GÖDEL'S COMPLETENESS AND INCOMPLETENESS THEOREMS

BEN CHAIKEN

Abstract. This paper will discuss the completeness and incompleteness theorems of Kurt Gödel. These theorems have a profound impact on the philosophical perception of mathematics and call into question the readily apparent strength of the system itself. This paper will discuss the theorems themselves, their philosophical impact on the study of mathematics and some of the logical background necessary to understand them.

Contents

1. Introduction

The completeness and incompleteness theorems both describe characteristics of true logical and mathematical statements. Completeness deals with specific formulas and incompleteness deals with systems of formulas. Together they serve to define how it is possible to ascribe the quality of truth to a mathematical or logical statement.

This has the interesting effect of illuminating how mathematical proofs acquire their strength, and also how it is possible for these proofs to lack unquestionable support. While this may not weaken or dismantle the general acceptance of mathematical "truths", it does raise some rather interesting thoughts on the philosophical underpinnings of the field of mathematics. These will be discussed after we look at the theorems themselves and the logical groundwork necessary to understand them.

2. Gödel's Completeness Theorem

2.1. **Introduction to Logic.** Before beginning to discuss completeness and incompleteness, it will be necessary to understand some basic concepts about logic. In particular we need to know how a *formal language* is constructed, how it differs from a natural language and how models of the language are constructed. Also we will need to be familiar with the concepts of truth and what it means for a formula to be deducible.

We will start with the idea of a formal language. A *formal language* is comprised of symbols arranged as follows:

A.) Logical Symbols

0.) parentheses: (,).

1.) sentential connective symbols: $\rightarrow$ (if... then...), $\neg$ (not).

2.) variables (one for each positive integer n): $v_1, v_2, \ldots$.

3.) equality symbol: $=$ (possibly).

B.) Parameters

0.) Quantifier symbols: $\forall$

1.) Predicate symbols: For each positive integer n, some (potentially empty) set of symbols called n-place predicate symbols.

2.) Constant symbols: Some (potentially empty) set of symbols.

3.) Function symbols: For each positive integer n, some set (potentially empty) of symbols called n-place function symbols

The concept of a formal language was developed in order to avoid the ambiguities that often arise when one tries to explain a concept in English. The idea of the formal language is that each arrangement of the symbols has a single, unique meaning that is not possible misinterpret. Before working with statements in a logical context we translate them from English into the formal language to avoid misinterpreting the meaning. For example, if we were to say that "

Which particular language we are working with depends on which parameters are employed as well as whether or not equality is included. The logical symbols are constant in every language. For example, the language of set theory contains equality, one two-place predicate parameter, $\in$, and no function symbols. Alternatively the language of number theory contains equality, the two-place predicate parameter $<$, the constant symbol $\mathbf{0}$, the one-place function symbol $\mathbf{S}$ (indicating the succession function) and the two-place function symbols $+$ (addition), $\cdot$ (multiplication) and $\mathbf{E}$ (indicating exponentiation).

In a language, the important statements are the terms of the language and formulas, which are formally called well-formed formulas, or wffs. These are the only relevant sentences logically speaking because these are precisely those sentences which make sense in the language, that is, they conform to the rules of the language. Clearly it is not useful to consider nonsensical statements in any endeavor much less in a logical context.

If a wff ϕ is provable from a language, then we say that it is deducible from the language and write $\Gamma \vdash \phi$. This is the same way we would write that Γ tautologically implies ϕ. This is because the idea of tautological implication means that the truth value of the statement is not contingent on the truth values of the component parts, and if a statement can be proved from the language itself then it can be proved from any form of the language (this will be clearer after we discuss completeness).

We also will need to know what a structure for a language is and what it means for that structure to satisfy that language. First, we will define a structure.

Definition 2.1. A *structure* (denoted by $\mathfrak{A}$) for a language is the set of values for which the language is defined (called the universe of $\mathfrak{A}$ and denoted by $|\mathfrak{A}|$) as well as how the set of parameters and functions of the language are defined.

