Gödel's Incompleteness Theorems and Self-Referential Structures in *Tractatus Logico-Philosophocus*

J. Plasmeier

Introduction

Written while he was serving time as a prisoner of war in World War I, *Tractatus Logico-Philosophocus* is Ludwig Wittgenstein's attempt to address many of the questions of philosophy in the form of a list of propositions. These propositions attempt to describe the limit of language and thought to describe facts about the world. Wittgenstein concludes that the problems of philosophy are merely confusions of language, and cannot possibly have factual answers. *Tractatus* consists of seven main propositions:

1. The world is everything that is the case.
2. What is the case (a fact) is the existence of states of affairs.
3. A logical picture of facts is a thought.
4. A thought is a proposition with a sense.
5. A proposition is a truth-function of elementary propositions. (An elementary proposition is a truth-function of itself.)
6. The general form of a proposition is the general form of a truth function, which is: $[\bar{p}, \bar{\xi}, N(\bar{\xi})]$. This is the general form of a proposition.
7. Whereof one cannot speak, thereof one must be silent.

The purpose of this paper is to examine the relationship between *Tractatus* and the results of some extraordinary theorems put forth (13 years after *Tractatus* was written) by the brilliant logician Kurt Gödel. Known commonly as Gödel's First and Second Incompleteness Theorems, these theorems express truths about the completeness and consistency of formalized systems of axioms. Both works have been hugely influential and carry philosophical implications beyond their original scope.

The First Incompleteness Theorem

Gödel's First Incompleteness Theorem (from this point until the second section of the paper referred to as simply Gödel's Theorem) proves that any formal axiomatic system (capable of expressing at least basic arithmetic) cannot be both consistent and complete. For a formal system to be inconsistent, there exists some valid theorem in the system that is both true and false. For a formal system to be complete, the system must provide a proof of every sentence within the language of the system (or prove its negation). Now, if the system is consistent, one may form a statement that is true but not provable. Denote the formal system T and such a statement G. In plain English, G has the meaning "G cannot be proved within T." If T were complete, T would provide a proof for G and thus simultaneously prove the negation of G, leaving T inconsistent. If T were consistent, then G cannot be proved and T is thus incomplete.

One of the results of Gödel's Theorem is the fact that one cannot prove, within a formal system of logic, that the system itself does not contain a contradiction. This is similar in language to Proposition 7 of *Tractatus*[1]: "Whereof one cannot speak, thereof one must be silent." The two works, crafted independently of each other, contain similar notions of "unspeakable" propositions. However striking these similarities may appear, they are dissimilar in terms of meaning, and the propositions do not imply each other. In Gödel's Theorem, these propositions are equivalent to creating a well-formed, valid formula that states, "(within this system) this formula is not provable in [this system]." For Wittgenstein, the only propositions that have meaning are propositions that express a truth about the world. These propositions will be referred to as "valid" throughout the paper. This implies that most statements in language are purely syntactical and meaningless. This is a much bolder statement, and includes the totality of mathematics as a syntactical construction with no direct relation to reality. This includes not only Gödel's Theorem and all of its implications but the very systems that Gödel's Theorem refers to.

Although Gödel's Theorem is regarded by *Tractatus* as senseless, the two works do have some elements in common. In Proposition 3.032 it is stated that, "To present in language anything which 'contradicts logic' is as impossible as in geometry to present by its coordinates a figure which contradicts the laws of space; or to give the coordinates of a point which does not exist." Wittgenstein brings to the discussion an entire theoretical class of propositions, which are so illogical that the propositions themselves cannot be conceived. By necessity, this implies paradoxical propositions like the true statement, "This statement is not true," are in fact logical because the statement can be precisely given. This is in accordance with Gödel's Theorem, as Gödel's Theorem hinges on the validity of paradoxical, self-referential propositions.

Further propositions in *Tractatus* are less immediately compatible with Gödel's Theorem. Proposition 3.332 states that, "No proposition can say anything about itself, because the propositional sign can be contained in itself." Wittgenstein then demonstrates in Proposition 3.333 that given a function F, within F(F(x)) the inner and outer functions differ. The letter F by itself signifies nothing, so the same result can be attained by writing F(F(x)) as:

$$\exists\, \phi : F(\phi x) . \phi x = Fx.$$

Wittgenstein boldly claims that this dismisses Russell's Paradox[2]. So long as this notion is consistent with and compatible with the methods used in Gödel's Incompleteness Theorem to denote self-reference, the two works will be consistent and compatible with each other.

[1] From here on, all propositions quoted from *Tractatus* will simply be introduced by "Proposition."

[2] To the Author of this paper, Wittgenstein fails to accurately address Russell's paradox here and merely provides a means of restating what can be still be a paradox. I will address this shortcoming later in the paper.

