Revisiting Chaitin's Incompleteness Theorem

Christopher P. Porter
Université Paris 7
LIAFA

Philosophy of Mathematics Seminar
Cambridge University
21 February 2013

Introduction

In his 1974 paper, "Information-Theoretic Limitations of Formal Systems", Gregory Chaitin proves a novel incompleteness theorem in terms of Kolmogorov complexity, a measure of complexity of finite strings.

In subsequent papers and books, Chaitin has made a number of claims of the significance of his incompleteness theorem (henceforth, CIT), for instance, that

(i) CIT shows that "if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms," and

(ii) CIT shows that the incompleteness phenomenon is "much more widespread and serious than hitherto suspected."

Chaitin's Critics

Chaitin's claims as to the significance of CIT have been subjected to much criticism:

- Chaitin's claims about "the amount of information of a formal theory" have been shown to be inaccurate (van Lambalgen 1989, Raatikainen 1998, Franzen 2005).
- Chaitin's claim that CIT shows that incompleteness is "much more widespread and serious than hitherto suspected" has also been severely criticized (Fallis 1996, Raatikainen 1998, 2001).

Why Revisit CIT? 1

Given these convincing refutations of Chaitin's claims, why bother revisiting CIT?

I have two main reasons for doing so:

(1) Recent work extending CIT may be construed as vindicating Chaitin's interpretation, at least in certain respects.

- One might argue on the basis of this work that Chaitin's undecidable statements have some sort of *priority* over other undecidable statements (when we restrict our attention to extensions of Peano arithmetic).
- I will argue that this argument for priority does not succeed.

Why Revisit CIT? 2

(2) I want to suggest an alternative account of the significance of CIT:

- On my account, CIT does not provide some sort of deep insight into the incompleteness phenomenon (beyond the work of Gödel, Turing, et al).
- Rather, CIT is one of a number of results that allow us to determine the formal costs associated with defining a sufficiently strong notion of randomness for finite strings.
- Seen in this light, CIT can be seen as part of a formal trade-off between the strength of a definition of randomness and certain properties that we might desire of a definition of randomness.

Outline

1. Formal Background

Kolmogorov Complexity: The Basic Idea, 1

In order to understand the statement and proof of CIT, we need to discuss the basics of Kolmogorov complexity of finite binary strings.

Suppose we fix a Turing machine M, viewed as a function from $2^{<\omega}$ to $2^{<\omega}$.

Suppose further that we want our machine M to output a given string σ; such a computation can be viewed as a construction of the string σ.

For each string τ such that $M(\tau)\downarrow = \sigma$, we can view τ as providing the blueprint for the construction of σ.

Kolmogorov Complexity: The Basic Idea, 2

Roughly, if σ has some regularity, then it should be fairly easy to construct. That is, using M, we should be able to construct σ from at least one short input.

- $\sigma = 000000000000\ldots$
- $\sigma = 010101010101\ldots$
- $\sigma = 010011000111\ldots$

Moreover, if σ lacks regularity, then it should be difficult to construct. That is, σ should not be constructible from any short inputs given to M.

The moral of the story: The complexity of σ is determined by the length of the shortest input given to M that yields the output σ.

Kolmogorov Complexity (relative to a machine M)

Let $M : 2^{<\omega} \to 2^{<\omega}$ be a Turing machine.

Definition

The *Kolmogorov complexity* of $\sigma \in 2^{<\omega}$ relative to M is

$$C_M(\sigma) = \min\{|\tau| : M(\tau) = \sigma\}.$$

(We set $C_M(\sigma) = \infty$ if σ is not in the range of M.)

Worry: For many Turing machines M, this does not appear to be a very meaningful notion.

Solution: Restrict to *universal* Turing machines.

Universal Turing Machines

We can effectively enumerate the collection of all Turing machines $\{M_i\}_{i \in \omega}$.

Then the function U defined by

$$U(1^e 0\sigma) = M_e(\sigma)$$

for every $e \in \omega$ and every $\sigma \in 2^{<\omega}$ is a *universal Turing machine*.

