MATH 400: Gödel's First Incompleteness Theorem

Lizi Shao

March 21, 2024

Abstract

Gödel's first incompleteness theorem was an ingeniously proved mathematical result that ended the hotly disputed philosophical debate on the completeness problem among mathematicans. In my first presentation, I briefly and comprehensively introduced components of Gödel's proof like consistency, representability, Gödel numbering, the diagonalization lemma, and how they converge to prove the theorem. In this paper, section 1 delves into my interest in Gödel's theorem and offers a heuristic explanation of the proof's components. Section 2 explores philosophical insights from Gödel's theorem, offering both popular and personal perspectives. Section 3 talks about examples of self-reference, an essential part of Gödel's proof, in different forms of art. The paper concludes with Section 4, where I reflect on my presentation and the feedback received.

1 Introduction

1.1 Topic Selection

My first exposure to Gödel's incompleteness theorems predated my college years, stemming from reading a Quora response about astounding mathematical facts that people know. Back then, I didn't have the mathematical expertise to understand even a repeatedly simplified version of the proof of Gödel's theorems. To the best of my recollection, Gödel's results were portrayed as a pinnacle of human intelligence in the 20th century in that response, but that was the superficial extent to which I had understood the importance of Gödel's work. As I became more familiar with Gödel's results, I started to realize that their importance isn't solely confined to mathematics and logic but also extends to philosophy, computer science, and other disciplines. The primary goal for my presentation was to outline parts of Gödel's proof to the first incompleteness theorem comprehensively. In this article, although I will first include a recap of the proof of the first incompleteness theorem, it mainly serves as a heuristic and a refined version complementing the content of my presentation. My primary objective will be to establish immediate or indirect connections to mathematics itself and other disciplines with the theorem itself and the methods Gödel used to prove the theorem.

1.2 A Helpful Refined Heuristic Recap of Gödel's Proof

Gödel's proof of the first incompleteness theorem involves abstract mathematical and meta-mathematical logic. From the feedback given by Professor Li and other students after the presentation, I realized that the presentation was a bit too hasty and unable to fully explain the intuition behind the components of Gödel's proof. Therefore, I devise a more intuitive and non-technical explanation of the components of Gödel's proof in this subsection. I believe this recap may help fellow students who attended the presentation to grasp Gödel's proof more deeply, as it has definitely enhanced my own understanding.

Theorem 1.1 (Gödel's First Incompleteness Theorem) *Any consistent formal system F within which a certain amount of elementary arithmetic can be carried out is incomplete, i.e., there are true statements of the language of F which can neither be proved nor disproved in F using the axioms and language of F. Equivalently,*

$$F \vdash G_F \leftrightarrow \neg Prov_F(\ulcorner G_F \urcorner)$$

where $\vdash$ is the meta-mathematical symbol suggesting what follows is provable in F, G_F is the self-referencing Gödel sentence, and $\neg Prov_F(\ulcorner G_F \urcorner)$ denotes: there doesn't exist a Gödel number $\ulcorner G_F \urcorner$ s.t. $\ulcorner G_F \urcorner$ is the Gödel number of a proof of the sentence G_F in F (This statement G_F can not be proved in F). [5]

Definition 1.2 (Consistency) When a formal system F is consistent, it is without contradictions. I started the presentation with consistency because it is the only assumption that Gödel needed for an interesting enough (capable of performing Peano arithmetic) formal system F to complete the proof. It is extremely important to note that consistency is a meta-mathematical assumption that we have for F, i.e., this assumption can't be realized inside F in any way. [5]

Definition 1.3 (Representability) Heuristically, when something is representable in F, F can express it using the language (symbols and rules) of F from within F, i.e.,

F can *talk about* it in F. Representability is a property that meta-mathematical statements mush have in F in order to prove a theorem related to meta-mathematics in F, such as the first incompleteness theorem (Theorem 1.1). [5]