The universe of $\mathfrak{A}$ determines what is meant by the quantifier symbol $\forall$. That is, when we see a formula such as $\forall x : \phi$, what that really means is $\forall x \in |\mathfrak{A}|$: ϕ. If a structure, $\mathfrak{A}$ *satisfies* a language (that is, if a statement in the language is true, or satisfiable, within the given structure) then we say that $\mathfrak{A}$ is a *model* of the language and write $\vDash_{\mathfrak{A}} \phi$ where ϕ is a wff of the language.

The structure for a language dictates what elements are used to define the components of the language. The function and predicate symbols of the language take on meanings according to the domain of the structure, and the constant symbols of the language correspond to constants within the set of elements of the structure.

To illustrate a familiar example, we will take the basic axioms of an algebraic field. These are:
1) $\exists 0$ such that $\forall a$: $0 + a = a$
2) $\forall a, b, c$: $a + (b + c) = (a + b) + c$
3) $\forall a, \exists(-a)$ such that $a + (-a) = 0$
4)$\forall a, b$: $a + b = b + a$
5) $\exists 1$ such that $\forall a$: $1 \cdot a = a$
6) $\forall a, b, c$: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$
7) For any a not equal to 0, there exists some b with $a \cdot b = 1$
8) $\forall a, b$: $a \cdot b = b \cdot a$
9) $\forall a, b, c$: $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$.

These axioms describe a language containing equality, the function symbols $\cdot$ and $+$ (defined as addition and multiplication), the constant 0 along with the other components mentioned above.

There are several common structures for this language. Those are the real numbers ($\mathbb{R}$), complex numbers ($\mathbb{C}$), natural numbers ($\mathbb{N}$), rational numbers ($\mathbb{Q}$) and integers ($\mathbb{Z}$). It is easy to see which structures of this language are models and which are not. The axioms hold for $\mathbb{R}$ and $\mathbb{C}$ but not for $\mathbb{Q}, \mathbb{N},$ or $\mathbb{Z}$.

In addition to these basic logical ideas, we will need several more specific theorems in order to prove completeness. These are:

Theorem 2.2. ***Rule T****: If $\Gamma \vdash \theta_{,} \ldots \Gamma \vdash \phi_n$, and $\{\phi_1, \ldots \phi_n\}$ tautologically implies β, then $\Gamma \vdash \beta$.*

And:

Theorem 2.3. ***Corollary to Generalization of Constants Theorem****: Assume that $\Gamma \vdash \phi^x_c$, where c is a constant that does not occur in Γ or in ϕ. Then $\Gamma \vdash \forall x : \phi$, and there is a deduction of $\forall x : \phi$ from Γ where c does not occur.*

2.2. **The Theorem.** Now we have enough to discuss the completeness theorem. The completeness theorem essentially asserts that true statements are the result of deductions (there is another theorem, the soundness theorem, that asserts the converse that all deductions lead to true statements). The statement of the theorem is that if ϕ satisfies a language, Γ, then ϕ is deducible from Γ.

Theorem 2.4. *(a) If $\Gamma \models \phi$ then $\Gamma \vdash \phi$*
(b) Any consistent set of formulas is satisfiable

The proof of the theorem is given in six different steps. The general idea is that we take a consistent set of formulas, Γ, and extend it to a set Δ of formulas such that

(i) $\Gamma \subseteq \Delta$

(ii) Δ is consistent and maximal (i.e. for any formula α either $\alpha \in \Delta$ or $\neg\alpha \in \Delta$)

(iii) For any formula ϕ and variable x, there is a constant c such that

$$(\neg\forall x\phi \longrightarrow \neg\phi^x_c) \in \Delta$$

Once we have this language, we create a structure $\mathfrak{A}$ in which members of Γ can be satisfied.

The proof works as follows, we take a consistent set of formulas and create an infinite consistent set of formulas from it. Then we create a structure for the expanded language and show that it satisfies both this and the original language. This shows that if a formula is provable in a language, then it will satisfy the language in any structure, which is precisely what the theorem states.

