DOI: 10.24906/isc/2021/v35/i4/210002

Feature Article

Gödel's Incompleteness Theorems: An Interdisciplinary Review

Aditya Dwarkesh[1] and Satbhav Voleti[2]

Abstract

In the following article, authors offer a modern proof of Gödel's incompleteness theorems. Authors then briefly recount what its immediate reception in the scientific community was like, and finally appraise the ultimate impact it has had, and issues it has raised, in a wide variety of fields—ranging from mathematics to philosophy of mind.

Key Words: Incompleteness, consistency, philosophy of mathematics, Turing machine, anti-mechanism

Introduction

Formal mathematical systems underlie most of modern mathematics. The 20th century and the various programs of formalizing previous mathematics brought about interest in understanding the possibility of formalization in principle. Frege and Russell made large leaps forward; they envisioned new foundations through predicate logic. Others sought other foundations. The question as to the scope of this formalization was an open question which eventually culminates in the fracturing of the optimism of early 20th century goals of complete and comprehensive formalization through Gödel's Incompleteness Theorems. The goal was to make a system where all statements are either provable or disprovable and no two contradictory statements could be proven. Gödel showed this to be impossible for sufficiently complicated formal systems.

Gödel's theorems are definitively the most major result in the foundations of mathematics. They are very general and incredibly profound even in the limited scope that one may afford the theorems in context of formal systems. Gödel was the first person ever to claim that certain statements of mathematics exist which cannot be proved or disproved (in the relevant systems). In fact, the second incompleteness theorem was the first ever proof that there was a particular statement that was 'independent' of the axioms of the foundations of mathematics, no matter how many extensions one makes; the particular statement being the consistency of the system itself. This was a big step for mathematics. Authors shall see exactly why in the course of this article.

[1]aditya.dwarkesh@gmail.com
[2]Corresponding author, satbhav.voleti@gmail.com
Indian Institute of Science, Education and Research, Kolkata, Mohanpur, Nadia- 741 246, West Bengal, India
ORCID: Aditya Dwarkesh: http://orcid.org/0000-0002-9534-0205
ORCID: Satbhav Voleti: http://orcid.org/0000-0002-2907-4113

Gödel first publicly talked about his incompleteness results on September 7, 1930, in a discussion in Konigsberg in the Second Conference on the Epistemology of the Exact Sciences. He had shown Carnap the same on the 26th of August of the same year. Gödel was present at the conference to present a short talk on his dissertation—the proof of the completeness of first-order logic. In the discussion between Scholz, Heyting, Carnap, von Neumann, and Hahn—after the talks—Gödel mentions how it is possible to prove ~(for all) x: not-F(x) in classical arithmetic even if one realizes through other considerations that all finite numbers have the property not-F. This is related to the idea of w-consistency. Further he says:

"One can (assuming the consistency of classical mathematics) even give examples of propositions (and indeed, of such of the type of Goldbach or Fermat) which are really contentually true but are unprovable in the formal system of classical mathematics. Therefore, if one adjoins the negation of such a proposition to the axioms of classical mathematics, one obtains a consistent system in which a contentually false propositionis provable." (Gödel, 1930)

In this article, authors survey the wide-ranging uses of the incompleteness theorems after an explanation of their origins. The incompleteness theorems have had a rich history consisting of heated debates and extravagant uses (and as is often said, abuses). Firstly, authors discuss the criticisms that the theorems received followed by their subsequent acceptance and eventual fossilization in logic text books. Of these criticisms, Wittgenstein's is of special interest. Authors also discuss the applications of the theorems in logic and metamathematics in the form of Tarski's undefinability theorem, Church's undecidability theorem, Hilbert's second and tenth problems. The various positions in the philosophy of mathematics all underwent some transformation in light of the incompleteness theorems. This is another point of authors' interest. More recently, Gregory Chaitin has shown an equivalent statement in information theory. But perhaps the most well-known use of Gödel's theorems (by Roger Penrose and John Lucas) comes in the philosophy of mind to argue that the mind can not be a machine. Towards the end, authors discuss this argument alongside its precursor in Gödel himself and a more elusive account of self-referentiality through Douglas Hofstadter.

The incompleteness theorems are broad and require a careful and nuanced understanding of their statements to proceed further. This is the aim of the next section.

Proof

It is crucial to comprehend two important basic notions in full formality before one begins trying to understand Gödel's theorems. These are *consistency* and *completeness*. In order to explain these properties, it will be necessary to first discuss what it is that they are properties *of*.

In first-order logic, a *theory* is a set of sentences which, in a certain sense, "leaves no stone unturned": If a collection of sentences within the set S imply the truth of another sentence s, then, in order for the set S to qualify as a theory, s must also be a part of it.

A sentence is any formula which is capable of having a truth-value. Therefore, "x<y", where x, y are variables, is not a sentence, because one cannot say of it cannot say of it whether it is true or false (both x and y are said to be 'free variables'); on the other hand, replacing the variable with constants to read "2<3" turns it into a sentence with a truth-value (which is usually true).

Consistency and completeness are properties of first-order theories.

The first of these feels rather simple. A theory is consistent if it doesn't affirm both a sentence as well as its negation. Completeness, however, appears to be something more ambitious: A theory is complete only if, *for every possible sentence s,* it affirms either s or its negation.

It will be worthwhile to further parse out precisely *when* a theory ends up affirming a given sentence. This is determined by the fact that the theory is "in" first-order logic.

First-order logic is associated with a collection of certain formal systems, which, in turn, are defined by a set of *axioms* coupled with certain *rules of inference*, which enable one to "derive"

one formula from a set of others. A formal system it typically treated as a totally abstract, syntactic object which merely instructs one on how to perform the "symbol-shunting". As such, a formal system is constructed in such a way as to make the formal deductions in it "emulate" natural-language proofs.

A theory is defined by its own set of axioms. For example, Peano arithmetic is the name given to the first-order theory which models natural numbers. The axioms of this theory can be thought of as sentences picked up on the basis of some intuitive motivation (for example, the sentence "For all x, $x + 0 = x$"). From them, one may deduct theorems in the theory using the rules of inference associated with the formal system one is working within.

While Gödel's theorem holds for the theory of Peano arithmetic, it is worthwhile to note that even theories 'weaker' than that (in the appropriate sense, which shall be explained) may succumb to the same.

The basic idea of Gödel's proof (let's say, for Peano arithmetic, abbreviated to P) takes the following steps:

1. Every recursive relation is representable in P.
2. Every formula in P can be arithmetized, such that for each formula corresponds a unique integer.
3. P_T defined in the following manner is recursive (and thus representable in P): (x, y) $\in$ P_T if and only if the formula associated with the integer *y* constitutes a proof for the formula associated with the integer *x*. The representation of this relation in P is called the "provability predicate", and shall be denoted by $\mathbf{P_T}$.
4. For every formula F(x) with exactly one free variable, there exists a sentence A such that F(a) is true in P if and only if A is (where a is the integer associated with the sentence A).
5. Let B(x) be a formula which reads "There exists a proof for the formula x." By 4, there exists a sentence G such that ~B(g) is true if and only if G is; that is, B(g) is *false* if and only if G is *true*.
6. If G or its negation are provable in P, P is not consistent; if neither are provable, P is not complete.

This representation of the argument in natural language should by no means hoodwink the reader into believing that any of the steps in the proof are "obvious"; they each require thorough and rigorous argumentation, and direct attempts to sum it up with one-liners such as "This sentence is true but unprovable" fall short of doing so by something more than a respectful distance.