Definition 1.4 (Gödel Numbering) a numbering system that Gödel invented to create an one-to-one correspondence between a sentence (a sequence of symbols in the language of F) and a natural number using the Fundamental Theorem of Arithmetic. Gödel uses Gödel Numbering to show that meta-mathematical statements are representable (talkable) from within F that's interesting enough. [5]

Lemma 1.5 (The Diagonalization Lemma) *Let $A(x)$ be an arbitrary formula in F about a variable x, then a sentence D can be mechanically constructed s.t.:*

$$F \vdash D \leftrightarrow A(\ulcorner D \urcorner)$$

Essentially, the diagonalization lemma allows for the construction of a formula $A(x)$ that says, "This statement D cannot be proved in F" or equivalently, $\neg \mathrm{Prov}_F(\ulcorner D \urcorner)$. All we need to do now is to find the sentence D that the lemma describes. [5]

Definition 1.6 (The Self-Referential Sentence G_F): Gödel's candidate for the sentence D is the self-referential statement G_F. As he showed, G_F is the very sentence that says, "This statement D cannot be proved in F" or equivalently, $\neg \mathrm{Prov}_F(\ulcorner D \urcorner)$. [5]

Proof of Theorem 1.1 The above is the climax of Gödel's proof. One may proceed to complete the proof in the following steps. First, we show that G_F is the suitable sentence that can be constructed s.t. $F \vdash G_F \leftrightarrow \neg \mathrm{Prov}_F(\ulcorner G_F \urcorner)$. By Definition 1.6, if $\neg \mathrm{Prov}_F(\ulcorner G_F \urcorner)$ is true, G_F is also true. Similarly, if G_F is true, then $\neg \mathrm{Prov}_F(\ulcorner G_F \urcorner)$ is also true. We have demonstrated the constructed sentence G_F leads to $F \vdash G_F \leftrightarrow \neg \mathrm{Prov}_F(\ulcorner G_F \urcorner)$, a case of the lemma. Second, by what we showed in the first step (essentially the diagonalization lemma), the truth values of G_F and $\neg \mathrm{Prov}_F(\ulcorner G_F \urcorner)$ are the same. All is left is to show $\neg \mathrm{Prov}_F(\ulcorner G_F \urcorner)$ indeed has a truth value that is true. Now, since we meta-mathematically assume F is consistent, claiming "This statement G_F can be proved in F" or equivalently, $\mathrm{Prov}_F(\ulcorner G_F \urcorner)$, leads to a contradiction (it's like saying G_F, which asserts itself to be unprovable, is provable). Therefore, $\neg \mathrm{Prov}_F(\ulcorner G_F \urcorner)$ must be true by the meta-mathematical assumption of consistency. It immediately follows that G_F must be true. This means that there are true statements (like the self-referential statement G_F) that F's axioms and language can't prove in F, which is what the first incompleteness theorem claims.

1.3 Mathematical Connections

A crucial insight from the heuristic proof outlined above is the capability to articulate meta-mathematics within F, such as expressing F's consistency assumption within itself. This enables the establishment of $\neg\mathrm{Prov}_F(\ulcorner G_F \urcorner)$'s truth using a meta-mathematical assumption (F is consistent) within F using the language of F. This ability to translate meta-mathematical talk into mathematical terms resolved a longstanding philosophical debate concerning meta-mathematics with rigorous mathematical arguments. Another significant insight is the ingenious use of self-reference, a concept not unique to logic and mathematics but also prevalent in various art forms. The remainder of this paper will explore the how Gödel's theorem and the concepts in his proof connects to philosophy and the arts, from the two perspectives highlighted above.

2 Some Philosophical Insights from Gödel's First Incompleteness Theorem

2.1 Acquiring Mathematical Knowledge: An Epistemological Question

Consider the epistemological query: How do we acquire knowledge in mathematics? If mathematics is considered a science, it stands out because its methods of investigation differ from those in natural sciences. Physicists and chemists, for example, use inductive reasoning, reaching conclusions and postulating theories using empirical evidence, which they then apply to different cases and on a broader scale. While there are instances in science where a theory temporally precedes the empirical evidence to support it, such as Einstein's Theory of General Relativity (proposed in 1915, with empirical support provided by English astronomers Eddington and Crommelin in 1919), inductive reasoning and empiricism remain cornerstones of the natural sciences [10].