Proof. **Step 1:** Expand the language by adding a countably infinite set of constants. The important thing to note about this expansion is that Γ remains consistent in the expanded language.

Step 2: For each wff, ϕ, in the new language, and each variable, x, we want to add another wff:

$$\neg\forall x\phi \longrightarrow \neg\phi^x_c$$

where c is one of the new constant symbols. We can do this in such a way that $\Gamma \cup \Theta$ (where Θ is the set of all added wffs) will still be consistent. We do this by adopting a fixed enumeration of the pairs $\langle\phi, x\rangle$ where ϕ is a wff of the expanded language and x is a variable :

$$\langle\phi_1, x_1\rangle, \langle\phi_2, x_2\rangle \dots$$

Let Φ_1 be the wff

$$\neg\forall x_1\phi_1 \longrightarrow \neg{\phi_1}^{x_1}_{c_1}$$

where c_1 is the first of the new constant symbols not occurring in ϕ_1. We define this wff for the pair $\langle\phi_2, x_2\rangle$ etc. so that we generally define Φ_n as

$$\neg\forall x_n\phi_n \longrightarrow \neg{\phi_n}^{x_n}_{c_n}$$

where c_n is the first constant not occurring in ϕ_n.

Let Θ be the set $\{\theta_1, \theta_2, \dots\}$, we claim that $\Gamma \cup \Theta$ is consistent. If not, then, because deductions are finite, for some $m \geq 0$, $\Gamma \cup \{\theta_1, \dots \theta_m, \theta_{m+1}\}$ is inconsistent. Take the least m, then $\Gamma \cup \{\theta_1, \dots \theta_m\} \vdash \neg\theta_{m+1}$.

As we defined Θ earlier, θ_{m+1} is

$$\neg\forall x\phi \longrightarrow \neg\phi^x_c$$

for some x, ϕ and c. So by Rule T we get

$$\Gamma \cup \{\theta_1, \dots \theta_m\} \vdash \neg\forall x\phi$$

$$\Gamma \cup \{\theta_1, \dots \theta_m\} \vdash \phi^x_c.$$

Since c does not appear on the left-hand side of either of these we can apply the corollary to the generalization of constants theorem to the second and get

$$\Gamma \cup \{\theta_1, \dots \theta_m\} \vdash \forall x : \phi.$$

This contradicts the consistency of Γ for $m = 0$.

Step 3: We now use Zorn's Lemma to extend the consistent set $\Gamma \cup \Theta$ to a maximal consistent set Δ (i.e. for any wff ϕ, either $\phi \in \Delta$ or $(\neg\phi) \in \Delta$). Let Λ be the set of logical axioms for the extended language. Since $\Gamma \cup \Theta$ is consistent, there is no formula β such that $\Gamma \cup \Theta \cup \Lambda$ tautologically implies both β and $(\neg\beta)$. Consequently, there is a truth assignment v for the set of all prime formulas that satisfy $\Gamma \cup \Theta \cup \Lambda$. Let $\Delta = \{\phi | \bar{v}(\phi) = T\}$, then clearly for any ϕ either $\phi \in \Delta$ or $(\neg\phi) \in \Delta$, but not both. Also, we have

$\Delta \vdash \phi \Rightarrow \Delta$ tautologically implies ϕ (since $\Lambda \subseteq \Delta$)

$\Rightarrow \bar{v}\phi =$ T (since v satisfies Δ)

$\Rightarrow \phi \in \Delta$.

Consequently Δ is consistent because both ϕ and $\neg\phi$ cannot be in Δ.

