Kompleteness, Incompleteness and Undefinability

D. Gunn
Institute for Theoretical Physics, University of Innsbruck, A–6020 Innsbruck, Austria
(Dated: September 5, 2022)

INTRODUCTION

This essay is about four theorems:

1. Gödel's Completeness Theorem
2. Gödel's First Incompleteness Theorem
3. Gödel's Second Incompleteness Theorem
4. Tarski's Theorem on the Undefinability of Truth.

I chose to write about these theorems because I realised I didn't really understand them. For instance, before researching this piece, I might have claimed:

"Gödel's Incompleteness theorem says there are true statements which cannot be proven"

This statement is not true[1]. The primary goal of this essay then is to clear away possible misconceptions regarding these theorems and to understand what they actually say.

The style of this essay reflects this goal. For example, I have chosen my own convention of referring to Completeness in Gödel's Completeness Theorem as "Kompleteness"[2] so as to distinguish it from Completeness in Gödel's Incompleteness theorems; I have included references to online discussions and blog-posts clarifying misconceptions; I have included Part I, which is a summary of First Order Logic; and I have included only a few proofs[3], focusing more on understanding the statements themselves. Finally, this work represents a current state of understanding. As a result, I would love if people would provide comments and corrections at david-kenworthy.gunn@uibk.ac.at.

With that said, this essay is organised as follows:

In Part I, I want to give an introduction, summary and overview of First Order Logic. This is the language and logic in which the four theorems above are formalised. The content of Part I is predominantly based on Ref.[1] and Ref.[2]. Unfortunately, in trying to be concise yet complete, it ended up being a bit "lecture-notes-ey". I included it in this essay because I knew very little about First Order Logic before researching this piece - and so figured others might also not know much about it - and because it is important for understanding the four theorems.

In Part II, I will then discuss Gödel's Kompleteness and First Incompleteness Theorem. In Part III, I will discuss Gödel's Incompleteness Theorem in the context of Tarski's Truth Schema and Theorem on the Undefinability of Truth. In Part IV, I will briefly discuss Gödel's Second Incompleteness Theorem, and then in Part V will present some concluding remarks.

PART I: AN ATTEMPT AT AN INTRODUCTION AND OVERVIEW OF FIRST ORDER LOGIC

I.a. An overview of Formal Systems

The four theorems that are the subject of this essay are ultimately about what can and can't be proven. To understand them, we need to understand ***Formal Systems***. A formal system is an abstract structure in which theorems can be proven. It consists of four components:

1. An ***"Alphabet"*** consisting of a finite set[4] of symbols (aka "letters").
2. A ***Grammar*** consisting of a set of rules about how symbols from the alphabet can be concatenated to form strings, referred to as a "formulae"[5]. Formulae that can be generated by the grammar are called ***"Well Formed Formulae"***.
3. A finite set of ***Inference Rules*** which may be used to infer formulae from finite sets of other formulae.
4. A finite set of ***Axioms*** or ***Axiom schemata***, consisting of well-formed formulae that are taken to be

[1] At least without clarifications
[2] As reflected in the title
[3] In fact sketches of proofs

[4] Given Logic is meant to be the foundation of Set Theory, if you are concerned, as I was, about the use of the word "set" so early on, please hold off your skepticism and just go with it until Section I.e. Meta-languages; though its worth noting that here we consider only finite sets, so naive notions of set theory tend to be consistent (e.g. no Russel's Paradox)
[5] Also referred to as "words"

true and are used as a starting point from which theorems of the formal system are derived.

(1) and (2) are typically referred to together as the ***Syntax*** of a ***Formal Language***[6]. (3) and (4) together are referred to as the ***Proof System***. The four components above correspond to a means by which we can formulate, manipulate and thereby prove theorems.

However, as described above, formal systems are only a means for manipulating symbols. The well formed formulae of formal systems also typically have meaning or ***Semantics***. As we will see, the distinction between proof systems and semantics is important. Very roughly speaking proof systems are about what can be "proven"[7] whereas semantics is about what is "true"[8] .

In the rest of Part I, we will discuss the formal system called ***First Order Logic***[9]. We will begin with the syntax of a first order language, followed by the semantics and then the proof system. We will then finish Part I by introducing some important ***First Order Theories***. These are formal systems which build on first order logic by adding further axioms.

So without further ado, lets introduce First Order Logic.

I.b. The Syntax of First Order Languages

The alphabet of First Order Logic consists of ***Variables***, ***Functions***, ***Predicates***[10], ***Logical Connectives*** and ***Universal quantifiers***. These are the symbols that we use to build sentences of the language. We take $x_1, x_2, ...$ as symbols for variables[11], $f_1, f_2, ...$ as symbols for functions and $P_1, P_2, ...$ as symbols for predicates. For the Logical connectives, we use the symbols

$$\neg,\ \wedge,\ \vee,\ \Rightarrow,\ \Leftrightarrow,\ ),\ (\tag{1}$$

We use the symbols $\exists$ and $\forall$ for the universal quantifiers. Both functions and predicates have a finite (possibly zero) number of inputs. We will discuss what all these symbols mean in the next section, but first we look at how they can be combined to make formulae.

A First Order Language consists of ***Well Formed Formulae***, which are defined as follows. Firstly, we must define what a ***Term*** is. Terms are represented with $t, u, v...$ and a defined recursively as follows:

1. a variable is a term
2. a ***constant*** symbol (a function with 0 inputs, which we denote with $c_1, ..., c_n$) is a term
3. if $t_1, ..., t_n$ are terms and f is an n-input function, then $f(t_1, ..., t_n)$ is a term

Then, well formed formulae, which we represent with $A, B, C...$, are defined recursively as

1. if $t_1, ..., t_n$ are terms and P is an n-input predicate, then $P(t_1, ..., t_n)$ is a (atomic) well formed formula.
2. If A is a well formed formula, then $\neg A$ is a well formed formula.
3. If A and B are well formed formulae, then $(A \wedge B)$, $(A \vee B)$, $(A \Rightarrow B)$, $(A \Leftrightarrow B)$ are well formed formulae
4. If x is a variable and A is a well formed formula, then $(\forall xA)$ and $(\exists xA)$ are well formed formulae

And that's it. Any string that can be composed from these rules, and only strings that can be composed from these rules are well formed formulae. To avoid writing "well formed formulae" each time, we will simply refer to these as formulae from now on.

Example 1. *The following are formulae:*

1. P; $P(x)$; $P(x_1, x_2, ...x_n)$, $P(f_1(x_1), f_2(x_2))$
2. $\neg P(x)$; $(P_1(x) \wedge P_2(x))$; $(\neg P_1(x_1) \vee (P_2(x_2) \Rightarrow P_1(x_2)))$
3. $(\forall xP)$; $(\forall x_1 P(x_1))$; $(\exists x_1 P(x_2))$; $((\exists x_1 P(x_1)) \wedge P(x_2))$

The following are not

1. x; $\forall x\ x$ *(a term is not a formula)*
2. $P(P(x))$ *(Predicates take terms as inputs, not formulae)*
3. $\forall P\ P(x)$ *(you cannot quantify over predicates)*

---

[6] Given an alphabet Σ, let Σ^* be the set of all finite strings of symbols from Σ. Then any subset of Σ^* is a **Formal Language**. Thus the set of well formed formulae in a formal system correspond to a formal language over Σ. Note the concept of a formal language doesn't strictly need a grammar. The grammar of a formal language can itself be formalised into a ***Formal Grammar***. Different formal grammars are categorised ***Chomsky's Hierarchy*** [3]. For the formal languages we look at in this essay, the corresponding grammar is categorised as "Context Sensitive" [4].

[7] Whatever that might mean

[8] Whatever *that* might mean

[9] To those already aware of Propositional Logic, First Order Logic can be seen as an extension of Propositional Logic. As we'll see later, there is also Second Order Logic, which similarly can be seen as an extension of First order Logic. Which Logic is *"the most fundamental"* is subject to debate, as we'll see later

[10] Also sometimes called Relation Symbols

[11] Although a Formal system has finitely many symbols, we assume we have as many variables as we need

So far, all we have are strings of symbols. We should give these strings some meaning. However, before we do, we have the following definition.