Self-reference is essential to the proof of Gödel's Theorem. In order to observe the details of the proof, Peano Arithmetic (a basic formal system of axioms for which the natural numbers and basic arithmetical operations are a model of) will serve as the system to prove Gödel's Theorem within. The Gödel number of a formula σ, denoted $\lceil\sigma\rceil$, is a unique numerical representation of the symbols of the formula σ. This allows formulae to be encoded from a finite string of symbols (e.g. 1+1) into a natural number. Then, this allows construction a function denoted $theorem\ (f)$ defined such that for every formula φ in PA:

$$\mathbb{N} \vDash theorem(\lceil\varphi\rceil) \leftrightarrow \vdash_{PA} \varphi$$

That is, for every formula φ in the system (PA), the function $theorem\ (\varphi)$ is logically equivalent to the statement:

"φ is a logical theorem in PA."

From there, Gödel constructs a formula denoted $\varphi\ (x_0)$ defined in plain English to mean:

x_0 *codes a formula* ψ *with the property: "*$\mathfrak{D}(\psi)$[3] *is not a logical theorem in PA."*

Then, define a formula $\sigma := \mathfrak{D}(\varphi)$. By definition of $\mathfrak{D}(\psi)$, the meaning of σ can be restated in plain English as:

"the Gödel number of φ *codes a formula* ψ *and* $\mathfrak{D}(\psi)$ *is not a logical theorem in PA."*

However, by definition, the Gödel number of φ codes the formula φ and so the meaning of σ can be stated as:

"$\mathfrak{D}(\varphi)$ *is not a logical theorem in PA."*

Now, recall that $\sigma = \mathfrak{D}(\varphi)$ by its original definition. Thus, the meaning of σ is "σ is not a logical theorem in PA," though we have taken care to define σ as a logical theorem in PA. Thus, PA is inconsistent.

At first, this proof appears to be inconsistent with Wittgenstein's rule that propositions may not contain themselves. One of the results of Gödel's Theorem is Gödel's self-referential lemma[4], which formally states:

$$\mathbb{N} \vDash \sigma \leftrightarrow \varphi(\lceil\sigma\rceil)$$

[3] $\mathfrak{D}(\varphi)$ is the diagonalisation of φ. Details of its precise definition is beyond the scope of this paper.

[4] The purpose of this paper is not to reconstruct the proof of Gödel's Theorem; I am taking its results as true.

In plain English: the set of natural numbers satisfies the closed formula σ, and σ is logically equivalent to some formula φ taking the Gödel number of σ as an argument. In this lemma it is especially clear how Gödel makes use of self-reference. In plain English, we can describe the formula $\varphi\ (x_0)$ as:

"the formula encoded by the Gödel number of the free variable x_0"

By construction, one could encode $\varphi\ (x_0)$ into a Gödel number and pass itself as a parameter to itself. This returns us to the subject of Proposition 3.3333. By Wittgenstein's explanation, $\varphi(x_0)$ the parameter value and $\varphi\ (x_0)$ the "proposition" are two separate propositions. However, this does not address the fact that both symbols share the exact same definition. It is not two separate instances of the same form but semantically the same object. For this reason, Gödel's Theorem and *Tractatus* do not appear to be fully compatible, as the proof of Gödel's Theorem relies on a notion of self-reference forbidden by *Tractatus*.

However, this only applies to valid propositions adhering to reality. As stated earlier, Gödel's Theorem (and all of mathematics) are not contained within the set of valid propositions and thus not subject to the restriction of self-reference. Strictly within the details of both works then, there is nothing in *Tractatus* that would be incompatible with the proof of Gödel's Theorem regarding the restriction of self-referential formulae. In fact, it is necessary for the propositions of Gödel's Theorem to be invalid. Thus, according to *Tractatus*, self-referential structure is absent within propositions describing the reality of the world.

Recursive programming in computer science relies on functions calling themselves to implement algorithms. In the way recursion is handled by the programming languages, we become aware of the difference between self-reference from within a nonsensical, metaphysical proposition and from within valid propositions. Semantically, a recursive function calls itself in the same way that a proposition becomes self-referential in *Tractatus*. However, in the actual runtime of the program, a new stack frame (space in memory) is allocated each time the function is called. The physical existence of the function is comprised of multiple instances of the same function definition. A specific area of memory in that computer has a specific (albeit very complicated!) pattern of electric charge displacements. This physical phenomenon is described by valid propositions, and accordingly avoids actual self-reference.

In this case, we are enabled to push closer to the limit of what can and cannot be described by valid propositions by examining the physical impossibility, but logical permissibility (and potential necessity) of self-reference as described by *Tractatus*. Observe that the previous sentence itself contains an invalid proposition. (As well as the majority of sentences in this essay!) Next we will show that only invalid propositions can describe the validity of (valid) propositions and the implications thereof.

The Second Incompleteness Theorem

Gödel's Second Incompleteness Theorem[5] states that if any formal axiomatic system (that includes basic arithmetic truths and notions of provability) proves itself consistent, then that system is in fact inconsistent. This implies that in order to determine the consistency of a system, one must do so by a separate, external system than the system in question. This notion of determining truth externally is found within the propositions of *Tractatus*.

Observe that one could not possibly attempt to actually apply either of Gödel's Incompleteness Theorems directly to *Tractatus*, as *Tractatus* does not meet the requirements for a formal axiomatic system. This is indirectly addressed in Proposition 6.54, which states:

> My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless, when he has climbed out through them, on them, over them. (He must so to speak throw away the ladder, after he has climbed up on it.) He must surmount these propositions; then he sees the world rightly.