Step 4: Now we create from Δ a structure $\mathfrak{A}$ for the new language, with equality replaced by a two place predicate symbol E.

$$(a)|\mathfrak{A}| = \text{the set of all terms of the new language}$$

(b) Define the binary relation $E^{\mathfrak{A}}$by: $\langle u, t\rangle \in E^{\mathfrak{A}}$iff $(u = t)$ belongs to Δ (i.e. $v(u = t) = T$)

(c) For each n-place predicate parameter P, define the n-ary relation $P^{\mathfrak{A}}$ by: $\langle t_1, \dots t_n\rangle \in P^{\mathfrak{A}}$iff $Pt_1, \dots t_n \in \Delta$

(d) For each n-place function symbol$f$, let $f^{\mathfrak{A}}$ be the function defined by: $f^{\mathfrak{A}}(t_1, \dots t_n) = ft_1, \dots t_n$

For the constant c, take $c^{\mathfrak{A}} = c$. Define $s : V \longrightarrow |\mathfrak{A}|$ as the identity $s(x) = x$ on V. It follows that for any term t of the language, $\bar{s}(t) = t$. For any wff ϕ, let ϕ^* be the result of replacing equality with E. Then $\models_{\mathfrak{A}} \phi^*[s]$ iff $\phi \in \Delta$. This can be checked by induction on the number of places at which a connective or a quantifier occur.

Step 5: If the original language did not include equality, then we are done. We simply restrict the structure $\mathfrak{A}$ to the original language to get a structure that satisfies every member of Γ.

Assume that equality was in the language, then $\mathfrak{A}$ will no longer work because we could have the formula $c = d$ (where c and d are distinct constants). In this case we need a structure $\mathfrak{B}$ in which $c^{\mathfrak{B}} = d^{\mathfrak{B}}$. We obtain $\mathfrak{B}$ as the quotient structure $\mathfrak{A}/E$ of $\mathfrak{A}$ modulo $E^{\mathfrak{A}}$.

$E^{\mathfrak{A}}$ is the equivalence relation on $|\mathfrak{A}|$. For each $t \in |\mathfrak{A}|$, let $[t]$ be its equivalence class. $E^{\mathfrak{A}}$ is a congruence relation for $\mathfrak{A}$. That is:

(i) $E^{\mathfrak{A}}$ is an equivalence relation on $|\mathfrak{A}|$,

(ii) $P^{\mathfrak{A}}$ is compatible with $E^{\mathfrak{A}}$ for each predicate symbol P:

$$\langle t_1, \dots t_n\rangle \in P^{\mathfrak{A}}, \text{ and } t_i E^{\mathfrak{A}} t'_i 1 \le i \le n \implies \langle t'_1, \dots t'_n\rangle \in P^{\mathfrak{A}},$$

(iii) $f^{\mathfrak{A}}$ is compatible with $E^{\mathfrak{A}}$ for each f:

$$t_i E^{\mathfrak{A}} t'_i \text{ for } 1 \le i \le n \implies f^{\mathfrak{A}}(t_1, \dots t_n) E^{\mathfrak{A}} f^{\mathfrak{A}}(t'_1, \dots t'_n).$$

Form the quotient structure $\mathfrak{A}/E$ defined as:

(a) $|\mathfrak{A}/E|$ is the set of all equivalence classes of members of $|\mathfrak{A}|$

(b) For each n-place predicate symbol P, $\langle [t_1], \dots [t_n]\rangle \in P^{\mathfrak{A}/E}$ iff $\langle t_1, \dots t_n\rangle \in P^{\mathfrak{A}}$

(c) For each n-place function symbol $f, f^{\mathfrak{A}/E}([t_1], \dots [t_n]) = [f^{\mathfrak{A}}(t_1, \dots t_n)]$.

Let $h : |\mathfrak{A}| \longrightarrow |\mathfrak{A}/E|$ be the map $h(t) = [t]$. Then h is a homomorphism of $\mathfrak{A}$ onto $\mathfrak{A}/E$ and $E^{\mathfrak{A}/E}$ is the equality relation on $|\mathfrak{A}/E|$. For any ϕ:

$$\begin{aligned} \phi \in \Delta &\iff \models_{\mathfrak{A}} \phi^*[s] \\ &\iff \models_{\mathfrak{A}/E} \phi^*[h \circ s] \\ &\iff \models_{\mathfrak{A}/E} \phi[h \circ s]. \end{aligned}$$

So $\mathfrak{A}/E$ satisfies every member of Δ (and hence Γ) with $h \circ s$.