Definition 2. *The occurrences of a variable in a formula are defined as follows*

1. *An occurrence of a variable, x, in a formula is **bound** if it occurs within a sub-formula of the form $\forall xA$ or $\exists xA$*
2. *An occurrence of a variable, x, in a formula is **free** if it is not bound.*
3. *A term/formula is **open** if it has free variables. Otherwise it is **closed**.*

Example 3. *We have*

1. *In $\forall x, \exists y P(x, y, z)$, x and y are only bound. z is only free*
2. *In $(\forall x P(x)) \wedge Q(x)$, the first occurrence of x is bound whilst the second is free. This situation can be avoided by renaming bound occurrences of variables, e.g. rewriting the formula as $(\forall y P(y)) \wedge Q(x)$*

Note, open Formulae can be used to define new predicates, e.g:

3. $P(y) \leftrightarrow_{df} \exists x Q(x, y)$

where $\leftrightarrow_{df}$ means the left hand side is a short hand for the right-hand side.[12]

I.c. The Semantics of First Order Languages

Ok, so having established the rules for what strings are allowed in our first order language - i.e. the syntax of our language - we now turn to what they actually mean - i.e. the ***Semantics*** of first order languages. The idea behind predicates is roughly that they they 'express a full sentence'. So, for example 1-input predicates, $P(x)$, refer to "being something", e.g. "x is a number" or "x is tall". A 2-input predicate, $P(x_1, x_2)$ correspond to binary relations, like "x_1 is greater than x_2" or "x_1 kicks x_2". 3-Input predicates correspond to tertiary relations and so on[13]. So, we begin by specifying the semantics of P

Definition 4. *Let $\mathcal{L} = \{x_1, .., .x_{n_1}; c_1, ..., c_{n_2}; f_1, ..., f_{n_3}; P_1, ..., P_{n_3}; \neg, \wedge, \vee, \Rightarrow, \Leftrightarrow; \forall, \exists\}$ be a first order language An **Interpretation** of $\mathcal{L}$ is a pair, (D, I), consisting of a non-empty[14], not necessarily finite set[15] of **elements**, D, referred to as the **Domain**, and a function I which maps the functions and predicates of the language to individuals in the domain or functions as follows:*

1. *if c is a constant, then $I[c]$ is an element of the domain*
2. *if f is an n-input function, then $I[f]$ is a function[16] that maps an n-tuple of elements in the domain to one element in the domain*
3. *if P is an n-input Predicate, then $I[P]$ is a function that maps an n-tuple of elements in the domain to True or False.*

In the wise words of Bill [5], "Things done without example, in their issue are to be fear'd". So lets create an example which we'll use throughout this semantics section.

Example 5. *Let $L = \{x_1, x_2, x_3, c, f, P, \neg, \wedge, \vee, \Rightarrow, \Leftrightarrow, \forall, \exists\}$. The following is an interpretation:*

- $D = \{"London", "Paris", "New\ York\}$,
- $I[c] = "London"$
- $I[f] : D \to D$ *defined by*

  $$"London" \mapsto "Paris"$$
  $$"Paris" \mapsto "London"$$
  $$"New\ York" \mapsto "London"$$

  ("the nearest city to x")
- $I[P] : D \to \{True, False\}$ *defined by*

  $$"London" \mapsto True$$
  $$"Paris" \mapsto True$$
  $$"New\ York" \mapsto False$$

  ("x is a capital city)

Note that the interpretation does not yet assign meaning to the variables. We do this next:

Definition 6. *A **Variable Assignment**[17], V, of $\mathcal{L}$ in an interpretation $\mathcal{I} = (D, I)$, is a function from the variables of $\mathcal{L}$ to the domain, D. We write $V_{x \mapsto a}$ for a variable assignment which maps x to a and maps all other variables as V does.*

---

[12] See discussion of Meta-language at end of Part I

[13] 0-input predicates are sentences without any internal structure. As we'll see later they can be evaluated as either True or False. First Order Logic with only 0-input predicates corresponds to Propositional Logic

[14] In First order logic, the domain must be non-empty

[15] Again you may object to the use of set so early on. Please again wait for the discussion of Meta-languages in Section I.e.

[16] Likewise if you have an objection to the use of the word function, again please wait until Section I.e.

[17] Also referred to as Valuation

Example 6. ***cont.*** *Continuing with the example let* $V : \{x_1, x_2, x_3\} \to D$*, be the variable assignment*

$$\begin{aligned} x_1 &\mapsto \text{"}Paris\text{"} \\ x_2 &\mapsto \text{"}New\ York\text{"} \\ x_3 &\mapsto \text{"}New\ York\text{"} \end{aligned} \tag{2}$$

Correspondingly, $V_{x_2 \mapsto \text{"}London\text{"}} : \{x_1, x_2, x_3\} \to D$ *would be the variable assignment*

$$\begin{aligned} x_1 &\mapsto \text{"}Paris\text{"} \\ x_2 &\mapsto \text{"}London\text{"} \\ x_3 &\mapsto \text{"}New\ York\text{"} \end{aligned} \tag{3}$$

We are now in position to assign meaning in First order Logic. We begin by assigning meaning to the terms:

Definition 8. *Let* $\mathcal{I} = (D, I)$ *be an interpretation of* $\mathcal{L}$ *and* V *a variable assignment. Then the meaning of a term,* t*, in* I *given* V*, represented by* $\mathcal{I}_V[t]$*, is defined recursively as follows:*

1. *if* t *is a variable, then* $I_V[t] = V[t]$
2. *If* t *is a constant, then* $I_V[t] = I[t]$
3. *If* $t = f(t_1, ..., t_n)$,
   then $I_V[t] = I[f][I_V(t_1), ...I_V(t_n)]$

Example 6. ***cont.*** *with* V *as above, we have* $I_V[f(x_1)] = I[f][I_V(x_1)] = I[f][\text{"}Paris\text{"}] = \text{"}London\text{"}$

Now we assign meaning to the formulae:

Definition 9. *Let* $\mathcal{L}$ *be a language. Given an interpretation* $I = (D, I)$ *and a variable assignment* V*, the meaning of a formula* A *in* I *given* V *is defined recursively as follows:*

1. $P(t_1, ...t_n)$ *is true in* I *given* V *if* $I[P][\mathcal{I}_V[t_1], ..., \mathcal{I}_V[t_n]] =$ *True. It is false in* I *given* V *if* $I[P][\mathcal{I}_V[t_1], ..., \mathcal{I}_V[t_n]] =$ *False.*
2. *Given two formulae,* A *and* B*, the meaning of logical combinations of* A *and* B *in* I *given* V *can be calculated via the following Truth table*[18]*.*

A	B	$\neg A$	$(A \wedge B)$	$(A \vee B)$	$(A \Rightarrow B)$	$(A \Leftrightarrow B)$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

3. *Given a formula* A*,* $\forall xA$ *is true in* I *given* V *if* $I_{V_{x \mapsto a}}[A] =$ *True for all* a *in the domain*[19]*.*
4. *Given a formula* A*,* $\exists xA$ *is true in* I *given* V *if there is an element* a *in the domain such that* $I_{V_{x \mapsto a}}[A] =$ *True.*[20]

We write "A *is true in* I *given* V*" as* $\models_{I,V} A$

Example 6. ***cont.*** *We have:*

- $\models_{I,V} P(x_1)$, $\not\models_{I,V} P(x_3)$, $\models_{I,V} P(x_1) \Rightarrow \neg P(x_3)$
- $\models_{I,V} \forall x_1 P(c)$, $\not\models_{I,V} \forall x_1 P(x_1)$
- $\models_{I,V} \exists x_3 P(f(x_1)) \wedge P(x_3)$, $\not\models_{I,V} \exists x_1 P(f(x_1)) \wedge P(x_3)$

Note that in First Order Logic, the meaning of a formula in an interpretation given a variable assignment is "True" or "False" and only "True" or "False".