As such, these steps call for some more elaboration.

1. The notion of a recursive relation was still very much in its formative stages when Gödel first concocted his proof. Ironically, the definition of a recursive function is given recursively. One calls the simple functions which map an input to the next number the successor function. The constant function maps an input to a fixed natural number. The projection map takes an array of natural numbers as input and maps it to a particular entry of the array (say the ith entry). These are the three basic recursive functions. The composition, i.e., putting together of any two recursive functions is recursive. Apart from this, a minimization of a recursive function is a function which returns the smallest value for which the recursive function outputs zero. A primitive recursion operator defines a new function from two known functions, which is defined as one function when zero is an input and defined as the second function's output for every successor of zero. Now a relation is recursive if there is a function whose inputs which give zero are exactly those in the relation. A relation is said to be representable over a theory when there exists a formula which can be proven in the theory when the inputs are exactly those of the relation and disproven otherwise. Through an explicit construction one can show that every recursive relation is representable in P. The Church-Turing thesis states that every predicate that can be computed using finite steps using only a set of instructions is recursive.

2. It is fairly easy to see explicitly how such an arithmetization is possible. One first encodes each symbol in the first-order language as a unique integer, and then a sequence of symbols via $p_1^{a1}*\ldots p_n^{an}$, where p_n is the nth prime and a_n is the code for the nth symbol. By the uniqueness of prime factorization, this function is injective, and so each formula will correspond to a unique integer.
3. Given that the formal system has finitely many axioms, it is intuitively clear that checking if a proof-procedure is valid or not will be a computable procedure. By the Church-Turing thesis mentioned in #1, this makes the formula P_T recursive and thus representable in the system.
4. By the Church-Turing thesis, one can show that there is a recursive function d such that it maps a number *a* to the Gödel number of the formula A(*a*), where *a* is the Gödel number of A(x), where x is a variable. Now since d is recursive, a formula D(x,y) represents the function.

Now one defines a formula ψ which says that for all y D(x,y)=>F(y), where F is some formula with one free variable. One calls c the Gödel number of ψ. Substituting c in ψ one defines ϕ. The Gödel number of ϕ is, say, q. It is clear that d(c)=q. One subtitutes this into ψ and notice that it can be derived from a particular tautology in logic. One finally arrives at the statement that there is a formula ϕ such that ϕ and F(G(ϕ)) are equivalent.

The final steps constitute the construction of B(x) and G, and the hangman's move is now revealed:

i) If G is true, then B(g) is false; in other words, G (whose Gödel number g was) will be true but unprovable, making the system incomplete.

ii) If G is false, then B(g) is true; in other words, G will be false but provable, making the system inconsistent.

This completes Gödel's first incompleteness theorem: *Any consistent recursively axiomatized extension of Peano arithmetic is incomplete.*

Gödel's second incompleteness theorem is as follows: *No consistent recursively axiomatizable extension of Peano arithmetic can prove its own consistency.*

Let C denote the statement, "No proof of a contradiction exists in the theory". This can be achieved by defining C as ~B(z), where z may the Gödel number of any contradictory formula Z, say, 0=1.

The crucial element in the proof of the second theorem is the fact that, if "B(a) implies A" is provable in the system—where a is the Gödel number of the sentence A—then, so is A.

Since one assumes the consistency of the system (and so the unprovability of Z in it), the provability of "B(z) implies Z" will have to imply the provability of ~B(z), which authors defined C to be. But the provability of "B(z) implies Z" also entails the provability of Z.

So, one has "1 is provable <=> 2 is provable", and "1 is provable => 3 is provable". From this, it is easy to see that authors have "2 is provable => 3 is provable". Replacing 1 with "B(z) implies Z", 2 with C and 3 with Z yields "C is provable => Z is provable". Since Z is a contradiction and the system is consistent, this means that C must be unprovable, thereby concluding the theorem's proof.

Criticism and Reception

Largely, the reception of the incompleteness theorems was not controversial[1]. It came to be soon accepted widely and by 1952, it was even a part of Kleene's classic text, *Introduction to Metamathematics* [2]. Later, Kleene even claimed that no one had doubted the second incompleteness theorem [3]. Although this may not completely reflect the reactions of the time, for mathematicians were starting to grow suspicious of the existence of unprovable statements before the incompleteness results; Brouwer for one had suggested that mathematics was inexhaustible and not completely formalizable. Gödel had wanted to provide another detailed proof for the second theorem in anticipation of any friction from the mathematical community at the time [4]. This was not necessary. As Gödel himself remarked, the prompt acceptance of his results was one of the reasons of abandoning the second proof.

The paper that formally presented the proofs was published in January 1931–the now famous, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I*[4]. At this time Bernays and Gödel corresponded on these results; discussing whether introducing an ω-rule (when $F(x)$ is a quantifier free formula, and if it is true for every particular x, then it is true for all x; this is more of an informal rule of inference) would save from incompleteness and whether it would be in the spirit of Hilbert's program (which shall be discussed in later sections).Von Neumann was impressed by these results and accepted them in stride even coming to the second theorem on his own. But of course, all was not sunshine and roses. In 1931, he also travelled to Bad Elster, Germany where he presented his paper. Ernst Zermelo was present at this conference.

Zermelo was perhaps the harshest critic of Gödel [1]. He criticized the incompleteness results by alleging that it put arbitrary finitistic restrictions on statements. He thought of quantifiers in terms of conjunction or disjunctions of statements. Proving all of such statements would itself involve induction which is not strictly finite in nature, which Zermelo thinks amounts to a proof of the general statement. This would remove the syntactical nature of the statements. This amounts to a misunderstanding of the *syntactic* results of Gödel. On this criticism, Zermelo dismissed the results. Gödel and Zermelo corresponded personally later, where Zermelo had claimed to find a loophole in Gödel's proofs. He had mistakenly assumed that truth would be definable within the system allowing to present a liar-paradox-type situation. However, this is not true as one will see that arithmetical truth isn't definable in arithmetic. Gödel wrote a ten-page response to Zermelo, explaining his proof. Zermelo thanked Gödel for the clarifications on the proof but it seems that he never fully grasped the gravity of the incompleteness results [6].

Emil Post might have anticipated Gödel's results in a real way, where he had independently come to the realization that there would be independent propositions in the system of the *Principia*. However, he did not present a complete proof and indeed said that he could not have replaced the "splendid actuality of your [Gödel's] proof" [7].

Paul Finsler, a German mathematician, wrote to Gödel in 1933 [1] implying that his work in 1926 [8] had prefigured that of Gödel himself. Gödel replied to Finsler and explained that Finsler had not been using a defined system at all and that his arguments would not be formalizable in a formal system. Finsler did not take Gödel's response well. In a flurry of anger, Finsler lashed out against Gödel, claiming that one need not define the system completely and sharply to make sense of the statements and that he could on similar grounds attack Gödel's proof because the he hadn't proved the consistency of Peano's axioms; even though the latter was shown to be impossible in the same paper! In an unsent paper that Gödel wrote to a student, he says that Finsler's paper "contains obvious nonsense" [1]. By 1939, Hilbert and Bernays' had presented the complete proof of the second incompleteness theorem in their *Die Grundlagen der Mathematik II*[9] which quashed the friction from within the logical community. The technical qualms about the theorems had no more life in them. The philosophical dimension, however, remained open.