In contrast, mathematical knowledge is obtained through deductive reasoning, logically deduced from axioms agreed upon by mathematicians. Proving or logically deducing a true statement or its negation is the primary way to acquire knowledge beyond axioms in mathematics. In other words, mathematical knowledge, or the truth or falsity of a mathematical statement, does not originate from inductive reasoning or empirical evidence, setting it apart from other scientific disciplines in terms of its methods of investigation. Although different schools of mathematical philosophy may disagree on what constitutes a mathematical truth and the nature of the relationship between a mathematical truth and its proof, as we will see in the next section, they would agree that the primary method of investigation in mathematics is fundamentally based on proofs. [4]

In this manner, the axiomatically important question *Can we ever know everything in mathematics?* introduced at the beginning of my presentation is equivalent to the consistency question that Gödel impressively answered.

2.2 Formalism vs. Intuitionism

Before Gödel, two schools of thoughts, the formalists and the intuitionists, competed for the right way to interpret mathematics. Alongside other topics, their discourse on the consistency question aforementioned existed on a purely philosophical ground. The main difference between intuitionists and formalists lies in that the formalists believed mathematical statements, true and false ones alike, to be syntactic strings without semantic meanings. A high schooler would reassuringly use the power rule to differentiate a polynomial in differential calculus not necessarily knowing why the rule works; in the same sense, formalists insisted following the principle above that mathematicians should philosophically ponder the consistency question with the conviction that any true proposition could be proven by advancing forward mechanically using syntactic rules in a formal system without considering their meaning. [1][4][6]

On the other hand, the intuitionists had more intricate views on the said matters, but the main difference they had with formalism was: they believed that mathematics is the product of mental construction and a reflection of human cognition. The intuitionists thought mathematics should not be viewed like a game of chess, with meaning only bestowed on it limitedly within the boundaries of its rules. [4][8]

In summary, the formalists believed in:

> (1) Treating mathematics as a whole like board games; mathematical statements have only syntactic meanings within the boundaries of its rules but not any semantic interpretations. [4]
>
> (2) Viewing the proofs of any mathematical statement also as semantically meaningless deductive steps intended to acquire further steps in a chosen formal system F. [1]
>
> (3) As a consequence of (2), the truthfulness of a statement in F is inextricably tied to its provability within F. In other words, a statement's truth value is not intrinsically meaningful; rather, it is only extrinsically significant insofar as

it could be derived through a sequence of symbol manipulations in accordance with the rules of F. Indeed, the formalists would contend that a statement in F is true if and only if it is provable in F in the sense that its meaning is exclusively contingent upon the existence of a proof. [1]

(4) Following the philosophical underpinnings of (2) and (3), all true statements in F are also self-evidently provable in F; this is also famously known as the consistency of a formal system F. As David Hilbert, the informal leader of the formalists, had suggested, consistency was a somewhat necessary result under and imperative to the formalist belief. [7]

In contrast, the intuitionists contended that:

(1) Mathematics is the product of mental construction. Communications between mathematicians primarily serve to create the same mental picture. [4]

(2) The law of the excluded middle (LEM) should be rejected. LEM is a fundamental law of logic stating that $p \vee \neg p$ for any proposition p. This means that p is binary in that it can only take on a true or a false truth value—there is no middle ground. To the intuitionists, there is be an intricate middle ground for theories and hypotheses yet to be proven or unproven, like the Riemann hypothesis. Time, for this reason, is crucial to them, as previously unprovable problems might become provable as time progresses. In a sense, it is this potential progress allowable by time that grants the plausibility of a "middle ground." [8]