Finally we have:

Definition 10. *Let* A *have no free variables. Then*

1. A *is True in* I *if* $\models_{\mathcal{I},V} A$ *for all variable assignment* V *in* I*. We write "*A *is True in* I*" as* $\models_{\mathcal{I}} A$*.*
2. *If* $\models_{\mathcal{I}} A$*, we say that* I ***satisfies*** A*. Moreover, if there is an interpretation that satisfies* A*, we say* A *is* ***satisfiable***[21]*.*
3. *A* ***model*** *of a set of formulae,* S*, is an interpretation that satisfies every formula in* S*.*
4. A *is true given a set of formulae* S *are true, if every model of* S *satisfies* A*. That is, if for all* I *such that* $\models_I S$*,* $\models_I A$*. We write "*A *is true given a set of formulae* S *are true" as* $S \models A$*.*
5. A *is a* ***tautology***[22] *if* $\models_{\mathcal{I}} A$ *for all interpretations,* I*. We write "*A *is a Tautology" by* $\models A$[23]*.*

Example 6. ***cont.*** *Finishing our example.*

1. $\models_I P(c)$, $\not\models_I P(x_1)$
2. $P(x_1)$ *is satisfiable.* $P(x_1) \wedge \neg P(x_1)$ *is not satisfiable*

---

[18] From the truth Table it is clear that we could have therefore written $A \vee B \leftrightarrow_{df} \neg(\neg A \wedge \neg B)$, $A \Rightarrow B \leftrightarrow_{df} \neg(A \wedge \neg B)$, $A \Leftrightarrow B \leftrightarrow_{df} ((A \Rightarrow B) \wedge (B \Rightarrow A))$

[19] Now you might notice this definition is almost circular; that $\forall$ has also slipped into our discussion. Again, please hold off skepticism till the Meta-languages section and appreciate that the defintion is none the less unambiguous.

[20] It is clear that therefore we may write $\exists xA \leftrightarrow_{df} \neg(\forall x \neg A)$

[21] This is exactly the same meaning as in 3-SAT

[22] Alternatively a set of formulae are "Valid"

[23] We might note that $A \models B$ if and only if $\models A \Rightarrow B$. For this reason, Introductions to Logic may concern themselves only with the Tautologies of a Formal System

3. *The intepretation of our example is a model for* $\{P(c)\}$
4. $P(x) \models P(f(x))$
5. $\neg(P(x_1) \wedge \neg P(x_1))$ *is a Tautology.*[24]

I.e. A Proof System for First Order Logic

Oof, that was a lot. But we now have a first order language equipped with semantics. To elevate it to First Order Logic we need one more component. We have defined what is true; however, we need a set of rules which allows us to *deduce* what is true - a set of ***Rules of Inference***.

Definition 11. *Given a set of Rules of Inference,* $\mathcal{R}$*, we write "Given the set of formulae S, we may deduce via* $\mathcal{R}$ *A" as* $S \vdash A$.

So lets start building a set of rules of inference for First Order Logic. Throughout, we will use S, S_1, S_2 to refer to sets of Formulae and $A, B, C, ...$ to refer to individual formulae. We begin with the painfully obvious: given A, we should be able to deduce A. We write this in generality as

(a) If A is an element of S, then $S \vdash A$

where (a) is the name we will use to refer to the rule (standing for Assumption).

The next most painfully obvious rule comes from combining formulae. Namely if we have A and B then we should be able to deduce $A \wedge B$ ("A and B"). Conversely, if we have $A \wedge B$, then we should be able to deduce A and to deduce B. Written in generality this yields:

$(\wedge i)$ If $S_1 \vdash A$ and $S_2 \vdash B$, then $S_1 \cup S_2 \vdash A \wedge B$.

$(\wedge e)$ If $S \vdash A \wedge B$ then $S \vdash A$ and $S \vdash B$.

Here, the "i" in $(\wedge i)$ stands for introduction and likewise the "e" $(\wedge e)$ stands for elimination.

Coming fourth place for painfully obvious, we have that from $\neg\neg A$ ("not not A"), we should should be able to deduce A[25]. Writing this in generality we have

$(\neg\neg e)$ If $S \vdash \neg\neg A$ then $S \vdash A$

Like $\wedge$, $\neg$ is also equipped with an introduction rule

$(\neg i)$ If $S_1 \cup \{A\} \vdash B$ and $S_2 \cup \{A\} \vdash \neg B$ then $S_1 \cup S_2 \vdash \neg A$

This is a good point to stop and see how these rules of inference can be combined.

Example 12. $\{A\} \vdash \neg\neg A$ *(the reverse of* $(\neg\neg e)$*)*

Proof. We have:

- From (a): (1) $\{A, \neg A\} \vdash A$
- From $(\neg i)$ with $\{(1)\}$: (2) $\{A\} \vdash \neg\neg A$

□

Example 13. $\{A, \neg A\} \vdash B$ *for all B (this is known as the "Principle of Explosion" - from a contradiction, everything follows)*

Proof. Let B be any Formula.

- From (a): (1) $\{A, \neg A, B\} \vdash A$
- From (a): (2) $\{A, \neg A, B\} \vdash \neg A$
- From $(\neg i)$ with $\{(1), (2)\}$: (3) $\{A, \neg A\} \vdash \neg\neg B$
- From$(\neg\neg e)$ with $\{(3)\}$: (4) $\{A, \neg A\} \vdash B$

□

These examples demonstrate how these rules of inference can be used to *syntactically deduce* formulae from other formulae. This is an important distinction. The proofs in Example 12 and 13 make no appeal to meaning at all. They simply manipulate the strings by the allowed rules. We will have more to say about the connection between $\vdash$ and $\models$ in Part II.

For completeness, the following set of rules of inference, in addition to (a), $(\wedge i)$, $(\wedge e)$, $(\neg i)$, $(\neg\neg e)$, constitute a proof system of First Order Logic [2]. Note, in the following $A_{x\mapsto t}$ corresponds to A where any free occurrence of x is substituted with t:

$(\vee i)$ If $S \vdash A$ then $S \vdash A \vee B$ and $S \vdash B \vee A$ for any B

$(\vee e)$ If $S_1 \vdash (A \vdash B)$, $S_2 \cup \{A\} \vdash C$ and $S_3 \cup B \vdash C$, then $S_1 \cup S_2 \cup S_3 \vdash C$

$(\Rightarrow i)$ If $S \cup \{A\} \vdash B$ then $S \vdash (A \Rightarrow B)$

$(\Rightarrow e)$ If $S_1 \vdash (A \Rightarrow B)$ and $S_2 \vdash A$, then $S_1 \cup S_2 \vdash B$

$(\forall i)$ For any closed term, t, that does not occur in Γ or A, if $\Gamma \vdash A_{x\mapsto t}$, then $\Gamma \vdash \forall x A$

$(\forall e)$ If $\Gamma \vdash \forall x A$ then $\Gamma \vdash A_{x\mapsto t}$ for any closed term t,

$(\exists i)$ For any closed term, t, $\Gamma \vdash A_{x\mapsto t}$ then $\Gamma \vdash \exists x A$

[24] It is the Law of non-contradiction

[25] Well I say painfully obvious. In fact there is a school of Logic called *Intuitionisitic Logic*[6] which does not allow this rule of inference. As an implication, it rejects mathematical proofs of the form "Assume the Theorem isn't True...."

($\exists e$) For any closed term, t, that does not occur in A,B or Γ_2, if $\Gamma_1 \vdash \exists x A$ and $\Gamma_2 \cup \{A_{x\mapsto t}\} \vdash B$, then $\Gamma_1 \cup \Gamma_2 \vdash B$

I say *a* proof system because it is not unique. For instance, we have chosen to codify our proof system in only rules of inference. We could have also used axioms instead[26].