The Vienna Circle was a group of philosophers who used to meet regularly in the 1920s-1930s. Kurt Gödel was a regular attendee along with Carnap, Reichenbach, Schlick and others. In January 1931, Gödel presented the incompleteness results in front of the Vienna Circle [10]. The members of the circle reacted differently to the theorems. Felix Kaufmann thought that consistency could only be seen through an 'intellectual insight'. Von Neumann believed that it had destroyed Hilbert's program and that intuitionism was vindicated. Carnap knew the results beforehand and still went through and presented logicism in the Königsberg lecture to inform of the origins of logicism and also as the impact of Gödel's results was not immediately obvious. Confusions about object and metalanguage seem to be abundant. Chwistek made this mistake and then quickly corrected himself.

Although he was not present at this meeting, having learnt of Gödel's results later, the obscure

genius Ludwig Wittgenstein made some of the most perplexing remarks ever to be made on the theorems.

In his notes on the philosophy of mathematics, later translated and published as *Remarks on the Foundations of Mathematics*[11], Wittgenstein commented about provability and truth in a section that is now colloquially known as the "notorious paragraph". Long before this, Wittgenstein in the *Tractatus Logico-Philosophicus* expresses that mathematical propositions lack *sinn* (sense); they are "pseudo-propositions" (cf. §6.2 [12]). They plainly state the equivalence of the expressions on either side of the equality; they do not refer to anything. And in *Tractarian*-Wittgenstein, this disqualifies them from being capable of truth or falsity. Although whether or not Wittgenstein was a logicist is up to debate, it seems fairly clear that he thought of mathematics as mainly formal and syntactic (cf. §6.22, 6.2321 [12] and II, §12 [13]). After his return to philosophy in 1929, Wittgenstein propounded a return to philosophy, moving into a finitism regarding mathematics. For the middle Wittgenstein, mathematical propositions are either proven to be true, proven to be false, have a decision procedure to be proven or disproven or not mathematical propositions at all (cf. II, §23 [13]). For Wittgenstein, there is no syntax-semantics distinction in mathematics, everything is syntax. He rejects any undecidability in mathematics (what is undecidable is not mathematical at all). Jumping ahead to his later work in the *Remarks*, for Wittgenstein's idiosyncratic constructivist attitude, 'true in a system' and 'provable in a system' are equivalent (cf. App. I §5-7 [11]). Here, Wittgenstein builds on this basis to make some comments which leave one either bewildered or angered. Gödel himself was the latter. He thought that Wittgenstein 'advanced a completely trivial and uninteresting interpretation' (as quoted in [1]).

From the *Remarks*, it seems that Wittgenstein believes that Gödel showed that there is a 'true but unprovable statement in the system of Principia Mathematica' (cf. App. I §8-19 [11]). Further, he tries to show that if the interpretation of the Gödelian sentence is that it is a true but unprovable statement, then if one finds a proof of it, the interpretation must be relinquished. In his notion of true, true in a system is the same as provable in that system. So, the only way he says that one can talk of true but unprovable might be that it is true in another system. He says "why should not proposition of physics–for e.g., be written in Russell's symbolism?" If someone proved a statement '*P*' (in his notation) that is to be interpreted as '*P* is true but unprovable in PM'. This naturally is the only way such a statement can be interpreted as. For if it were false, then it would be provable and then, (because it is proved) true. If it was proved, then it would not be unprovable. Hence, it can only be true and unprovable. But then again, he asks 'true in what sense?'. This is the plan to go beyond or rather, to 'bypass' Gödel. If it is true in Russell's sense then it must be provable and hence the interpretation of '*P* is not provable' has to be given up. If one proves the unprovability of *P*, then by this proof one must have proved *P*, he says. The only thing that this tells one is that *P* is not a part of PM at all. The existence of 'profitless performances' like the liar's paradox does not make language less usable.

At least on reading the comments that Wittgenstein makes alone, it seems easy to see why the *Remarks* came to be so heavily criticized. Wittgenstein seems to misunderstand the method and the statements of Gödel. The theorems are syntactical and really tell one the undecidability of arithmetic given its consistency. It seems plausible that Wittgenstein would reject the Gödelian statement from being a mathematical proposition in the first place; that this is just a statement whose proof, if and when discovered, would have to be part of some other mathematical system, a new calculus altogether. This is a very unpropitious position to take since it can be shown that all Gödelian sentences are provably equivalent, and mathematicians have found lots of these statements that are not just pathological but a part of mathematical problems of copious amounts of interest. This is still not uncharacteristic of Wittgenstein considering that he took similar positions in relation to "undecided" propositions like the Goldbach Conjecture and the then-unproven Fermat's Last Theorem (cf. §189 [14]).

Floyd and Putnam [15] have argued in favor of Wittgenstein's paragraph. Firstly, they claim that

Wittgenstein's claim about the untranslatability of *P* as '*P* is true but unprovable in PM' is correct.

A system ω-inconsistent if, for some property F,it is provable that F(n) holds for every standard natural number n, but itis also provable that there is some natural number n such that P(n) fails. Since the provability predicate has the form,"There exists a p such that $\mathbf{P}_T(x,p)$", if *P* is proven, then PM has to be ω-inconsistent (setting aside inconsistency in general for now). Hence, the variable pmust be a nonstandard natural number; therefore, the translation of *P* as '*P* is true but unprovable in PM' is misguided.

They interpret Wittgenstein as saying that just because a proposition is unprovable doesn't mean it is not true in some *other* sense. They say that Wittgenstein is actually battling against the logicism of Frege and Russell, and that he actually is rejecting the idea that an ideal language can provide one a standard of truth. PM or the *Begriffschrift* don't provide an ideal language and foundation for mathematics, but are formal systems themselves. They claim that Wittgenstein was instead attacking the idea that Gödel's results show that there is a standard metaphysical truth about arithmetic.

Others have also commented on these paragraphs. Priest [16] argues that these remarks of Wittgenstein should not be understood in terms of the modern model-theoretic account of truth (indeed Wittgenstein himself gives a redundancy theory of truth in the same section) and that it is improper to say that he misunderstood Gödel. Priest remains agnostic on whether Wittgenstein understood Gödel but says that one must read the text in light of his idiosyncratic identification of truth and provability, and senselessness of mathematics, even claiming that a hint of acceptance of contradictions lies dormant in these remarks. Steiner [17] and Rodych [18] have argued that these accounts of the paragraph are too sympathetic and that textual evidence, especially within the other areas of his oeuvre makes it likely that Wittgenstein's remarks are mainly mistaken.

Whatever the immediate reactions in the communities of logic, philosophy and mathematics may be, the Incompleteness Theorems have strongly held their grip on the imagination of these communities for years to come and have had far-reaching results across the widest domains of human knowledge.

Discussion

Logic and Mathematics

The historical context of Gödel's Incompleteness Theorem is before model theory and almost only completely reliant on the techniques of proof theory. Tarski's work in model theory and beyond shows some very interesting and closely related ideas concerning formal systems. Of these closely related ideas, the closest one—so close that Gödel himself arrived at it independently while proving the Incompleteness theorems—is now known as "Tarski's Undefinability Theorem".

Theorem (Tarksi's Undefinability). *The set of Gödel numbers of the sentences of a consistent extension of* ***Q****, namely arithmetic, that are true under it is not arithmetically definable.*

This means that "truth" cannot be defined inside the system. Truth is to be understood as satisfaction of the formula under the standard model. Further, since all recursive sets are definable in arithmetic, the set of the Gödel numbers of true sentences of arithmetic is not recursive. If one assumes Church's thesis, then there also doesn't exist any algorithm such that one can decide whether any sentence of arithmetic is true or not.