(3) Because of (2) and other reasons, intuitionism is often considered a deviation from classic mathematical philosophy. [8]

Figure 1: *We must know, We shall know*: inscription of Hilbert's unfulfilled formalist dream on his tombstone [12]

2.3 Shattering the Formalist Dream

Gödel's first incompleteness theorem proves that there are true statements in F which are unprovable from within F. Compared to its impacts on intuitionism, the influences the theorem had on formalism were more obvious and direct. The primary and direct consequence of the theorem was that it dealt a groundbreaking blow to one of the central beliefs of formalism that all true statements in a formal system F are provable within F directly and rigorously using mathematics. Indeed, it ends a hotly contended philosophical discussion in mathematics using strict mathematics. [5][8]

But beyond dismantling the ideal formalist mathematical world promised by consistency, the value to the first incompleteness theorem is precisely that it demonstrates the truth value of a mathematical statement may have meaning beyond its property of being (possibly) provable in F, as mathematical statements and proofs of the statements are no longer coextensive. Indeed, there is intrinsic value in the truth values of at least some mathematical statements in F insofar as these truth values *must* exist, *must* take on to be true, and *must* be determined outside F. [2][6]

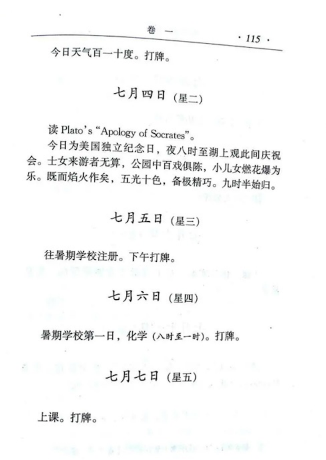
The ability for Gödel to discuss meta-mathematics, i.e, discuss topics like provability, in mathematical language, or, precisely, in the language of the formal system F, is instrumental to Gödel's proof. In a way, both Gödel and the formalists transformed semantically "meaningful" sentences into precise mathematical symbols, but they did so on completely different grounds. The formalists reasoned that mathematical sentences are semantically meaningless, so they treated the sentences as pure symbols as a result. Whereas Gödel treated semantically meaningful sentences in meta-mathematics as pure symbols inside the formal system and, as a result, was able to prove that mathematical sentences should not be viewed as meaningless symbols. Perhaps a little ironically, Gödel had utilized the mathematical instrument or idea central to the formalists to shatter the formalist dream.

3 Self-Reference Outside Mathematics

3.1 Personal Encounters With Self-Reference

Self-reference, the magic, clever, yet convoluted tool that Godel used in his proof of the first incompleteness theorem, is not exclusive to mathematics. In common language, self-reference involves the creation of an object that talks about itself or its own characteristics. It is an intriguing phenomenon precisely because of its mindpuzzling nature. We see its presence in literature, daily language, the arts, and different forms of media. I'd like to start by offering two personal encounters with self-reference which I only came to realize after I started working on the presentation.

Example 3.1 (Diary Entry) The earliest example of self-reference in my memory dates back to when I was in grade school. Throughout the entirety of second grade, we were asked to write a 300-word-long diary every week. One week towards the end of the school year, I had finally run out of things to write about. The demand to constantly filling in a diary is a challenging endeavor, which I claim to be notably echoed by Hu Shih (1891-1962), a famous Chinese literary scholar. Legend has it that when Hu studied in the U.S., he kept a diary in which he logged "played cards" for three consecutive days because it was the only thing he did in those three days. Troubled, I consulted a friend on the topic he planned on writing. "Simple, just write about how you are having trouble writing the diary," he proposed. "It's gibberish, but hey, I call it honest work." Gibberish indeed. Nevertheless, I went along with it. Unbeknownst to me, the very act of writing a literarily meaningless log for the sake of completing schoolwork was actually a great example of an important concept in mathematics and logic.