We will now conclude Part I by discussing briefly first order theories and the Meta-language of First Order Logic

I.d. First Order Theories

In this essay, we will uses the term ***First Order Theory*** (symbolized by the pair $(\mathcal{L}, T)$) to apply to a formal system consisting of a first order language, $\mathcal{L}$, equipped with the rules of inference from First Order Logic and a set of additional ***Axioms***, T. By this definition, First Order Logic is also a first order theory (with an empty set of additional axioms). Another such example of first order theory, which we will have a lot more to say about, is ***Peano Arithmetic***

Example 14. ***Peano Arithmetic (PA)*** *The first order theory of Peano Arithmetic is defined as follows*

- $\mathcal{L}_{PA} = \{x, y, ...z; 0; S, +, \times; =; , \neg, \wedge, \vee, \Rightarrow, \Leftrightarrow ; \forall, \exists\}$*, with syntax of a first order language, and where* 0 *is a constant,* S*,* $+$ *and* $\times$ *are 2-input functions and* $=$ *is a 2-input predicate*
- *The rules of inference are those of first order logic*
- *In addition we have following* ***Axioms****,* T_{PA}*:*
  1. *for all variables* x*,* $x = x$
  2. *for all variables* x*,*y $(x = y \Rightarrow f(t_1, ..., x, ...t_n) = f(t_1, ..., y, ...t_n)))$ *for any n-input function*
  3. *for all variables* x*,*y $(x = y \Rightarrow (A \Rightarrow A_{x\mapsto y})$
  4. $\forall x(0 \neq S(x))$
  5. $\forall x, y(S(x) = S(y) \Rightarrow x = y)$
  6. $\forall x(x + 0 = x)$
  7. $\forall x, y(x + S(y) = S(x + y))$
  8. $\forall x(x \times 0 = 0)$
  9. $\forall x, y(x \times S(y) = x \times y + x)$
  10. $\forall \vec{y}\ \big((A(0, \vec{y}) \wedge \forall x(A(x, \vec{y}) \Rightarrow A(S(x), \vec{y}))) \Rightarrow \forall x A(x, \vec{y})\big)$ *for any formula* A *with occurring variables* $x, y_1, ..., y_k$*.* $\Rightarrow A(S(x), y_1, ...y_k))) \Rightarrow \forall x A(x, , y_1, ..., y_k)$ *for any formula* A *with occurring variables* $x, y_1, ..., y_k$*.*

(1-3) are equality axioms, (4-5) successor function axioms, (6-9) addition and multiplication axioms and (10) is an ***Axiomatic Schema****, corresponding to a recursively enumerable*[27] *set of axioms.*

Thus we see Peano Arithmetic is a formal system which extends First Order Logic with additional axioms. In was constructed such that it has arithmetic on the Natural Numbers (i.e. $\mathbb{N} = \{1, 2, 3, ...\}$ with $1 + 1 = 2$, $1 \times 2 = 2$ etc) as a ***model*** (i.e. $\models_{(\mathbb{N}, +\times)} PA$).

In this essay, we will use the word ***Theorem*** to refer to a formula in a first order theory that can be derived from the Axioms. That is if T are the set of additional axioms, A is a theorem if $T \vdash A$.

Example 15. *The Theorems of First Order Logic are Tautologies.*

Proof. See Gödel Kompleteness Theorem in Part II. □

Example 16. *Let* $1 \leftrightarrow_{df} S(0)$ *and* $2 \leftrightarrow_{df} S(1)$[28]*. Then* $1 + 1 = 2$ *is a Theorem of Peano Arithmetic:*

Proof. Proof: $1 + 1 = S(0) + S(0) = S(S(0) + 0) = S(S(0)) = S(1) = 2$ □

I.e. The Meta-language of First Order Logic

Finally, we should note that the symbols $\vdash$, $\models$ and $\leftrightarrow_{df}$ do not belong to the formal language of First Order Logic. They are not symbols in the alphabet. Rather, they belong to the ***Meta-language*** [1] - that is they belong to the language used to describe the formal system (in these notes, predominantly standard English, augmented by symbols such as $\vdash$, $\models$ and $\leftrightarrow_{df}$).

Generally, a Meta-language is used to describe an ***Object Language***. Indeed, in order to define the semantics of the Object Language, we must use a Meta-language [9, 10]. The Meta-language can be chosen independently of the Object Language (we could for instance have written these notes in French). However, the Meta-language should have the following properties [11]

1. The Meta-language should contain a copy of the Object Language, L (it must be able to 'say' anything L can say)
2. The Meta-language should be able to talk about the Formulae of L (it must be able to refer to them with a 'name'), as well as the syntax of strings of symbols from the Alphabet.

[26] See Discussion in Ref [7]

[27] It can be shown PA is not finitely axiomatizable [8]

[28] Alternatively, add $1, 2$ to the constants of the language and add the axioms $S(0) = 1$, $S(1) = 2$

Property (1) is the reason why words like 'set' appear in this text. Typically, the Meta-language used to describe a first order theories contains some set theory [9][29]. Its also why the semantic definitions of $\forall$ and $\exists$ seem almost circular. We will have more to say about this when we discuss Meta-language in more detail in Part III.

PART II: GÖDEL'S KOMPLETENESS AND FIRST INCOMPLETENESS THEOREM

Aaaaand Breath. Ok, having done all that work, we are now in a position to efficiently state and understand Gödel's Completeness Theorem and Gödel's First Incompleteness Theorem. This is precisely the goal of Part II.

Before that, however, I would like to say that I find the naming of these theorems very unfortunate. The Completeness in Gödel's Completeness theorem is not the same as Completeness in Gödel's Incompleteness Theorems. Therefore, I introduce here my own convention of referring to the former as Komplete and the latter as Complete.

We now begin with two definitions relating the concepts of $\vdash$ and $\models$.

Definition 17. ***Soundness and Kompleteness***[30]

1. *A Formal System is* ***Sound*** *if $S \vdash A$ implies $S \models A$ for any set of Formulae S and Formula A*
2. *A Formal System is* ***Komplete***[31] *if $S \models A$ implies $S \vdash A$ for any set of Formulae S and Formula A*

Soundness and Kompleteness are a property of $\vdash$, i.e. our rules of inference. Obviously, they are very desirable properties to have. They ensure anything we can derive is true in all models, and anything always true is derivable. Note that constructing a sound proof system is easy. For example, the proof system consisting of only the rule of inference (a) is obviously sound. However, such a proof system is also trivial. The challenge is to construct a sound proof system complicated enough that it is Komplete. To this end we have the following remarkable theorems:

Theorem 18. *First Order Logic is Sound.*

Theorem 19. ***Gödel's Kompleteness Theorem***[32] *[13]: First Order Logic is Komplete*

Thus, in first order theories, we have $S \vdash A$ if and only if $S \models A$. Colloquially, we can express this as "In first order theories, something is provable if and only if it is *always* true"[33].

Next, we move onto Gödel's First Incompleteness Theorem. First we need another two definitions

Definition 20. ***Consistency and Completeness***

1. *A first order theory, $(\mathcal{L}, T)$ is* ***Consistent*** *if there is no formula A in $\mathcal{L}$, such that $T \vdash A$ and $T \vdash \neg A$. That is the Axioms do not lead to a* ***contradiction***
2. *A first order theory, $(\mathcal{L}, T)$, is* ***Complete*** *if, for all formula A in $\mathcal{L}$, either $T \vdash A$ or $T \vdash \neg A$*

As we saw earlier in Example 13, if a theory is not consistent, then any proposition is derivable. Thus consistency is certainly something to be desired[34]. However, First Order Theories are generally not complete.[35].

We can now state Gödel's First Incompleteness Theorem[36].

Theorem 21. ***Gödel's First Incompleteness Theorem*** *[14, 15]: Let $(\mathcal{L}, T)$ a first order theory containing Peano Arithmetic*[37] *be consistent. Then $(\mathcal{L}, T)$ is incomplete.*

Colloquially, "In any complicated enough, consistent theory, there will be a statement which can neither be proved nor disproved". This statement "which can neither be proved nor disproved" is often referred to as the theories' Gödel Statement. Note, from now on, we will say "**1OT extending PA**"[38] as shorthand for "first order theory containing axioms of Peano Arithmetic".