Note that there are many other ways to produce a universal Turing machine (by using a different enumeration of all Turing machines, by using a different mechanism for coding machines, etc.).

Kolmogorov Complexity

Let U be a universal Turing machine.

Definition

The *Kolmogorov complexity* of $\sigma \in 2^{<\omega}$ is
$$C(\sigma) = \min\{|\tau| : U(\tau) = \sigma\}.$$

Another worry: How can we justify the restriction to some fixed universal Turing machine when we could have chosen one of infinitely many other universal machines?

Optimality and Invariance

Theorem (The Optimality Theorem)

Let U be a universal Turing machine. Then for every Turing machine M, there is some $c \in \omega$ such that

$$C_U(\sigma) \leq C_M(\sigma) + c$$

for every $\sigma \in 2^{<\omega}$.

Consequently, we have:

Theorem (The Invariance Theorem)

For every two universal Turing machines U_1 and U_2, there is some $c_{U_1,U_2} \in \omega$ such that for every $\sigma \in 2^{<\omega}$,

$$|C_{U_1}(\sigma) - C_{U_2}(\sigma)| \leq c_{U_1,U_2}.$$

As we'll see shortly, this doesn't completely resolve the worry about our choice of universal machine.

Incompressible Strings

Since there is some Turing machine M such that $M(\sigma) = \sigma$ for every $\sigma \in 2^{<\omega}$, it follows that

$$C(\sigma) \leq |\sigma| + c$$

for some $c \in \omega$.

But observe that for every n, while there are 2^n strings of length n, there are

$$\sum_{i=0}^{n-1} 2^i = 2^n - 1$$

strings of length less than n.

Thus, for each n, there is at least one string τ of length n such that

$$C(\tau) \geq |\tau|.$$

We call such strings *incompressible*.

c-Incompressible Strings

Let $c \in \omega$. If σ satisfies

$$C(\sigma) \geq |\sigma| - c,$$

then we say that σ is *c-incompressible*.

For each n, there are at least $2^n - (2^{n-c} - 1)$ c-incompressible strings of length n.

Most Strings are Incompressible

Fix $n \in \omega$. Then from the last fact on the previous slide, one can show:

- At least $\frac{1}{2}$ of the strings of length n are 1-incompressible.
- At least $\frac{3}{4}$ of the strings of length n are 2-incompressible.

$$\vdots$$

- At least $1 - \frac{1}{2^c}$ of the strings of length n are c-incompressible.

Thus, if we want to produce a 10-incompressible string of length 100, by tossing a fair coin 100 times, we will obtain one with probability greater than 1023/1024.

Incompressibility and Randomness

In the theory of algorithmic randomness, one defines $\sigma \in 2^{<\omega}$ to be random if

$$C(\sigma) \approx |\sigma|.$$

But we should be cautious here: There is no precise dividing line between the random and non-random strings on this approach.

For instance, one can take all of the 1-incompressible strings to be the random strings, but why not also include the 2-incompressible ones, and so on?

Despite this lack of a precise dividing line between random and non-random strings, in the literature on algorithmic randomness one often finds the claim that the above definition of randomness is "absolute".

The Stability Problem

The Invariance Theorem implies that the complexity values assigned by two different universal machines can only differ by some fixed finite amount.

However, this doesn't guarantee that the class of c-incompressible strings remains stable under changes of the universal machine used to define Kolmogorov complexity.

For every $\sigma \in 2^{<\omega}$, there is a universal machine $U^{\uparrow}_{\sigma}$ such that

$$C_{U^{\uparrow}_{\sigma}}(\sigma) \geq |\sigma|.$$

For every $\sigma \in 2^{<\omega}$, there is a universal machine $U^{\downarrow}_{\sigma}$ such that

$$C_{U^{\downarrow}_{\sigma}}(\sigma) = 1.$$

I refer to this phenomenon as the *stability problem*.

Nonetheless...

Despite the stability problem, there is still a sense in which incompressible strings are the sort of strings we'd expect to be produced by a random process (one that outputs a 0 or 1 with equal probability).

- As we've seen, the vast majority of strings are c-incompressible for a fixed c.
- More significantly, Martin-Löf proved that the collection of c-incompressible strings coincides with the collection of strings that pass all computably enumerable statistical tests for randomness.

