odifreddi (P.), *Classical Recursion Theory*, vol. II. North Holland Publishing, 1999.

[Oll08] Ollinger (N.). « Universalities in cellular automata ». personal communication (submitted to JAC 2008), 2008.

[Raa98] Raatikainen (P.), « On interpreting Chaitin's incompleteness theorem », *Journal of Philosophical Logic*, vol. 27, 1998, p. 569–586.

[Rad62] Radó (T.), « On non-computable functions », *Bell System Technical Journal*, vol. 41, May 1962, p. 877–884.

[Rog58] Rogers (H.), « Gödel numberings of partial recursive functions », *Journal of Symbolic Logic*, vol. 23, 1958, p. 331–341.

[Rog67] Rogers (H.), *The Theory of Recursive Functions and Effective Computability*. MIT Press, 1967.

[Ros36] Rosser (J. B.), « Extensions of some theorems of Gödel and Church », *Journal of Symbolic Logic*, vol. 1, 1936, p. 87–91.

[SFKM$^+$86] S. Feferman (W. D.), Kleene (S.), Moore (G.), Solovay (R.) et van Heijendort (J.), éditeurs, *Kurt Gödel: Collected Works*, vol. 1. Oxford University Press, Oxford, 1986.

[Sho67] Shoenfield (J. R.), *Mathematical logic*. Addison-Wesley, Reading, Massachusetts, 1967.

[Smo77] Smorynski (C.), *Handbook of mathematical logic*, chap. The incompleteness theorems, p. 821–865. Amsterdam, North-Holland, 1977.

[Smu93] Smullyan (R. M.), *Recursion Theory for Metamathematics*. Oxford University Press, New York, 1993.

[Vop66] Vopěnka (A.), « A new proof of the Gödel's result of non-provability of consistency », *Bulletin de l'Académie Polonaise des Sciences*, vol. 14, n° 3, 1966, p. 111–116.

[VS03] Vereshchagin (N.) et Shen (A.), *Computable Functions*. American Mathematical Society, 2003.