Due to unfortunate notation, two notions of "undecidability" exist. A sentence σ is said to be undecidable in theory **T**, if neither it nor its negation are derivable in **T**. To avoid confusion with the other notion, authors shall use–as has become custom– "independent" instead. A set is called decidable if there exists an algorithm (which stops in finite time) which can decide whether an input is part of the set or not. They are also called recursive sets. A theory is said to be decidable if the set of theorems of that theory is decidable. Authors shall explore these notions more in the third section.

Another closely related theorem is Church's Undecidability Theorem, which states that *the set of valid sentences of arithmetic is not decidable.*

Gödel's First Incompleteness Theorem, Tarski's Undefinability Theorem and the Church's Undecidability Theorem are all heavily reliant on the techniques of the arithmetization of syntax and the diagonal lemma. This is a novel technique whose use by Gödel showed one the limitations of formal systems which can represent basic arithmetic. This is often considered to be real legacy of the Incompleteness Theorems. However, there is even more that is true. Gödel's Incompleteness Theorem's were the first theorems which established the existence of statements that cannot be proven and at the same time whose negations can also not be proven.

In the year 1900, David Hilbert gave one of the most influential lectures at the International Congress of Mathematics. Hilbert proposed twenty-three problems that would become a guide map for mathematicians well into the rest of the century [19]. Some of these problems are still unsolved. Running with the establishment of rigorous argumentation in the work of Weierstrass, Cauchy and Minkowski, Hilbert explains how a satisfactory solution to these problems would need one "to investigate the principles underlying these ideas and so to establish them upon a simple and complete system of axioms". Further, as one will see in later sections, Hilbert also emphasized on solutions being based on a finite number of steps on finite hypotheses–a deduction to the exact formulation of the problem. Gödel's theorems have a direct bearing on the second of these problems.

The second problem proposed by Hilbert was that of the proof of the compatibility of the axioms of arithmetic. To establish an exact and complete description of arithmetic, one must show that no finite deductions from the axioms of arithmetic lead to a contradiction, in other words, that it is consistent.

Hilbert explains how one may construct a suitable field of numbers for any geometrical system and then embed the geometry into the field and reduce the consistency of geometrical theory into that of the arithmetic. This meant that proving the consistency of arithmetic is a priority; most other things depend on arithmetic. If a proof of the consistency of arithmetic is given on Hilbert's conditions, then a lot of higher mathematics which is based on arithmetic, can also be shown to be consistent (analogous to the example of geometry). The problem also is contextually significant for Hilbert's philosophy of mathematics.

One must note a few things before one proceeds. Firstly, this notion of "finite methods" is quite elusive. Secondly, developments of model theory had not yet taken place. If one understands the problem to be a proof of the consistency of arithmetic using methods which are formalizable inside arithmetic, say in Peano Arithmetic, then this is proven to be impossible by the Second Incompleteness Theorem. However, this does not mean that some acceptable proof of consistency cannot exist. Gentzen surprisingly provided one with a consistency proof of Peano arithmetic in 1936 [20].

The proof was not formalizable within Peano arithmetic, meaning that this is in no way a violation of Gödel's theorems. Primitive Recursive Arithmetic is simpler than PA and it is likely that Hilbert would have been satisfied with a proof of PA's consistency through PRA. However, it is unclear whether the method satisfies Hilbert's criteria. Further, one can even show the consistency of PA within something stronger like ZFC set theory. Still, the Second Incompleteness Theorem applies to ZFC, making it impossible to prove the consistency of ZFC itself inside ZFC. Hence, it remains open whether Hilbert's second problem was solved positively by Gentzen or negatively by Gödel's Incompleteness.

However, if one accepts it as a proper proof method, then one need not worry about the consistency of the arithmetic. One needs not be thrown out of seats into a confused delirium about the consistency of arithmetic; it is safe and secure. Where does this leave one? First, one is left to reconsider the Second Incompleteness Theorem more precisely. The only thing that the Second Incompleteness states concerning the consistency of arithmetic is that one cannot find a proof of a statement expressing the consistency of the first-order theory *inside* the language and methods of the first-order theory–by the theory itself; this is a crucial nuance and can easily be misrepresented as saying that no proof can exist in a theory "weaker" than PA. Strictly, Gentzen's

system does not contain PA and isn't "stronger" than PA.

Another one of Hilbert's problems has been impacted by the Incompleteness Theorems, although more indirectly. This is the tenth problem—*To find a finite algorithm by which one can determine whether a Diophantine equation is solvable in the integers*. A Diophantine equation is a polynomial equation of finite unknowns and integer coefficients. The tenth problem was proven to be impossible through the work of Yuri Matiyasevich, Hilary Putnam, Julia Robinson and Martin Davis [21]; the last of whom used Gödel's method of encoding statements about a system into the system. Later, Maityasevich found a way to show that recursively enumerable sets can be represented by Diophantine equations. From incompleteness one knows that the set of provably true statements is recursively enumerable by definition, but not recursive since it is incomplete. Hence, there are recursively enumerable sets that are not recursive. From this, it follows that there is no such algorithm in general.

Other examples of independent statements of formal systems have popped up across the years. The Paris-Harrington theorem, Goodstein theorem, Kruskal's theorem and many more have been proven to be of this sort. Gödel's legacy in showing the first statement independent of its formal theory and the limitations of formal systems in general remains unshakeable.

Philosophy of mathematics & science

After logic and mathematics themselves, it is the philosophy of mathematics which was of most immediate relevance to Gödel's work.

At the time, the debate over the foundations of mathematics and its nature had given rise to three distinct schools of thought: Formalism, logicism and intuitionism. The impact of Gödel's theorem on each of them will be discussed here.

It is already seen in the previous section the consequences Gödel's theorem had on Hilbert's second problem. But in fact, the reach of the former went beyond just this: It struck a fatal blow to what is now known as *Hilbert's program*, and thus had major implications upon formalism itself.

To understand formalism, one must first understand what made formalism *necessary*. In standard mathematics, one seems to be forced to presuppose the metaphysical existence of an infinite jungle of objects: Numbers, relations, sets, and so on. The desire to make precise what ontological commitments mathematics required gave birth to formalism.

And so, the formalist goes to the other extreme, and claims that mathematical utterances are metaphysically *meaningless*. To put it succinctly, "mathematics is not a body of propositions representing an abstract sector of reality, but is much more akin to a game, bringing with it no more commitment to an ontology of objects or properties than ludo or chess." [22]

This kind of 'game formalism', however, was subjected to a series of devastating attacks (the details of which need not concern one now) by Frege. It is interesting to note that Gödel himself had objections to interpreting mathematics as pure syntax on other grounds as well: He believed that any purported reduction to pure syntax will end up with the concepts involved themselves being presupposed implicitly in the syntax, unless one restricted oneself to absurdly elementary systems. [23]

And so now is when David Hilbert, with his aforementioned program, joins the fore. He soon became the leading figure of formalism. To begin with, Hilbert divided mathematics into two domains: the finitary and the infinitary. Drawing the boundary between the two in a rough manner, Hilbert described the former as those objects which are "intuitively present as (irreducible) immediate experience prior to all thought." This would include the finite integers (and perhaps the rationals) but exclude the irrationals. Hilbert is considered to be a realist with respect to this sector. As for the rest, they were but meaningless symbols in the formalist sense. [24]

Quite apart from the problem of making rigorous the line between the two domains, the rise and downfall of Hilbert's program can both be traced down to this one thing: Hilbert's certainty that finitary means are consistent, and that the whole of the infinitary domain can be reduced down to them.