2. CIT: Its Proof and Chaitin's Interpretation

Representing Kolmogorov Complexity in a Theory

Let T be a computably axiomatizable theory that interprets Robinson arithmetic Q (that is, T is an $\mathcal{L}$-theory for some language $\mathcal{L}$ expressive enough to formulate the axioms of Q).

Further, let us fix some primitive recursive coding c of binary strings as natural numbers, which induces a primitive recursive map that sends the code of a string to its length.

By our assumption on T, there is a Σ^0_1 $\mathcal{L}$-formula $\psi(x, y)$ such that

$$U(\sigma) = \tau \text{ if and only if } \mathbb{N} \models \psi(c(\sigma), c(\tau)).$$

Thus, there is an $\mathcal{L}$-sentence $\phi_C(x, y)$ such that

$$C(\sigma) \geq n \text{ if and only if } \mathbb{N} \models \phi_C(c(\sigma), n).$$

Statement of CIT

Let us say that T is C-sound if

$$T \vdash \phi_C(c(\sigma), n) \text{ implies } \mathbb{N} \models \phi_C(c(\sigma), n).$$

Theorem

Let T be a computably axiomatizable, C-sound theory that interprets Q. Then there is some $N \in \omega$ such that

$$T \nvdash \phi_C(c(\sigma), N)$$

for any $\sigma \in 2^{<\omega}$.

That is, there is a threshold N such that T cannot prove of any individual string σ that it has complexity greater than N.

Proof Sketch

If the conclusion of the theorem does not hold, then for every $N \in \omega$, there is some $\sigma \in 2^{<\omega}$ such that

$$T \vdash \phi_C(c(\sigma), N).$$

Now we consider a machine M such that, given input a suitably chosen input, enumerates theorems of T until it finds a proof of $\phi_C(c(\sigma), k)$ for some sufficiently large k, and then outputs σ.

Since $T \vdash \phi_C(c(\sigma), k)$, by C-soundness, it follows that

$$C(\sigma) \geq k.$$

However, by virtue of being the output of M (with a carefully chosen input), it will also follow that

$$C(\sigma) < k,$$

yielding the desired contradiction.

The Proof of CIT

1. Suppose for every $N \in \omega$, there is some $\sigma \in 2^{<\omega}$ such that
$$T \vdash \phi_C(c(\sigma), N).$$
2. We define a Turing machine M as follows. Given any input τ, M looks for the first pair (σ, k) such that
   (i) $k > 2|\tau|$ and
   (ii) $T \vdash \phi_C(c(\sigma), k)$,

   and then M outputs σ.
3. Let $d \in \omega$ be such that
$$C(\sigma) \leq C_M(\sigma) + d$$
   for every $\sigma \in 2^{<\omega}$.
4. Now given input δ of length d, M outputs a string σ such that $T \vdash \phi_C(c(\sigma), k)$ for some $k > 2d$.
5. By C-soundness, this implies that
$$2d < C(\sigma) \leq C_M(\sigma) + d \leq d + d = 2d$$

Chaitin's Interpretation

"[I]f one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms."

Idea: For each theory T satisfying the conditions of the theorem, let N_T be the least natural number such that

$$T \nvdash \phi_C(c(\sigma), N_T)$$

for any $\sigma \in 2^{<\omega}$.

According to Chaitin, the number N_T can be seen as the measuring the information content of the theory T.

Moreover, Chaitin claims that any string σ such that $C(\sigma) > N_T$ has information content larger than N_T, and thus the above statement seems to follow.

A Serious Problem with Chaitin's Interpretation

The map $T \mapsto N_T$ does not depend entirely on T, but also on our choice of universal machine used to define Kolmogorov complexity.

What's more, due to this dependence of N on the choice of universal Turing machine, we have an analogue of the stability problem:

For a fixed theory T, we can make the constant N_T as small or as large as we like by changing the underlying universal machine.