卷 一 · 115 ·

今日天气百一十度。打牌。

七月四日（星二）

读 Plato's "Apology of Socrates"。

今日为美国独立纪念日，夜八时至湖上观此间庆祝会。士女来游者无算，公园中百戏俱陈，小儿女燃花爆为乐。既而焰火作矣，五光十色，备极精巧。九时半始归。

七月五日（星三）

往暑期学校注册。下午打牌。

七月六日（星四）

暑期学校第一日，化学（八时至一时）。打牌。

七月七日（星五）

上课。打牌。

Figure 2: An excerpt of Hu Shih's *Overseas Diary* in Chinese. The legend turns out to be true: He *did* play cards three days in a row. [3]

Example 3.2 (*All You Zombies*) (spoil alert until the end of the paragraph) Another example came from reading Robert A. Heinlein's *All You Zombies* in freshman year in college, a short science fiction story. *All You Zombies* is about a bartender who is a secretly a time-travel agent working to keep time-travelling safe and free from threats. In the plot, the agent meets a young man who was born female, impregnated by a mysterious stranger, and later became male because of a benign tumor in her reproductive system. The bartender takes the young man back in time to confront the stranger. Through a series of unexpected twists, the bartender realizes that the young man, the stranger, the baby, and himself are all the same person at different points in his life. [9]

In *All You Zombies*, it is an attribute of the protagonist that is self-referenced, namely, his own creation. The protagonist is both the cause and effect of his existence and life experiences. In the setting of the story, a key question that remains unanswered even with the introduction of time-travelling is: Where and When does the loop start? Although there might not a coherent physical theory to explain the never-ending loop, Heinlein's story definitely challenges readers' imagination insofar as it defies the usual understanding of cause-and-effect. An interesting comparison is: Unlike Heinlein's story wherein the possibility for such a figure to exist is logically questionable in a physical/biological sense, the Godel sentence also defies one's usual understanding but is logically permissible in the mathematical world.

3.2 Self-Reference in Arts and Popular Culture

Example 3.3 (*Drawing Hands*) This is one of M.C. Escher's best-known artworks. It is a lithograph containing two hands, each drawing the other. Much like the self-referential birth of the bartender in Heinlein's story, the creation of any of the two hands in this artwork is contingent upon it drawing another hand which draws the former. Aside from self-drawing hands, Escher also explored never-ending self-referential paradoxes in the form of staircases and waterfalls. [13]

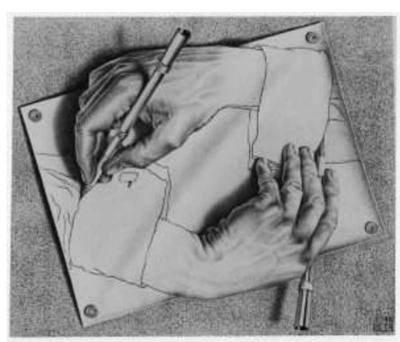


Figure 3: *Drawing Hands* by M.C. Escher [13]

Example 3.4 (Self-Referencing Graffiti) As much as this is an example of self-referential graffiti, it is a self-referential joke. Hence, it is best not to analyze it.

Figure 4: *Sorry About Your Wall*: graffiti by Uncleduke, position unknown [14]

Example 3.5 (Self-Referencing Joke) *I enjoy using the comedy technique of self-deprecation - but I'm not very good at it.*

—Arnold Brown [11]

4 Reflections

4.1 Personal Reflections

Godel's first incompleteness theorem is definitely one of the most abstract concepts in mathematics that I have had explored on in undergrad. It was one that I had to constantly revisit because my presentation focused on explaining parts of its proof, and every time I go back to it, I believe I understand it a bit more and gain something a bit more. This demonstrates just how much complex this theorem is. My most recent understanding came from unravelling a nuanced question which had puzzled me since day one: why does the very process of claiming "There are unprovable true statements in F" not considered a proof in F? After coming back to this many times, I realized it is because the proof relies on the meta-mathematical assumption of consistency which is made and exists outside F. This realization really fills up perhaps the biggest loophole in my current understanding of the theorem. Therefore, I feel the necessity to include in this paper a carefully refined heuristic version of the proof based entirely on this new understanding. Therefore, subsection 1.2, I believe, will clear up some confusions that other students who listened to my presentation may have regarding the proof.