For the sake of clarity, authors enumerate some of the chief goals of Hilbert's program:

1. Finiteness: Any result concerning infinitary objects should be provable using a formalism of finitary objects exclusively.
2. Consistency: No contradiction should be obtainable in the formalism.
3. Completeness: All true mathematical statements should be provable in this formalism.

Of these, the second was the most important to validate Hilbert's philosophy: Contradictions *must* be unobtainable from finitary objects. And it is at this stage that Gödel enters the picture: His incompleteness theorems showed that, for these finitary objects, achieving completeness was impossible, and proving consistency (while staying within the formalism) was impossible—effectively shattering Hilbert's goal.

So where does the school of formalism itself stand now?

Most attempts to keep formalism alive take on after Hilbert's vision of it, typically by tweaking the allowance associated with the notion 'finitary', and then observing the strength and the properties of the resultant formalism. (Recall that Hilbert's hope had been that his notion of finitary would result in a formalism which would have been as strong as it gets.) These so-called relativized Hilbert programs have found great utility in proof theory and reverse mathematics. [25]

Similarly, formalism itself is also far from dead. One resurrection of it, distinct from both game formalism and Hilbert's formalism, does away with Hilbert's dichotomy and, committing ontologically only to metamathematics, holds mathematics to be a collection of formal systems. This "relativized" formalism, echoing the process above, reduces Gödel's second theorem to just another formal result.

In truth, however, accounts of direct attempts to retain formalism does not do its legacy justice. Hilbert remains relevant and revolutionary by the way in which he changed the manner in which axioms were treated in mathematics. They were no longer a set of self-evident statements; rather, they were any arbitrary set of statements which one may clump together according to their wishes in order to examine what they may give rise to. Since this experimental treatment of axioms also threw the consistency of the system into doubt, this shed the spotlight on yet another revolution in the philosophy of mathematics: That consistency is existence.

To conclude the section on formalism, *"... of the 'big three' foundational programs of the early 20th century, logicism and intuitionism retain supporters but are definitely special and minority positions, whereas formalism, its aims adjusted after the Gödelian catastrophe, has so infused subsequent mathematical practice that these aims and attitudes barely rate a mention. That must count as a form of success."*[26]

With this, authors now move on to logicism—which was, if anything, even more directly impacted by Gödel's theorems than formalism was.

After identifying a now-famous paradox in Frege's original conception of logicism, Russell's vision of refinement for the same is described in the preface to his 1903 'Principles of Mathematics': *"...all pure mathematics deals exclusively with concepts definable in terms of a very small number of fundamental logical concepts, and that all its propositions are deducible from a very small number of fundamental logical principles."* In this case, it becomes the logician's job to define logic in the necessary manner and prove the veracity of this claim. This was the goal striven to achieve with the project *Principia Mathematica.*

The punchline to *this* story is already well-known: Gödel proved the existence of formally undecidable statements in Principia Mathematica (as one knows it was this particular formal system which bore the original application of his theorems), and effectively ended the dream of a complete and consistent logical axiomatization for all mathematics.

But like Hilbert's program was for formalism, *Principia Mathematica* was but a figurehead representing the ideal of the logicist then; logicism itself has absorbed Gödel's attack and metamorphized, and *Principia Mathematica* itself remains important and relevant—although not

just exactly for the reasons its authors thought it would.

One may say that difference of opinion between formalism and Russell's logicism lies in the fact that the division between finitary and infinitary sectors was replaced with that of logic and non-logic (and the new metaphysics this replacement entailed)—the rest of the project remained the same (and thus failed in the same way). However, a crucial difference arises in the fact that, in spite of the aims of *PM* in particular, logicism itself is not inherently inconsistent with the fact that there exist undecidable statements in mathematics.

To begin with, instead of identifying formal derivability (in logic) with mathematical truth, the logicist may merely identify it with *knowable* mathematical truth (with a suitable notion of knowability). This is often referred to as the *weak* version of logicism, and essentially attempts to relegate independent statements to the realm of unknowable mathematical truth by making the required logical formal system an infinitary one [27]. One will soon see a similar theme in the different context of Paul Benacerraf's discussion of Gödel's theorem in the philosophy of mind.

Apart from this, one may even maintain *Russell's* logicism (which identifies formal derivability with *just* mathematical truth) by weakening the claim to just saying that all mathematical truths are derivable not in any *one* formal system, but in a set of them. This parallels the "relativized" formalism mentioned above.

However, there remain various other reasons to criticize logicism. One need not concern oneself with those here.

Intuitionism is, perhaps, the most enigmatic of the three doctrines being discussed. If Hilbert's essence is encapsulated by the maxim "Consistency is existence", then Brouwer's maxim is the following: *To exist is to be constructed.*

An ordinary garden-variety proof in classical mathematics consists of assuming the negation of *A*, obtaining a contradiction, and concluding *A*. It was seen, however, that assuming the law of the excluded middle and other such tools led to various metaphysical difficulties—which were what led to formalism and the like in the first place. Therefore, Brouwer rejected the validity of all such classical derivations and baptized the *'constructive proof'* to be the central notion in logic and mathematics, in place of 'truth'.

And so, intuitionistic logic rejects the law of the excluded middle; intuitionistic mathematics rejects the notion of 'actual' infinity and replace it with a constructible version: 'potential' infinity.

From these considerations stems the idea that mathematical objects are merely constructions of the human mind. Brouwer's intuitionism was an idea far more radical than formalism or logicism; indeed, many objected against it on the grounds that a philosophical idea should not purport to dictate so strongly what a mathematician should and shouldn't do.

Coming to the matter of its relationship with the incompleteness theorems: Brouwer was always of the opinion that mathematics was inexhaustible. The continuum is infinite and man's capacity to construct is finite; "one must always again draw afresh from the 'fountain of intuition'" [28]. And as a matter of fact, it seems to be that it was *Gödel* who was at least partly influenced by Brouwer's notions and his constructivism.

One has, from Carnap's diaries, in the years just preceding the publication of the incompleteness theorems [29]:

> [Gödel talked to me that day] about the inexhaustibility of mathematics. He was stimulated to this idea by Brouwer's Vienna lecture. Mathematics is not completely formalizable. He appears to be right.

Carnap writes down what Gödel told him:

> We admit as legitimate mathematics certain reflections on the grammar of a language that concerns the empirical. If one seeks to formalize such a mathematics, then with each formalization there are problems, which one can understand and express in ordinary language, but cannot express in the given formalized language. It follows (Brouwer) that mathematics is inexhaustible: one must always again draw afresh from the 'fountain of intuition'. There is, therefore, no *characteristica universalis* for the whole

> mathematics, and no decision procedure for the whole mathematics. In each and every closed language there are only countably many expressions. The continuum appears only in 'the whole of mathematics' . . . If we have only one language, and can only make 'elucidations' about it, then these elucidations are inexhaustible, they always require some new intuition again.

Therefore, it should come as no surprise that intuitionism is the only one of the three projects which was almost *emboldened* by Gödel's theorems; and while it has not dethroned classical mathematics, it remains the one generating the most active interest presently.