In addition, one can construct universal machines U and U' such that

$$N_{PA}^{U} < N_{ZFC}^{U} \text{ and } N_{ZFC}^{U'} < N_{PA}^{U'}.$$

3. Vindicating Chaitin's Interpretation?

How Might One Vindicate Chaitin's Interpretation?

We've seen that Chaitin's claims about the information content of formal theories do not withstand scrutiny.

That is, we cannot appeal to the information content of a theory T to explain why there are statements that T does not decide.

Similarly, Chaitin's claims that CIT shows how "widespread and serious" the incompleteness phenomenon is have been shown to be exaggerated.

Still, one strategy for vindicating Chaitin's interpretation, at least in spirit, is to establish that there is a sense in which Chaitin's undecidable sentences are *prior to* or *more fundamental than* other undecidable sentences.

Π^0_1-Completeness

In very recent work of Bienvenu, Romashchenko, Shen, Taveneaux, and Vermeeren ("The Axiomatic Power of Kolmogorov Complexity"), we find a somewhat surprising result.

Let PA^* be the theory obtained by adding to Peano arithmetic all true statements of the form

$$C(\sigma) \geq n,$$

that is,

$$PA^* = PA \cup \bigcup \{\phi_C(c(\sigma), n) : \mathbb{N} \vDash \phi_C(c(\sigma), n)\}.$$

Theorem

For every true Π^0_1 sentence ϕ in the language of arithmetic,

$$PA^* \vdash \phi.$$

A Notion of Complexity for Π^0_1 Statements

Further, to each Π^0_1 statement ϕ, we can associate a string σ such that

$$PA \vdash \phi \leftrightarrow U(\sigma)\uparrow,$$

where U is a fixed universal Turing machine.

This yields a notion of complexity for Π^0_1 statements:

$$C(\phi) := \min\{|\sigma| : PA \vdash \phi \leftrightarrow U(\sigma)\uparrow\}.$$

"Local" Π^0_1-Completeness

Using this notion of complexity for Π^0_1 statements, Bienvenu et al. prove the following:

Theorem

For each $n \in \omega$, there is a string σ_n such that

$$PA + \phi_C(c(\sigma_n), n) \vdash \phi$$

for every true Π^0_1 statement ϕ with $C(\phi) \leq n - c$, where c is independent of n.

Further, this result is fairly resistant to the stability problem: by choosing a different universal Turing machine, this may result in a change of the collection $\{\sigma_n\}_{n\in\omega}$, but the statement still holds.

What Do These Theorems Tell Us?

One might be tempted to conclude from these theorems that there is a sense in which Chaitin's undecidable sentences uniformly "control" all other undecidable universal statements in PA.

Even if we were to grant this, it wouldn't follow that Chaitin incompleteness somehow explains or accounts for the incompleteness phenomenon in general.

In particular, these theorems have no bearing on undecidable Π^0_2 statements such as the Paris-Harrington Theorem.

Moreover, they are only applicable in the context of Peano arithmetic, whereas the incompleteness phenomenon occurs much more widely.

Upon Closer Examination...

But there is reason to be skeptical about this alleged priority of Chaitin undecidable sentences over other undecidable universal statements in the context of PA.

Let us look more closely at the special strings $\{\sigma_n\}_{n\in\omega}$.

σ_n is defined to be the first string σ of length n such that $C(\sigma) \geq n$.

Now let $t(n)$ be the number of steps needed to verify that $C(y) < n$ for all strings of length n preceding σ_n.

Then one can show for every τ with $|\tau| \leq n - c$, either $U(\tau)\downarrow$ in at most $t(n)$ steps, or $U(\tau)\uparrow$.

Upon Closer Examination...(continued)

Thus, the statements $\phi_C(c(\sigma_n), n)$ are so powerful because they encode finite chunks of information about the halting problem.

Further, we can exploit this information *in PA*, thus allowing us to derive all true Π^0_1 statements with complexity at most $n - c$ from $\phi_C(c(\sigma_n), n)$.

Lastly, most statements of the form $\phi_C(c(\tau), n)$ don't have this property: it appears that the only statements that give us this proof-theoretic strength are ones that encode the halting problem.