Step 6: Then simply restrict $\mathfrak{A}/E$ to the original language and this satisfies Γ with $h \circ s$. This shows that every wff that satisfies the language satisfies the language for any possible structure of the language.

□

2.3. Implications of Completeness.

This completeness theorem is really a statement about the relationship between languages and structures. The theorem itself states that if a set of formulas is consistent, then it is satisfiable. What this essentially means is that if a formula satisfies every structure for the language then there is a formal proof for the formula.

This theorem implies that the only way a language can be incomplete is if there is a model of the language in which a particular statement is true, and another in which the statement is false. For example, we can see that the language (which was described earlier) comprised of the symbols 0,1,+, -, and ·, then the statement that $\exists a a \cdot a = 2$ is true in the if we take the structure to be $\mathbb{C}$ or $\mathbb{R}$, but not if we choose $\mathbb{Q}$. So it is clear that t he formula $a * a = 2$ is not true in every model of the language and the thus the language is incomplete. What the completeness theorem asserts is that this is the only way that a theory (set of formulas) can be incomplete and that every formula that satisfies every structure is provable in the language.

This makes intuitive sense. It is natural to assume that a statement that is really true, and provable within a language would also be true in any possible set of elements chosen to further define that language.

3. Gödel's First Incompleteness Theorem

When we consider the first incompleteness theorem along with the completeness theorem, we get a very interesting logical result. The theorem is stated as follows

Theorem 3.1. *Any consistent set of formulas cannot be complete, in particular, for every consistent set of formulas there is a statement that is neither provable nor disprovable.*

This means that every set of formulas that does not contain a contradiction is incomplete, that is there is a formula, X such that neither X nor $\neg X$ are provable from the theory.

The proof of this is contingent on the idea of Gödel numbering. Arithmetization of metamathematics statements *about* a language when written mathematically become true statements within the language itself. Metamathematics is the collection of statements not *of* mathematics, but *about* mathematics. What Gödel showed is that true statements *about* mathematics, when arithmetized become true mathematical statements within the language.

The proof consists of translating the statement "This sentence is false" into a mathematical statement within any language (Gödel uses that of the *Principia Mathematica*) and then showing that this is demonstrable (provable) if, and only if, the opposite is also provable. This would make the language inconsistent thus proving the theorem.

3.1. Completeness and Incompleteness.

When we consider incompleteness and completeness together, something interesting becomes clear. For any consistent mathematical language, there is some formula (in fact an infinite number of them) that are true in one structure, but the opposite is true in a different structure.

The implication of this is fairly surprising. In linguistic and analytic philosophy, it is easy to criticize any argument by adopting the view that whatever the proposition is that is being proved, it is only proved within a particular set of philosophical axioms. It is easy for us to question statements that are proved using axioms derived from the English language for one (or more) of several reasons, most likely that we are so familiar with the structure of the language, and that the axioms devised for philosophical purposes generally are not so simple to understand and are not familiar to us. On the other hand, it is not quite so typical that one would question the fundamental statements of mathematics solely on the grounds that they are proved using a particular set of axioms.

This, however, is precisely what Gödel shows with the completeness and incompleteness theorems. The implication of these two theorems is that the truths found through mathematics are, in fact, not infallible. They rest on the *assumption* that the axioms we choose are true. We choose to question these mathematical axioms far less frequently than those of philosophy, and as such it is far more surprising when the idea that math is not infallible is put forward.

It is exactly this that makes the completeness and incompleteness theorems so interesting. They challenge the general perception of mathematics as a field of absolutes. By showing that the very same statements can at once be proved true and false if the axioms are changed, Gödel made it more possible to question the axioms upon which all of accepted mathematics is based.

Acknowledgments. I would like to thank my mentors Asaf Hadari and Rita Jimenez Rolland for their comments and help in writing this paper.

References

[1] Herbert B. Enderton. A Mathematical Introduction to Logic. Harcourt Academic Press. 2002.
[2] George Boolos. Logic, Logic and Logic. Harvard University Press. 1998.
[3] Ernest Nagel and James R. Newman. Gödel's Proof. New York University Press. 2001