Observe that this is similar to how mathematics uses axioms. Mathematics takes (axiomatic) propositions to be true with consideration, but not proof, just as is requested by Wittgenstein here. Wittgenstein acknowledges that the propositions comprising *Tractatus* are precisely those deemed invalid by the work. Thus, whether or not the propositions are inherently true (or false) is apparently left for the reader to determine.

It is immediately clear that we cannot apply Gödel's Theorem directly to *Tractatus* (as an axiomatic system) because there is no sensical model for *Tractatus*. *Tractatus* does not include relational symbols or functional symbols[6]. Hence, *Tractatus* is not afforded the formality of Gödel's Theorems. However, many similar ideas, specifically within Gödel's Second Incompleteness Theorem, can be found within *Tractatus*.

At the core of Wittgenstein's philosophy of mathematics (in *Tractatus*) is the distinction between valid propositions about reality and mathematical propositions. To Wittgenstein, mathematical propositions fall within the same class of senseless propositions mentioned above because the truthiness or falsehood of a mathematical proposition is not contingent on "the state of affairs." Proposition 2.224 states, "It cannot be discovered from the picture alone whether or not the picture is true." Proposition 4.12 states, "To be able to represent the logical form, we should have to be able to put ourselves with the propositions outside logic, that is outside the world." This echoes the very same requirement of some external means to determine the truth of a proposition.

---

[5] Gödel's Theorem now refers to Gödel's Second Incompleteness Theorem.

[6] There are in fact a number of propositions (not introduced in this paper) that involve the definition of a primitive arithmetic. Those propositions do not comprise a language for *Tractatus* as a whole.

Furthermore, Proposition 6.54 (stated above) is reliant on an idea similar to the result of Gödel's Theorem. [As an aside, consider the validity of a set of *Tractatus* propositions to rely on the validity of individual statements. Thus, a set of propositions is valid if and only if each proposition is valid on its own.] If we substitute provability with validity (that is, a proposition with a sense is a valid proposition that represents a "state of affairs"), the theorem states that if a set of propositions is capable of describing themselves as being valid, then that set does not have validity on the whole. This conflates the rigorous definition of Gödel's Theorem, but the idea is present. We can further reduce this statement to apply to individual propositions and state that if a proposition in *Tractatus* were to describe itself as being a valid proposition, it would inevitably not be a valid proposition. This statement is eerily similar to the result of Gödel's Second Incompleteness Theorem.

Earlier, the potential necessity of self-reference in *Tractatus* was mentioned. Since then, we have shown that in order to establish the notion of valid propositions, invalid propositions were required to define validity. Proposition 6.54 clearly demonstrates self-reference; by referring to "[all of] my propositions" as invalid, this includes Proposition 6.54 itself. Taking this as a subset, we can consider part of Proposition 6.54 to state, "This proposition is not valid." This proposition is nearly identical to the Gödel sentence, "this theorem is not provable," with the essential difference being the *Tractatus* proposition is not guaranteed to be a valid proposition by its construction, unlike the Gödel sentence which is actually a provable theorem. It is now in plain sight that the proof of Gödel's Theorem is about more than mere self-reference; it also involves the ability to guarantee certain conditions for which the self-referential statement can contradict. This is what leads to proof of incompleteness in PA and similar systems.

We now have another argument against the possibility of applying Gödel's Incompleteness Theorems directly to *Tractatus*: in order to construct an equivalent "Gödel Proposition," that is, a valid proposition (in accordance with the rules of *Tractatus*) which states, "This proposition is invalid." This is plainly untenable as a possible "state of affairs" and obviously lacks a corresponding logical picture, not to mention the restriction imposed by Proposition 3.3332 that (valid) propositions cannot contain themselves. Thus, we are unable to form the construction and apply Gödel's Incompleteness Theorem (specifically the First).

Thus, we conclude that when examined in detail, *Tractatus Logico-Philosophicus* and Gödel's Incompleteness Theorems have little technical overlap or implication toward each other. However, many of the ideas required to formulate both works are shared. Gödel's Incompleteness Theorems and *Tractatus* are both captivating achievements of human intellect that have changed the way in which humans think about math and philosophy. Neither the results of Gödel's Incompleteness Theorems nor the conclusions of *Tractatus* are considered to be terribly practical, but more so profound. The method of thought present in both works has implications beyond the scope of their immediate application and indicates the presence of some deeper, more general underlying structure of self-reference and reality yet to be discovered.

References

- All propositions from *Tractatus* are referenced from:
  Wittgenstein, Ludwig, and Bertrand Russell. *Tractatus Logico-Philosophicus*. New York, NY: Cosimo Classics, 2007. Print.
- All equations in the proof sketch of Gödel's Incompleteness Theorem are referenced from:
  Goldstern, Martin, and Haim Judah. The Incompleteness Phenomenon: A New Course in Mathematical Logic. Wellesley, Mass.: K Peters, 1995. Print.