So far, authors have talked about the various philosophies of mathematics that have been impacted by the Incompleteness Theorems. However, an evergreen position regarding the ontology of mathematics is Platonism. The two theses that largely characterize mathematical Platonism are: (1) mathematical objects exist independently of one and the mathematical language and (2) mathematical objects are non-spatiotemporal and causally inert. These two are basic commitments and a lot of variety arises from differences in how much the philosopher sticks to these claims. Gödel himself was some sort of a strong mathematical Platonist.

Analyticity can be understood in such a way where all the axioms and theorems boil down to the law of identity, A is A. But arithmetic is essentially undecidable (incompleteness), and the set of analytic statements/theorems would be recursively enumerable. Gödel here rules out this analyticity in mathematics. Gödel instead uses a different notion of analyticity, where propositions are true by the 'nature of what the concepts that occur in it mean'. Gödel's realism about mathematics allows him to maintain that the concepts that make these analytic statements or the axioms of *Principia* true are objective and independent of one. He even says that one may assume axioms based on their success in applications (but still analytic axioms). Gödel says

> "... it is correct that a mathematical proposition says nothing about the physical or psychical reality existing in space and time, because it is true already owing to the meaning of the terms occurring in it, irrespectively of the world of real things. What is wrong, however, is that the meaning of the terms (that is, the concepts they denote) is asserted to be something man-made and consisting merely in semantical conventions. The truth, I believe, is that these concepts form an objective reality of their own, which we cannot create or change, but only perceive and describe." (p. 320 [30])

Putnam [31] argues similarly to say that one can also establish "syntheticity" in mathematics. He took Gödel's theorems in conjunction with the broadly naturalistic attitude brought about by (neo-)pragmatism to transform it into a Platonism and to build a realistic attitude to mathematics. This is strikingly anti-Gödelian in the sympathies for naturalism but is very close to the method of Gödelian Platonism. There is a notable tension between two things that Putnam builds his argument from. Firstly, that mathematics *works*. It is used all the time–the so-called unreasonable effectiveness of mathematics in natural science. Secondly, these formal systems used in mathematics (ZFC, PA etc.) cannot prove their own consistency finitistically as the second incompleteness theorem tells one.

A strictly analytic conception of mathematics seems *prima facie* untenable in the light of the incompleteness theorems. Putnam argues here that some sort of "synthetic" truths must exist in mathematics and quasi-empirical methods are the best chance at getting them. Calculus, to take an example, was quasi-empirical before its formalisation by Weierstrass. The formalization doesn't *justify* calculus, calculus was already justified by other means. Mathematics (and the need for it) is necessarily embedded in its applications. Putnam argues that a realistic interpretation, namely that mathematical statements are made true by something external to one, allows one to understand the consistency and the creative applications of mathematics in the best way. He even argues that one may learn things from 'mathematical experiments'. A shift of attitude in broad scale mathematics, with the advent of 'computer experiments/assisted-proofs' and brute-force verification, might fortify the

attractiveness of this position. Mathematics is fundamental to the world and Putnam argues that one might as well take it as a *part* of the world.

Automata theory & Information theory

In the first section of the discussion, authors saw two sister-theorems (so to speak) to Gödel's: Tarski's undefinability theorem and Church's undecidability theorem. In this section, authors shall first look at another fundamental result developed by Alan Turing: Namely, the unsolvability of the halting problem. Subsequently, authors shall briefly appraise an important development of this result, due to Gregory Chaitin.

The Turing machine is an entity scans a tape (upon each square of which is printed a letter from the alphabet) which extends infinitely to the left and right and, according to the 'instruction' delivered by a function, moves left or right (or stays) and changes its state. A Turing machine will keep computing until it reaches a state upon for which it has no instructions. At this point, one says that the machine has *halted*; and furthermore, that the machine *accepts* the string it was originally fed.

Now, Turing machines can also be viewed as devices used to compute functions. The initial input string is the argument for the function, and the expression on the tape when the machine halts (if it halts) is taken to be the value of the function at that argument. In fact, the kind of functions Turing machines compute were posited to capture a certain specific kind of function, which may be referred to as *algorithmic* or *effectively computable* or also *partial recursive* functions.

In fact, all these three classes of functions were posited independently in different ways by Church, Turing and Gödel respectively, in an attempt to capture the informal notion of an 'effectively calculable function'; that they do indeed coincide was proven by Church and Turing, and the statement that they do capture the aforementioned notion is known as the Church-Turing thesis [32].

Before moving on, one must observe that the set of functions from the set of naturals to itself is non-enumerable, and that the set of Turing machines is enumerable (from which it follows that each Turing machine can be given a natural number encoding). From these considerations alone, one can see that there must exist natural number functions which are not Turing computable.

The halting problem is the question of whether the following function is Turing computable: $h(m, n) \rightarrow \{1, 2\}$ such that $h(m, n) = 1$ if the Turing machine with the natural number encoding n halts on the input tape with the natural number encoding n, and $h(m, n) = 2$ if the Turing machine fails to halt on the input tape.

Theorem. *The halting function h is not Turing computable.*

Proof. One first constructs the following Turing machines:

1. The copying machine, C: Given a tape containing a block of n strokes, and otherwise blank, if the machine is started scanning the leftmost stroke on the tape, it will eventually halt with the tape containing two blocks of the n strokes separated by a blank, and otherwise blank, with the machine scanning the leftmost stroke on the tape.
2. The dithering machine, D: Started on the leftmost of a block of n strokes on an otherwise blank tape, D eventually halts if $n > 1$, but never halts if $n = 1$.

Now, suppose one had a Turing machine H which computed the function h. This means that one can easily create a combined machine C+H=G which first performs the processes of the machine C and then, upon halting, begins the processes of the machine H. This machine G computes the function $g(n) = h(n, n)$. One now further combines G with D to get a machine G+D=M, which first goes through the operations of G and then the operations of D.

Now, if the Turing machine numbered n halts when fed the tape corresponding to its own number, one would have $h(n, n) = g(n) = 1$, and consequently, the machine M would not halt when started on the number n, since one describes D as halting only if $n > 1$. On the other hand, if the Turing machine numbered n does *not* halt when fed the tape

corresponding to its own number, one would have $h(n, n) = g(n) = 2$, and consequently, the machine M *would* halt when started on the number n.

But what would happen if one fed the machine M its own number? In this case, it would halt on m only if m does not halt on m, that is, if it does not halt on m. This is a contradiction. One must conclude that h is not computable.

One can see once again the diagonalization original to Cantor, a certain flavour of *self-referential* argumentation inspired by Gödel, lying at the heart of the argument. As a historical aside: The proof of the negative answer to the halting problem was Turing's first step in proving the unsolvability of the *Entscheidungsproblem*, also known as Church's undecidability theorem, and which Turing proved independently of and parallel to Church [33].

The relationship between automata theory and Gödel's theorem has also been exploited by Putnam in an article[34], wherein he shows that if scientific epistemology—what Chomsky referred to as the "scientific competence"—could ever be represented by a Turing machine, then humans could never, given that level of scientific competence, know this fact (in the sense of having evidence which provides justification for it).

An extended application of the halting problem in the field of algorithmic information theory was first given by Gregory Chaitin in terms of what is now known as *Chaitin's constant* and *Chaitin's incompleteness theorem*. Chaitin's constant can be thought of as the probability that a random program of a fixed finite length will halt. Its value is highly machine-dependent.