Gödel's Kompleteness Theorem and First Incompleteness Theorem can seem contradictory[39]. This appearance of a incongruity can be cleared up considering the semantics more carefully. Recall a model of a a set of

---

[29] Or its extensions. See also Ref.[12] for further discussion
[30] Using my convention to avoid confusion with Completeness later on - typically called "Completeness"
[31] Again, using my convention - typically this is called "Complete"
[32] Again, using my convention. Typically this is called "Gödel's Completeness Theorem"
[33] Compare this to the statement about Gödel incompleteness theorem in the introduction.
[34] See also concluding remarks
[35] In fact, First Order Logic is only complete if the allowed predicates are strongly restricted
[36] As presented [14]
[37] Note, in fact we do not need all 10 axioms. The 10th axiom can be replaced by $\forall y(y = 0 \vee \exists x(S(x) = y)$. This axiomatization of arithmetic is referred to as ***Robinson Q***
[38] Similarly PA could be changed to Q throughout these notes
[39] Not least because of their usual names

formulae, S, is an interpretation, I, that satisfies every formula in S , i.e. $\models_I A$ for all A in S. Then we have the following semantic conditions for consistency:

Theorem 22. ***Semantic Definition of Consistency*** *A theory is consistent if and only if it has a model* [40]

This result is not at all obvious[41] and I find it to be quite a beautiful result. Colloquially, it states that at first order theory cannot reach a contradiction if and only if we can find a concrete example of it.

We may now re-state Gödel's Kompleteness and First Incompleteness Theorem in terms of consistency and semantics

Theorem 23. ***Gödel's First Incompleteness Theorem [Semantic version]*** *Let $(\mathcal{L}, T)$ be a 1OT extending PA. Then there is a formula in $\mathcal{L}$, A, such that*

1. *there exists a model of T, M_1 (i.e $\models_{M_1} T$) such that $\models_{M_1} A$*
2. *there exists a model of T, M_2 (i.e $\models_{M_2} T$) such that $\models_{M_2} \neg A$*

Proof. By Gödel's Kompleteness Theorem, it follows from Gödel's Incompleteness Theorem that neither $T \models A$ nor $T \models \neg A$. This can happen two ways: either T has no model, or there is a model of T that satisfies A and model which doesn't. But, by the semantic definition of consistency, $(\mathcal{L}, T)$ has at least one model. Hence the result. □

In fact Gödel's First Incompleteness as originally published in 1931 showed that the model M_1 can be taken to be the arithmetic on the Natural Numbers, $(\mathbb{N}, +, \times)$. That is to say Gödel's First Incompleteness theorem tells us that there are *truths about arithmetic on the Natural Numbers* which cannot be proven from Peano Arithmetic or its extensions[42], i.e. there is a G such that $PA \models_{(\mathbb{N},+,\times)} G$ but $PA \nvdash G$

As a corollary, we have

Corollary 24. *There are non-standard models of Peano Arithmetic (i.e. models other than $(\mathbb{N}, +, \times)$)*

This gives us a new way of looking at Gödel's First Incompleteness Theorem. What it tells us is that in any first order theory that contains Peano Arithmetic, the axioms will never be strong enough to pin down a unique model of the theory.[43]

As an example, consider adding the Gödel statement to the axioms of Peano Arithmetic. This defines a new first order theory. This new theory is still consistent - it still has at least one model; indeed the Gödel statement is true in $(\mathbb{N}, +, \times)$. However, this new theory also still satisfies the conditions for Gödel's First Incompleteness Theorem. Thus this new theory has its own Gödel statement and so on and so on[44].

Hopefully, it is now clear what Gödel's Kompleteness and First Incompleteness Theorem say. In the next two parts, we will first investigate the proof of Gödel's First Incompleteness Theorem and its connection to "The Liar", and then answer some remaining questions about it.

PART III: GÖDEL'S FIRST INCOMPLETENESS THEOREM IN THE CONTEXT OF TARSKI'S TRUTH SCHEMA

In this Part, we give a sketch of the proof of Gödel's First Incompleteness Theorem from Tarski's Theorem on the Undefinability of Truth. To begin, we have to return to a more detailed discussion of Meta-languages.

III.a. Meta-languages and Tarski's Truth Schema

Recall a Meta-language, M, is the language used to discuss the Object Language, L and should have the following properties:

1. It should contain a copy of the Object Language, L - it should be able to "say" everything L can say
2. It should be able to talk about the formulae of L - it should be able give a "name" to any formula in L - as well as discuss syntax

Here we are making the distinction between the name of a sentence and the sentence itself. We make this distinction as we want to talk about the truth of a sentence. Talking about truth in general is difficult [18]. However, Tarski proposed the following criterion for the truthfulness of sentences themselves:

A sentence is true if and only if its content is true.

---

[40] For this reason, some Logicians use "satisfiable" as a synonym for consistent. That T is consistent implies it has a model is called the "Model Existence Theorem"

[41] In fact Thm 19 is a corollary of the forward direction of this claim [2, 15]

[42] This is the accurate modified form of the statement from the introduction

[43] See Ref. [16] for an excellent discussion.

[44] This is discussed nicely in Ref. [17]

So for example

$$\text{"1+1=2" is true if and only 1+1=2} \tag{4}$$

This criterion seems painfully trivial, but lets break it down to see its content. First we remember the Meta-language must be able to express everything L can say. Hence the proposition $1+1=2$ appears both in the Object Language and the Meta-language. Secondly, the Meta-language must be able to name each proposition in the Object Language. Here we happen to name it "1+1=2", but we could also call it "Sentence 1" or "My First Equation".

We might notice therefore that (4) is a sentence written completely in the Meta-language. The sentence label "$1+1=2$" and $1+1=2$ are both strings of the meta-alphabet. Taking this further, the phrase "if and only if" is a logical connection of the Meta-language. Indeed, it expresses the same concept as $\Leftrightarrow$ in the Object Language. To make this distinction clear, in this essay we will use $\Rightarrow$ and $\Leftrightarrow$ for the Object Language and $\rightarrow$ and $\leftrightarrow$ for the *same concept* in the Meta-language. Finally, we realise "is true" in a predicate in the Meta-language.[45]

This leads us to Tarski's Criterion of Truth. We denote the name of a sentence, A, is with $\ulcorner A \urcorner$. Then Tarski's criterion is defined as follows.

Definition 25. ***Convention T*** *[18] A one variable predicate $\mathbb{T}$ for the truth of a sentence must satisfy*

$$\mathbb{T}(\ulcorner A \urcorner) \leftrightarrow A \tag{7}$$

for all sentences A.

As a final comment, we reiterate that we can have the case that the Object Language and the Meta-language are the same. That is, the language is capable of self-reference. English, for instance, is a self-referential language. We can say things like:

$$\text{"This is sentence is seven words long"} \tag{8}$$

It may seem counter-intuitive for a language to contain a copy of itself. However, this tension is somewhat resolved by realising the number of formulae in any such formal language is countably infinite [20]. In such self-referential cases, a predicate satisfying Convention-T would satisfy

$$\mathbb{T}(\ulcorner A \urcorner) \Leftrightarrow A \tag{9}$$

As we will see, this leads to paradoxes. But first, we will briefly discuss Gödel's encoding

III.b Gödel's Encoding

A key insight of Gödel is that Peano Arithmetic can talk about itself. Namely, we are able to assign a unique number - i.e. a unique name - to every formula in Peano Arithmetic. In fact, not only can we assign each formula in Peano Arithmetic a unique number, we can assign a unique number to finite sequences of formulae and therefore to each proof in Peano Arithmetic. I won't go into detail here as to how this is done but good explanations are found here: Ref.[21, 22]. From now on, we will use $\ulcorner \urcorner$ to symbolize this encoding, i.e. in a 1OT extending PA, $\ulcorner A \urcorner$ is the Gödel number of A.

III.c. Tarski's Truth Schema, Diagonal Lemma, and the Undefinability of Truth

We begin investigating these self-referential theories with an extremely important Lemma:

Lemma 26. ***The Diagonal Lemma***[46]. *Let $(\mathcal{L}, T)$ be 1OT extending PA. For every formula $A(x)$ with one free variable, there is a formula B such that:*

$$T \vdash (B \Leftrightarrow A(\ulcorner B \urcorner)) \tag{10}$$

The Diagonal Lemma leads immediately to Tarski's indefinabilty of Truth.