Thus, Chaitin's undecidable statements have no more priority than undecidable statements about which Turing computations fail to halt.

4. An Alternative Interpretation: A Formal Trade-Off

CIT and Definitions of Randomness

Discussions of the significance of CIT in the philosophical literature have focused on the question: What does CIT tell us about the incompleteness phenomenon in general?

Critics of Chaitin rightly point out that CIT is an interesting result but that it tells us nothing deep about the incompleteness phenomenon that we couldn't have already gathered from the work of Gödel, Turing, etc.

However, I want to suggest that CIT is an instance of a more general phenomenon one encounters in the task of providing definitions of randomness for individual objects such as finite strings.

The Classical Approach to Randomness

In classical probability and statistics, randomness is attributed to processes that generate certain outcomes, and then an individual string of events is counted as random in virtue of being produced by a random process.

For instance, on this approach, a random finite string is simply one that is obtained by some paradigm random process such as the repeated tosses of a fair coin.

With such a randomly obtained string, we can be reasonably certain that it satisfies those properties that are satisfied by a large majority of strings (e.g. roughly equal distribution of 0s and 1s, of the blocks 00, 01, 10, 11, etc.).

A Different Approach

Alternatively, we might first specify a collection of properties that are "typical", i.e. properties that are held by most randomly generated strings, and then define a string to be random if it satisfies all of those properties.

One problem with this approach is that it is notoriously difficult to isolate these "typical" properties.

In the theory of algorithmic randomness, one studies definitions that result under different formalizations of the class of "typical properties".

Comparing the Strength of Definitions of Randomness

As we vary the choice of "typical properties", definitions of randomness can vary in strength.

For instance, if we require of our random strings that they pass every computably enumerable statistical test for randomness, the resulting definition will be stronger than a definition given in terms of computable statistical tests.

As we include more and more properties among the "typical properties", the collection of strings counted as random will get smaller.

The Cost of a Strong Definition

For certain purposes, we might require a sufficiently strong definition of randomness.

But often this comes with a cost.

Many results in algorithmic randomness can be seen as showing the costs that are associated with working with sufficiently strong definitions of randomness.

CIT: One Consequence of a Strong Definition

CIT is precisely such a result.

If we require of our random sequences that they be incompressible by all Turing machines, which is equivalent to requiring that they pass all computably enumerable statistical tests for randomness, this comes at a cost:

We lose the ability to certify the randomness of our strings.

Certification of Randomness

What does it mean to certify the randomness of a string?

The certification of the randomness of a string is simply a formal proof of its randomness.

Thus, if we require of the random strings that they be incompressible, then string is certified as random if it is *provably* incompressible.

Moreover, certification should be carried out uniformly.

Each incompressible string is provably incompressible in *some* formal system, but we require certification to be carried out in one fixed formal system.

CIT Recast in Terms of Certifiability

Let T be a computably axiomatizable theory that interprets Q.

For any choice of universal machine U, for any choice of $\mathcal{L}$-formula to express U-computations, and for every $c \in \omega$, only finitely many c-incompressible strings are provably incompressible in T.

Notice: the stability problem has no bearing on this formulation of the result.

An Example of Certifiable Randomness

Not every definition of randomness for finite strings has this problem of certifiability.

For example, if we require of our random strings that they be incompressible within some time-bound, we get a completely different outcome:

Let $t : \omega \to \omega$ be a computable function. We define the t-bounded Kolmogorov complexity of $\sigma \in 2^{<\omega}$ to be

$$C^t(\sigma) = \min\{|\tau| : U(\tau) = \sigma \text{ in less than } t(|\sigma|) \text{ steps}\}$$

If $C^t(\sigma) \geq n$ is true, one can verify it in any computably axiomatizable theory that interprets Q.

Thus, CIT doesn't apply in this case.

In Conclusion

The significance of CIT can thus be understood in light of a trade-off for definitions of randomness.

On the one hand, if we require a sufficiently strong definition of randomness, then certain desiderata for our definition may have to be sacrificed.

On the other hand, if we don't want to sacrifice these desiderata, then we must be willing to accept a weaker definition of randomness, one that counts more strings as random than stronger definitions do.