Chaitin's constant is computably enumerable, but algorithmically random. This is to say that each halting probability is uncomputable by any algorithm, and in some cases, it has been proven that not even a single bit of Chaitin's constant is computable. Without going into too many details, Chaitin's incompleteness theorem states that there exists a position of the decimal expansion of Chaitin's constant beyond which the value of the numeral is undecidable.

Authors conclude this section by emphasizing the point that the halting problem and Chaitin's construction have been widely impactful. This is because many important problems in fields such as number theory amount to solving the halting problem for special programs, or alternatively, to knowing enough bits of Chaitin's constant.

Philosophy of mind & cognitive science

In 1951, the eponymous logician formulated the following:

Gödel's disjunction: *Either mathematics is incompletable in this sense, that its evident axioms can never be comprised in a finite rule, that is to say, the human mind infinitely surpasses the powers of any finite machine, or else there exist absolutely undecidable problems.*[23]

Of all the various extensions in application Gödel's theorem has found across disciplines, the temptation to extract conclusions regarding the nature of consciousness from it is undoubtedly the most seductive one of them all. It is with good reason that J.R. Lucas, in one of the most well-known articles regarding this matter, says immediately after declaring in the opening lines his belief that Gödel's theorem disproves mechanism: "Almost every mathematical logician I have put the matter to has confessed to similar thoughts, but has felt reluctant to commit himself definitely until he could see the whole argument set out, with all objections fully stated and properly met. This I attempt to do." [35]

Gödel's own thought process taking him to the aforementioned disjunction was clear enough. One starts off with the following tautology: *Either the human mind is not a Turing machine, or it is a Turing machine.* If the latter is true, then, since it is known that

1. Every Turing machine capable of performing arithmetic is incomplete
2. The human mind is capable of performing arithmetic,

it follows that the human mind is *essentially* incomplete with respect to mathematics, that is, there exist some questions in it which are *absolutely* undecidable, whose answers one cannot know even in principle. And so, mathematics is incompletable.

If the former is true, then there *may* not exist any 'absolutely' undecidable problems in mathematics. However, since this means that the

human mind cannot be captured by any finite Turing machine, it follows that neither can the axioms of mathematics—for Gödel held that mathematical axioms were just those sentences which were necessarily evident by virtue of basic mathematical intuition in humans. Since the axioms of mathematics are then uncapturable by a finite rule, one once again arrives at the conclusion that mathematics is incompletable—albeit in a different sense than the first.

In the same piece, Gödel went on to remark that it seems to be the first alternative which is in good agreement with leading figures in physiology (and authors will soon look into explicit proponents of this alternative and its naysayers); and so perhaps Gödel was, in general, inclined to believe that the human mind surpassed the Turing machine. However, he adds that it is the philosophical conclusions of the *second* alternative (which was, for him, Platonism) which seem to gel better with modern developments in the foundations of mathematics.

Now, it is interesting to note here that while this discussion is on Gödel's Gibbs lecture which was delivered in the year 1951, the notes pertaining to the same were published only posthumously, and by the time they came out, Nagel & Newman had already put forward an informal anti-mechanist suggestion based on Gödel's theorem in their 1958 exposition titled 'Gödel's proof' [36], Lucas had already published 'Minds, Machines and Gödel' in 1961, and people had already begun attacking Lucas' argument.

Lucas' article itself was a wholehearted espousal of the anti-mechanist conclusion, that is, of the first leg of Gödel's disjunction. His idea was, at heart, quite a natural one: In spite of the fact that there exist formally undecidable propositions in mathematics, one is able to 'see' that these propositions are true—something which the formal system in question *cannot* do. It follows that the human mind transcends formal systems.

Perhaps the most interesting analysis/criticism of it was given by Paul Benacerraf in his 1967 article 'God, the Devil and Gödel' [37]. Benacerraf fleshed out better and made more rigorous the assumptions and steps in Lucas' argument, concluding that there *is*, inarguably, a contradiction in saying both that the human mind is a Turing machine and that humans can *know* the independent statements of a Turing machine to be true.

Benacerraf diverged from Lucas at this stage, and concluded instead that humans cannot know the truth of certain independent statements; that there was an in-principle limitation in the power of the human mind to perceive the truth or falsity of mathematical statements. In fact, what Benacerraf did was nothing other than *reveal* the other leg of Gödel's disjunction—and emphasis is placed on the use of the word 'reveal' here, for as it most interestingly turns out, Benacerraf's article also came out *before* the publication of Gödel's Gibbs lecture and his other manuscripts.

The last major player in this milieu whom authors will discuss is Sir Roger Penrose, whose advocation of anti-mechanism (indeed, it is now referred to as the Penrose-Lucas argument) is, perhaps, closest in emotion to Gödel himself (Penrose himself is also a self-declared Platonist). For while Penrose recognizes the 'choice' offered to one by the disjunction, he insists that to say that there is an algorithm to decide mathematical truth which is too complicated to ever be known to one is a contradiction in itself. As he says in *The Emperor's New Mind*: "But this flies in the face of what mathematics is all about! The whole point of our mathematical heritage and training is that we do *not* bow down to the authority of some obscure rules that we can never hope to understand." [38]

Needless to say, however, such an argument is far from what may be required to settle the matter once and for all, and a standing consensus among various experts in the relevant fields is that it ultimately fails (for example, Putnam, among others, has contended that there is no reason to believe that the human mind can ultimately always prove the consistency of an arbitrarily complex Turing machine, which is the conditional which must be satisfied to apply Gödel's theorem and find the independent statement). [39]

As it so happens, one of these critics was the American scholar Douglas Hofstadter—who had some other ideas of his own. A decidedly less bold attempt to use Gödel's theorem in making advances in the understanding of consciousness

owes itself primarily to him, one of the biggest names in thefield of cognitive science.

In his book *Gödel, Escher, Bach: An Eternal Golden Braid*, Hofstadter comments [40]:

> If one uses Gödel's theorem as a metaphor, as a source of inspiration, rather than trying to translate it literally into the language of psychology or of any other discipline, then perhaps it can suggest new truths in psychology or other areas. But it is quite unjustifiable to translate it directly into a statement of another discipline and take that as equally valid. It would be a large mistake to think that what has been worked out with the utmost delicacy in mathematical logic should hold without modification in a completely different area.

Having given his readers a word of caution, Hofstadter now goes on to say:

> I think it can have suggestive value to translate Gödel's theorem into other domains, provided one specifies in advance that the translations are metaphorical and are not intended to be taken literally. That having been said, I see two major ways of using analogies to connect Gödel's theorem and human thoughts…

While Penrose and Lucas were more focused on the *result* of the incompleteness theorems and the consequences that they entailed, Hofstadter's interest lay in the *process* by which said results were obtained.

More precisely, it was the idiosyncratic element of *self-reference* in Gödel's proof which so captured Hofstadter's attention. As a matter of fact, Hofstadter sees such self-referential phenomena *everywhere*, christening this general abstract structure with the phrase '*Strange loop*', and believes it to be the essential mechanism out of which the psychological self emerges. For Hofstadter, the real takeaway from Gödel was the idea that any 'system', if it surpasses a certain level of complexity, has the ability to refer to itself.

The way the system had two 'levels' of description (for example, authors had instantiating the provability predicate both P_T and $\mathbf{P}_\mathrm{T}$) is not something unique to it: One deals with multiple levels of description and representation all the time.

Both these levels form a hierarchy in the sense that the P_T level was, logically speaking, constructed after the—which is to say, based on the—$\mathbf{P}_\mathrm{T}$ one via the Gödel numbering and arithmetization.