Theorem 27. ***Tarski's Theorem on the undefinable of Truth*** *Any 1OT extending PA containing a predicate satisfying Convention-T is inconsistent.*

Proof. Assume T is consistent and contains a predicate, $\mathbb{T}$, that satisfies convention-T, i.e.

$$\mathbb{T}(\ulcorner A \urcorner) \Leftrightarrow A \tag{11}$$

Consider the predicate $\neg\mathbb{T}$. By the Diagonal Lemma, $\exists\Lambda$ such that

$$T \vdash (\Lambda \Leftrightarrow \neg\mathbb{T}(\ulcorner \Lambda \urcorner)) \tag{12}$$

However, $\mathbb{T}$ satisfies Convention-T, so we also have

$$T \vdash (\Lambda \Leftrightarrow \mathbb{T}(\ulcorner \Lambda \urcorner)) \tag{13}$$

---

[45] As a further comment $\vdash$ and $\models$ are in fact Meta-language symbols [1]. Hence the string

$$\vdash \neg(A \wedge \neg A) \tag{5}$$

is technically in-fact a Meta-language statement. It is staying

$$\neg(A \wedge \neg A) \tag{6}$$

is derivable from First order Logic. This reminds me of the distinction in Kantian Philosophy between the "Noumenon" ("the [unknowable] thing itself") and the "Phenomenon"[19]. It is curious that this distinction might apply to languages, given the Object Language isn't physical.

[46] Also called "Fixed point lemma" or "self referential lemma"

Combining these yields:

$$T \vdash (\mathbb{T}(\ulcorner \Lambda \urcorner) \Leftrightarrow \neg \mathbb{T}(\ulcorner \Lambda \urcorner)^{\urcorner}) \tag{14}$$

But of course we also have[47]

$$T \vdash \neg(\mathbb{T}(\ulcorner \Lambda \urcorner) \Leftrightarrow \neg \mathbb{T}(\ulcorner \Lambda \urcorner)^{\urcorner}) \tag{15}$$

thus we derive a contradiction. Thus T is not consistent. □

Note in the above Λ takes the role of:

"This sentence isn't true" (16)

This referred to as 'The Liar'[23]. Thus $(\mathbb{T}(\ulcorner \Lambda \urcorner) \Leftrightarrow \neg \mathbb{T}(\ulcorner \Lambda \urcorner)^{\urcorner})$ expresses:

The Liar is True if and only if The Liar is False (17)

Tarski's Theorem shows that, in First Order Arithmetic and its extensions, there cannot be a satisfactory[48] definition of truth [23].

III.d. A proof sketch of Gödel's First Incompleteness Theorem

We're now in a position to give a sketch of Gödel's First Incompleteness Theorem. Firstly, Gödel's encoding allows us to encode both formulae and proofs in arithmetic. In particular, we can construct a 2-input predicate $Proof(\ulcorner A \urcorner, \ulcorner B \urcorner)$ (read "A is the a proof of the formula B") with the property[49][14].

$$T \vdash A \leftrightarrow T \vdash \exists\, \ulcorner B \urcorner\ Proof(\ulcorner B \urcorner, \ulcorner A \urcorner) \tag{18}$$

We define

$$Prov(\ulcorner A \urcorner) \leftrightarrow_{df} \exists\, \ulcorner B \urcorner\ Proof(\ulcorner B \urcorner, \ulcorner A \urcorner) \tag{19}$$

We can now prove Gödel First Incompleteness Theorem:

Theorem 28. ***Gödel's First Incompleteness Theorem:*** *Any consistent 1OT extending PA is incomplete.*

Proof. (sketch [8, 23]) Assume T is consistent and Complete. As it is consistent and complete, by Eq.18 we have:

$$T \vdash (Prov(\ulcorner A \urcorner) \Leftrightarrow A) \tag{20}$$

That is $Prov(\ulcorner A \urcorner)$ satisfies Convention T. Thus by Tarski's Theorem, T is inconsistent. □

In the context of Gödel's Theorem, the fixed point of Tarski's Theorem (Λ) is known as "the Gödel Sentence of T", G_T. For this sentence, we have

$$T \vdash (G_T \Leftrightarrow \neg Prov(\ulcorner G_T \urcorner) \tag{21}$$

Thus Gödel's Theorem can be stated succinctly: For any consistent 1OT extending PA, there is a Gödel sentence G_F such that

$$T \nvdash G_T \text{ and } T \nvdash \neg G_T \tag{22}$$

As explained in Section IV, G_F is true in the standard model of Arithmetic.

PART IV LEFT OVER FAQ REGARDING GÖDEL'S FIRST INCOMPLETENESS THEOREM

Pretty reasonably, one might have many left over questions about Gödel's First incompleteness Theorem. In this section, I aim to quickly answer a few that came to my mind.

Can you give me an example of an unprovable true sentence about the Natural Numbers ?

So the Gödel sentence is a true unprovable sentence about the Natural Numbers. However, it is pretty artificially constructed. It is reasonable to ask: are there any important theorems which are unprovable? The answer is yes!

For example, there is the strengthened finite Ramsey theorem about colouring subsets of the Natural Numbers. This theorem is true for the Natural Numbers. However, the Paris-Harrington Theorem states this theorem is unprovable from Peano Arithmetic [24, 25]. There is also a very nice discussion about this in Ref.[17]

How do you prove such a sentence is True - or indeed unprovable?

Well ok, but how do you prove the strengthened finite Ramsey Theorem is true then? It turns out this was done from ZFC set theory [26]. ZFC set theory is another first order theory, which can be used to construct models of Peano Arithmetic (see later). However, it can also express considerably more than Peano Arithmetic and as such can be used to prove statements.

Of course being another first order theory which extends Peano Arithmetic, ZFC has its own Gödel sentences. There is no escape from Gödel.

---

[47] Check Truth table of $\neg(A \leftrightarrow \neg A)$ and then apply Gödel Kompleteness

[48] According to Tarski

[49] Note, proving this property is non-trivial

One might also be curious how one proves that the strengthened finite Ramsey theorem is unprovable. The idea here was to show that if it were provable, Peano Arithmetic could demonstrate its own consistency - which we will see in the next section, it cannot.

Non-Standard Models

The fact that theories like the Gödel Sentence and the strengthened Ramsey Theorem are unprovable means - via Gödel's Kompleteness Theorem - there are models of Peano arithmetic in which these theories are false. These are referred to as "Non-Standard Models". In fact, the sheer range of non-standard models of PA is mind boggling. For instance, the Löwenheim-Skolem Theorem [27] tells us there is a model of PA for every infinite cardinality. That is there is a non-standard model of the Natural Numbers with cardinality of the Reals!

"What are ... the Numbers?"[28]

By this point, you may have started to lose your grasp on what a number is. After all, way back in Part I, I simply introduced the standard model of arithmetic on the Natural Numbers as "$\mathbb{N} = \{1, 2, 3, ...\}$ with $1 + 1 = 2$, $1 \times 2 = 2$ etc". But what exactly is this standard model? How do we know its a model of PA?

To answer this, I would first recommend the excellent series of blog posts: Ref. [16, 29, 30]. They discuss how one can come to define the standard model of the Natural Numbers through **Peano Axioms**.

Peano Axioms are **not** the same as Peano Arithmetic. Peano Arithmetic is a first order theory, whereas Peano's Axioms include the axiom of induction

$$\forall P(P(0) \wedge \forall k(P(k) \Rightarrow P(k+1)) \Rightarrow \forall n P(n)) \qquad (23)$$

i.e. for any 1-input predicate, if P(0) is true and P(k) implies P(k+1) is true, then we can conclude P(n) is true for all. This axiom belongs to **Second Order Logic**. In First Order Logic, we can only quantify over variables; we can write $\forall x$ but not $\forall P$. Second Order Logic allows us to quantify over predicates too, i.e. we can write $\forall P$ ("For all Properties...").

Remarkably, Dedekind proved that any two models of Peano's Axioms are isomorphic [28]. Thus Peano Axioms really do pinpoint arithmetic on the Natural Numbers.