Some lessons learned from this presentation would be to really be more precise about timing because I had gone overtime for this presentation and was left with only a few minutes for discussion. Although parts of the proof can be dull, I feel it is necessary to comprehensively introduce them in the presentation for the sake of rigorousness and because the connections I made in this paper really depends on a basic understanding of the techniques used in the proof (like self-reference).

4.2 Peer Feedback

Reading through feedback from the discussion board I feel like the presentation was well-received. Several students mentioned that they have heard about the theorem before but never had the chance to understand it on a deeper level and that they thought that I did a good job explaining how the proof works on a basic level. I am extremely glad to hear this and hope that I am providing with this paper even a better explanation of the proof in subsection 1.2. I also take note that some students are interested in how theorem connects to other disciplines. For this reason, I have devoted much of this paper to introduce how the theorem and parts of its proof connect to philosophy and arts/popular culture. One student also asked if there are any "meaningful" theorems that are true but unprovable in an interesting enough F. I had expected this question and prepared for this, but I didn't have enough time to actually introduce any example in the presentation because of time constraints. I was able to provide the fellow student with an example in the reply section. In the future, I'd like to maybe do more research on the second incompleteness theorem. I will also focus a bit more on time control, as suggested by Professor Li and others.

References

[1] Alan Weir, https://plato.stanford.edu/, Formalism in the Philosophy of Mathematics, https://plato.stanford.edu/entries/formalism-mathematics/, accessed March 23, 2024

[2] Alfred Driessen, https://philpapers.org/, Philosophical consequences of the Gödel theorem, https://philpapers.org/archive/DRIPCO.pdf, accessed March 24, 2024

[3] Juzhong Ju, https://zhuanlan.zhihu.com/, Does Hu Shih Really Playing Cards, https://zhuanlan.zhihu.com/p/23311772, accessed March 25, 2024

[4] Leon Horsten, https://plato.stanford.edu, Philosophy of Mathematics, , accessed March 23, 2024

[5] Panu Raatikainen, https://plato.stanford.edu/, Gödel's Incompleteness Theorems, https://plato.stanford.edu/entries/goedel-incompleteness/, March 25, 2024

[6] Quinn Crawford, https://elischolar.library.yale.edu/yurj/, Incomplete? Or Indefinite? Intuitionism on Gödel's First Incompleteness Theorem, , accessed March 23, 2024

[7] Richard Zach, https://plato.stanford.edu/, Hilbert's Program, , accessed March 24, 2024

[8] Rosalie Iemhoff, https://plato.stanford.edu/, Intuitionism in the Philosophy of Mathematics, https://plato.stanford.edu/entries/intuitionism/, accessed March 23, 2024

[9] Robert Heinlein, https://emilkirkegaard.dk/, All You Zombies, , Accessed March 24, 2024

[10] Sidney Perkowitz, https://www.britannica.com, Experimental evidence for general relativity, https://www.britannica.com/science/relativity/Philosophical-considerations, accessed March 23, 2024

[11] The Newsroom, https://www.scotsman.com/, Joke of the week: Arnold Brown on putting yourself down, https://www.scotsman.com/arts-and-culture/theatre-and-stage/joke-of-the-week-arnold-brown-on-putting-yourself-down-1480850, Accessed March 24, 2024

[12] Unknown Author, https://commons.wikimedia.org/, File:David Hilbert Grab Wissen.jpg, , accessed March 24, 2024

[13] Unknown Author, https://moa.byu.edu/, M.C. Escher's "Drawing Hands", , accessed March 23, 2024

[14] Uncleduke, https://en.wikipedia.org/, File: Paradox.jpg, , March 24, 2024