But the crucial property of this hierarchy would be its *tangledness*: Not only is the P_T level determined by the $\mathbf{P}_\mathrm{T}$ level, but also, the P_T level, in its own way, determines the $\mathbf{P}_\mathrm{T}$ level; for after arithmetization, it was in the arithmetically-encoded description that one constructed P_T, before importing it back to the $\mathbf{P}_\mathrm{T}$ level and considering the implications of this crossing-over.

Hofstadter also refers to this as a 'tangled hierarchy'. One of his favorite illustrations to offer as an example is the lithograph 'Drawing Hands' by Dutch artist M.C. Escher.

In terms of human cognition, Hofstadter's ultimate thesis is that the various phenomena emergent from it—ideas, feelings, analogies, and finally, self-consciousness—are all based on a strange loop, a Gödelian level-crossing, a tangled hierarchy involving the neurons of the brain and the symbols of the mind.

Acknowledgement

The authors would like to thank their mentor Prof. Mihir Chakraborty, whose invaluable guidance made this article possible. They would also like to thank the anonymous reviewers for their helpful comments in improving the article.

Bibliography

1. J W Dawson Jr, The Reception of Gödel's Incompleteness Theorems, in PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Philosophy of Science Association, Vol 2, page 253–271, 1984.
2. S C Kleene, N De Bruijn, J de Groot, and A C Zaanen, Introduction to Metamathematics, van Nostrand, New York, Vol 483, 1952.
3. S C Kleene, The work of Kurt Gödel, Journal of Symbolic Logic, Vol 41, No 4, page 761–778, 1976.
4. K Gödel, Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatshefte fúr mathematik und physik, Vol 38, No 1, page 173–198, 1931.
5. J W Dawson Jr, Discussion on the Foundation of Mathematics, History and Philosophy of Logic, Vol 5, No 1, page 111–129, 1984.

6. I Grattan-Guinness, In memoriam Kurt Gödel: His 1931 correspondence with zermelo on his incompletability theorem, Historia mathematica, Vol 6, No 3, page 294–304, 1979.
7. M Davis, The undecidable: Basic papers on undecidable propositions, unsolvable problems and computable functions, Courier Corporation, 2004.
8. P Finsler, Formale beweise und die entscheidbarkeit, Mathematische Zeitschrift, Vol 25, No 1, page 676–682, 1926.
9. D Hilbert, Die Grundlagen der Mathematik II, Springer, 1939.
10. P Mancosu, Between Vienna and Berlin: The Immediate Reception of Gödel's Incompleteness Theorems, History and Philosophy of Logic, Vol 20, No 1, page 33–45, 1999.
11. L Wittgenstein, Remarks on the Foundations of Mathematics, 1956.
12. L Wittgenstein, Tractatus logico-philosophicus (trans Pears and McGuinness), 1961.
13. L Wittgenstein, Philosophical Grammar, University of California Press, 2005.
14. L Wittgenstein, Philosophical Remarks, University of Chicago Press, 1980.
15. J Floyd, H Putnam, A Note on Wittgenstein's Notorious Paragraph about the Gödel Theorem, The Journal of Philosophy, Vol 97, No 11, page 624–632, 2000.
16. G Priest, M Kolbeland, B Weiss (eds), Wittgenstein's Remarks on Gödel's Theorem, in Wittgenstein's Lasting Significance, Routledge, Taylor Francis, London, chapter 8, page 207–227, 2004.
17. M Steiner, Wittgenstein as his Own Worst Enemy: The Case of Gödel's Theorem, Philosophia Mathematica, Vol 9, No 3, page 257–279, 2001.
18. V Rodych, Misunderstanding Gödel: New Arguments about Wittgenstein and New Remarks by Wittgenstein, Dialectica, Vol 57, No 3, page 279–313, 2003.
19. D Hilbert, Mathematical Problems, Bull Amer Math Soc, Vol 8, page 437–479, 1901-1902.
20. G Gentzen, Die widerspruchsfreiheit der reinen zahlentheorie, Mathematische annalen, Vol 112, No 1, page 493–565, 1936.
21. M Davis, Hilbert's tenth problem is unsolvable, The American Mathematical Monthly, Vol 80, No 3, page 233–269, 1973.
22. A Weir, Formalism in the Philosophy of Mathematics, in The Stanford Encyclopedia of Philosophy, E N Zalta (ed), Metaphysics Research Lab, Stanford University, 2020.
23. F A Rodriguez-Consuegra, Some basic theorems on the foundations of mathematics and their philosophical implications, in Kurt Gödel: Unpublished Philosophical Essays, F A Rodriguez-Consuegra (ed), Basel: Birkhäuser Basel, page 129–170, 1995.
24. M Detlefsen, Hilbert's formalism, Revue Internationale de Philosophie, Vol 47, No 186, page 285–304, 1993.
25. L Horsten, Philosophy of Mathematics, in The Stanford Encyclopedia of Philosophy, E N Zalta (ed), Metaphysics Research Lab, Stanford University, 2019.
26. P Simons, Formalism, in Philosophy of Mathematics, A D Irvine (ed), Amsterdam, North Holland, page 291–310, 2009.
27. G Hellman, How to Gödel a Frege-Russell: Gödel's Incompleteness Theorems and Logicism, Nous, Vol 15, No 4, page 451–468, 1981.
28. M van Atten, Essays on Gödel's Reception of Leibniz, Husserl, and Brouwer, ser. Logic, Epistemology, and the Unity of Science, Springer International Publishing, 2014.
29. R Tieszen, Dirk van Dalen. Mystic, Geometer, and Intuitionist: The life of L E J Brouwer. Vol 2: Hope and Disillusion, Oxford: Clarendon Press, 2005. pp x+ 441–946. ISBN 0-19-851620-7 (hardcover), Philosophia Mathematica, Vol 15, No 1, page 111–116, 2007.
30. K Gödel, S Feferman, Kurt Gödel: Collected Works: Volume III: Unpublished Essays and Lectures, Oxford University Press on Demand, Vol 3, 1986.
31. H Putnam, What is Mathematical Truth?, Historia Mathematica, Vol 2, No 4, page 529–533, 1975.
32. B J Copeland, The Church-Turing Thesis, in The Stanford Encyclopedia of Philosophy, E N Zalta (ed), Metaphysics Research Lab, Stanford University, 2020.
33. A M Turing, On Computable Numbers, with an Application to the Entscheidungsproblem, Proceedings of the London Mathematical Society, Vol s2-42, No 1, page 230–265, 1937.
34. H Putnam, After Gödel, Logic Journal of the IGPL, Vol 14, No 5, page 745–754, 2006.
35. J R Lucas, Minds, Machines and Gödel, Philosophy, Vol 36, No 137, page 112–127, 1961.
36. E Nagel, J Newman, Gödel's Proof, ser Routledge Classics, Routledge, 2005.
37. P Benacerraf, God, the Devil, and Gödel, Etica E Politica, University of Trieste, Department of Philosophy, Vol 5, No 1, page 1–15, 2003.
38. R Penrose, M Gardner, The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics, ser Popular Science Series, OUP Oxford, 1999.
39. H Putnam, Minds and machines, in Dimensions of Minds, S Hook (ed), New York University Press, New York, USA, page 138–164, 1960.
40. D Hofstadter, Gödel, Escher, Bach: An Eternal Golden Braid, ser Pulitzer Prize in letters: General non-fiction, Basic Books, 1999.