For this reason, an alternate definition of the standard model comes from ZFC set theory. The rough idea here is that 0 is identified with the empty set, $\emptyset$. The successor of t is then defined as $s(t) = t \cup \{t\}$. Thus we have:

$$0 \leftrightarrow_{df} \emptyset \qquad (24)$$
$$1 \leftrightarrow_{df} S(0) = \emptyset \cup \{\emptyset\} = \{\emptyset\} = \{0\} \qquad (25)$$
$$2 \leftrightarrow_{df} S(1) = \{\emptyset\} \cup \{\{\emptyset\}\} = \{\emptyset, \{\emptyset\}\} = \{0, 1\} \qquad (26)$$
$$\vdots$$

It is through this model, one can prove statements like the strengthened finite Ramsey theorem for the Natural Numbers. For further discussion see Ref, [31]

One Logic to Rule them all

It might occur to you - well if Peano Axioms uniquely specify Arithmetic on the Natural Numbers, why don't we just use Second Order Logic instead of First Order? One problem is that Second Order Logic has not been demonstrated to be Komplete. However, this is all I will have to say on Second Order Logic. For a further discussion see Ref.[9, 32, 33]

Is all of Maths incomplete?

Ok ok, coming back to First Order Logic, so Peano Arithmetic and its extensions are incomplete. Does that mean all first order maths is incomplete? No! For example, it can be demonstrated that certain axiomatization of Euclidean Geometry [34] are consistent and complete.[50]

Closer to Peano Arithmetic, another complete theory is Presburger Addition [35] - a theory with the model of addition on the Natural Numbers (but no multiplication). Before you think multiplication is somehow unique important to self-reference, there is also complete theories with multiplication and no addition[36][51].

As a final delightfully strange note, it turns out the theory of real numbers - the First Order Theory of Real Closed Fields - is also complete. [38][52]

---

[50] I find this very beautiful as Euclidean Geometry was in some sense the beginning of mathematicians attempts to axiomatize mathematics. It's also where I started being interested in the consistency of axioms

[51] See Ref.[37] for a good discussion.

[52] See Ref.[39] for discussion

IV GÖDEL'S SECOND INCOMPLETENESS THEOREM

In this final section, we now turn our attention to Gödel's Second Incompleteness Theorem. In Gödel's First Incompleteness theorem, we saw there are Gödel statements, which can neither be proven nor disproven; namely:

$$\text{T is consistent} \rightarrow T \nvdash G_T \tag{27}$$

where G_T is the Gödel statement following from the Diagonal Lemma that satisfies:

$$T \vdash (G_T \Leftrightarrow \neg Prov(\ulcorner G_T \urcorner) \tag{28}$$

and $Prov(\ulcorner A \urcorner)$ states there is a proof of A in the theory.

Gödel's Second Incompleteness Theorem give a very dramatic example of another such a statement. Specifically, consider a statement Cons(T) that encodes the statement "T is consistent". Specifically, we choose $\perp$ to symbolize an inconsistent statement (typically $\perp \leftrightarrow_{df} 0 = 1$[53][14]). We then define:

$$Cons(T) \leftrightarrow_{df} \neg Prov(\ulcorner \perp \urcorner) \tag{29}$$

Theorem 29. ***Gödel's Second Incompleteness Theorem*** *[14, 15] Let $(\mathcal{L}, T)$ be a 1OT extending PA. If $(\mathcal{L}, T)$ is consistent, then*

$$T \nvdash Cons(T) \tag{30}$$

Colloquially, "Complicated enough, consistent First Order Theories cannot prove their own consistency".

Proof. (sketch [14]) The idea of the proof is that we first formalise Eq.27 within T, i.e.

$$Cons(T) \Rightarrow \neg Prov(\ulcorner G_T \urcorner) \tag{31}$$

Remarkably, this formula does indeed hold. Namely:

$$\vdash (Cons(T) \Rightarrow \neg Prov(\ulcorner G_T \urcorner)) \tag{32}$$

Thus from the definition of the Gödel statement:

$$T \vdash Cons(T) \Rightarrow G_T \tag{33}$$

Therefore if $T \vdash Cons(T)$, then by $(\Rightarrow e)$ $T \vdash G_T$ which contradicts Gödel's First Incompleteness Theorem. Hence $T \nvdash Cons(T)$ □

[53] This contradicts axiom 4 of PA

Semantic Second Incompleteness Theorem

As before, we can express a semantic version.

Theorem 30. ***Gödel's Second Incompleteness Theorem [Semantic]*** *Let $(\mathcal{L}, T)$ be a 1OT extending PA. If T is consistent (equivalently, if T has a model [Thm. 22]) then*

1. *$T \nvDash Cons(T)$ (by Gödel Kompleteness)*
2. *(equivalently) There exists a "non-standard" model of T, M_2, such that $\models_{M_2} \neg Cons(T) = Prov(\ulcorner \perp \urcorner)$*

This was hard for me to get my head around[54]. Taking the example of Peano Arithmetic again. PA has a model and is therefore consistent. But this means there are non-standard models in which $Prov(\ulcorner \perp \urcorner)$ is true[55]. However, by the axioms, $T \vdash \neg\perp$, and therefore by consistency $T \nvdash \perp$[56]. Therefore, for me, this means $Cons(T)$ is not doing what it says it is. Perhaps an important observation is that $Cons(T)$ is just one encoding of consistency. We could have chosen a different $\perp$. I guess one way to look at it is no encoding of consistency is enough to guarantee consistency. I will leave this here in the hope someone can explain it to me[57]

Moving on, again we might think we could solve this problem by adding Cons(T) to our axioms, and thereby tautologically restrict ourselves to consistent models? Nope! Once again Gödel's Second Incompleteness theorem also applies to this new theory, and so on.

Well, can we be sure at least the standard model - Peano's Axioms - is consistent? The good news is, yes! We can prove the standard model is consistent from ZFC...but the bad news is ZFC is a first order theory extending PA, so it itself cannot prove its own consistency!

To finish, we ask one more question. Is there any theory which can prove its own consistency. Remarkably, Yes! These are curious theories that are expressive enough to be able to self-reference, yet not enough for the diagonal lemma to apply [42].

V CONCLUDING REMARKS

This essay turned out to be a lot more than I expected. We began with an attempt at a very quick

[54] and I'm not 100% sure I have
[55] For support of this understanding, see Ref. [40]
[56] Alternatively we must have $\nvDash \perp \leftrightarrow_{df} 1 = 0$ and therefore by Gödel Kompleteness $\vdash \perp$
[57] See Ref[40, 41] for further discussion

introduction to Classical Logic. This allowed us to understand Gödel's Kompleteness theorem[58]: $\vdash \leftrightarrow \models$. I don't think the power of Gödel's Kompleteness theorem can be overstated. For me, it is the lynch-pin for my understanding of Gödel's other statements.

Next we dived into Gödel's First Incompleteness theorem - that there are unprovable statements in PA and its extension- and saw its connection to the Liar Paradox through Tarski's Undefinability of Truth. Finally, we discussed one such unprovable statement in the form of Gödel's Second Incompleteness theorem - that PA is consistent.

I would like to finish by simply listing some questions I find interesting:

1. When approaching Classical Logic, I started with a belief that formal logic would constitute a foundation from which to build more complex ideas. However, we find that we get words like 'set' and 'function' appearing very early in our text. As we saw, this is a feature of the fact the Meta-language should be able to say everything the Object Language can say (it includes the semantics of the Object Language). This has a circular feel to it. A way out of this is to remove the assumption that formal logic is more foundational and instead realise the point of a Formal Language is to remove any ambiguity. Indeed, formal systems go so far as to completely remove meaning and reduce logic to generation and manipulation of strings. From the well defined formulae, we can then put back on meaning in a completely unambiguous way and justify our rules of inference by their preservation of truth (as understood in the Meta-language)[59].

2. The rules of inference treat logic as mechanical computation. To this observation, I have three comments

   (a) This perspective reminds me of John Searle's "Chinese Room" thought experiment in the field of AI and human consciousness [43]. The thought experiment is as follows: Say scientists have produced a program that speaks Chinese so well it can pass the Turing test (in Chinese). Then put a person who cannot speak Chinese in a closed room and feed pieces of paper with sentences in Chinese through a slit in the door. The non-Chinese-speaking person then manually implements all the steps of the algorithm and feeds the output back through the door. The question is: does that mean the person now understands Chinese? After all, they will pass the Chinese Turing Test... In our context, what do the rules of logic have to do with understanding truths? Logic seems to be how we verify knowledge but not necessarily how we create it.

   (b) If logic is mechanical, what is provable is constrained by the laws of physics. As expressed by David Deutsch [44]: "Though the truths of logic and pure mathematics are objective and independent of any contingent facts or laws of nature, our knowledge of these truths depends entirely on our knowledge of the laws of physics"

   (c) In fact, if rules of inference are just computation, why should that computation be classical? Could we construct a quantum rules of inference. Moreover, why shouldn't semantics be quantised? (see for instance Ref. [44])

3. The semantics of Classical Logic asserts that formulae (in an interpretation given a value assignment) are either True or False. However, we know natural language has paradoxical sentences, such as The Liar, which can be shown to be both True and False. Should we then develop a semantic system which allows such contradictions? As advocated by Priest [45]: "Suppose we stop banging our heads against a brick wall trying to find a solution, and accept the paradoxes as brute facts. That is, some sentences are true (and true only), some false (and false only", and some both true and false". Note, that if we take this approach, in order to avoid triviality, such a system must avoid the principle of Explosion (see E.g. 13) - that is to say it must be a *para-consistent* Logic [46]. What would be the implications of a mathematics founded on such a logic?

[58] with Soundness
[59] For further discussion see Ref. [**?** **?**]

[1] L. C. Paulson, "Lecture notes: Logic and proof," (2008).
[2] S. Shapiro and T. Kouri Kissel, "Classical Logic," https://plato.stanford.edu/archives/fall2022/entries/logic-classical/ (2022).
[3] Wikipedia, "Chomsky hierarchy," http://en.wikipedia.org/w/index.php?title=Chomsky%20hierarchy&oldid=1091751708 (2022).
[4] https://math.stackexchange.com/q/120568 ().
[5] W. Shakespeare and J. Fletcher, "Henry viii," (1623).
[6] J. Moschovakis, "Intuitionistic Logic," https://plato.stanford.edu/archives/fall2021/entries/logic-intuitionistic/ (2021).
[7] https://math.stackexchange.com/questions/2376279/minimal-set-of-rules-of-inference ().
[8] S. G. Simpson, "Lecture notes: Mathematical logic," (2005).
[9] J. Väänänen, "Second-order and Higher-order Logic," https://plato.stanford.edu/archives/fall2021/entries/logic-higher-order/ (2021).
[10] A. Tarski, J. Woodger, and J. Corcoran, *Logic, Semantics, Metamathematics - Papers from 1923 to 1938* (Hackett Publishing Company, Indianapolis, 1983).
[11] W. Hodges, "Tarski's Truth Definitions," https://plato.stanford.edu/archives/fall2018/entries/tarski-truth/ (2018).
[12] https://math.stackexchange.com/questions/121128/when-does-the-set-enter-set-theory ().
[13] K. Gödel, Montatshefte für Mathematik und Physik , 349–360 (1930).
[14] P. Raatikainen, "Gödel's Incompleteness Theorems," https://plato.stanford.edu/archives/spr2022/entries/goedel-incompleteness/ (2022).
[15] K. Gödel, Monatshefte für Mathematik Physik , 173–198 (1931).
[16] "Godel's Completeness and Incompleteness Theorems - LessWrong — lesswrong.com," https://www.lesswrong.com/posts/GZjGtd35vhCnzSQKy/godel-s-completeness-and-incompleteness-theorems (), [Accessed 14-Aug-2022].
[17] "Gödel's Incompleteness Theorem - Numberphile — youtu.be," https://youtu.be/O4ndIDcDSGc?t=644 (2017), [Accessed 14-Aug-2022].
[18] M. Glanzberg, "Truth," https://plato.stanford.edu/archives/sum2021/entries/truth/ (2021).
[19] N. F. Stang, "Kant's Transcendental Idealism," https://plato.stanford.edu/archives/spr2022/entries/kant-transcendental-idealism/ (2022), section 6.1.
[20] K. Easwaran, A. Hájek, P. Mancosu, and G. Oppy, "Infinity," https://plato.stanford.edu/archives/win2021/entries/infinity/ (2021).
[21] "Math's Fundamental Flaw — youtube.com," https://www.youtube.com/watch?v=HeQX2HjkcNo&t=954s, [Accessed 15-Aug-2022].
[22] P. Raatikainen, in *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta (Metaphysics Research Lab, Stanford University, 2022) Spring 2022 ed.
[23] T. Bolander, "Self-Reference," https://plato.stanford.edu/archives/fall2017/entries/self-reference/ (2017).
[24] J. Paris, in *HANDBOOK OF MATHEMATICAL LOGIC*, Studies in Logic and the Foundations of Mathematics, Vol. 90, edited by J. Barwise (Elsevier, 1977) pp. 1133–1142.
[25] Wikipedia, "Paris–harrington theorem," (2022).
[26] Wikipedia, "Zermelo–fraenkel set theory," (2022).
[27] T. Skolem, Videnskapsselskapet Skrifter, I. Matematisk-naturvidenskabelig Klasse , 1 (1920).
[28] R. Dedekind, (1888).
[29] "Logical Pinpointing - LessWrong — lesswrong.com," https://www.lesswrong.com/posts/3FoMuCLqZggTxoC3S/logical-pinpointing (2012), [Accessed 14-Aug-2022].
[30] "Standard and Nonstandard Numbers - LessWrong — lesswrong.com," https://www.lesswrong.com/s/SqFbMbtxGybdS2gRs/p/i7oNcHR3ZSnEAM29X (), [Accessed 14-Aug-2022].
[31] https://math.stackexchange.com/questions/186506/consistency-of-peano-axioms-hilberts-second-problem ().
[32] E. Yudkowsky, "Second-Order Logic: The Controversy - LessWrong — lesswrong.com," https://www.lesswrong.com/s/SqFbMbtxGybdS2gRs/p/SWn4rqdycu83ikfBa (2013), [Accessed 14-Aug-2022].
[33] S. Shapiro, "The "triumph" of first-order languages," in *Logic, Meaning and Computation: Essays in Memory of Alonzo Church*, edited by C. A. Anderson and M. Zelëny (Springer Netherlands, Dordrecht, 2001) pp. 219–259.
[34] A. Tarski, (1959).
[35] Wikipedia, "Presburger arithmetic," https://en.wikipedia.org/wiki/Presburger_arithmetic ().
[36] Wikipedia, "Skolem arithmetic," https://en.wikipedia.org/wiki/Skolem_arithmetic ().
[37] https://philosophy.stackexchange.com/questions/38128/why-does-multiplication-lead-to-incompleteness-where ().
[38] Wikipedia, "Complete theory," https://en.wikipedia.org/wiki/Complete_theory ().
[39] https://math.stackexchange.com/questions/362837/are-real-numbers-axioms-a-consistent-or-complete-system ().
[40] https://math.stackexchange.com/questions/2038565/completeness-and-incompleteness ().
[41] https://math.stackexchange.com/questions/1383286/godels-second-incompleteness-and-the-assumption-of ().
[42] Wikipedia, "Self-verifying theories," (2022), [Online; accessed 14-August-2022].
[43] D. Cole, "The Chinese Room Argument," https://plato.stanford.edu/archives/win2020/entries/chinese-room/ (2020).
[44] D. Deutsch, A. Ekert, and R. Lupacchini, Bulletin of Symbolic Logic **6**, 265 (2000).
[45] G. Priest, Journal of Philosophical Logic **8**, 219 (1979).
[46] G. Priest, K. Tanaka, and Z. Weber, "Paraconsistent Logic," https://plato.stanford.edu/archives/spr2022/entries/logic-paraconsistent/ (2022).