

11 What Can We Know

Rowan Brad Quni ORCID <https://orcid.org/0009-0002-4317-5604>

[[Philosophy of Science]]

Chapter 11: What Can We Know? Gödel, Computation, Paradoxes, and the Limits of Formal Systems

11.1 Introduction: Probing the Boundaries of Knowledge

Previous chapters have critically examined the foundations of modern physics, revealing deep challenges to our understanding of reality, substance, causality, structure, methodology, and even the process of scientific knowing itself. These challenges suggest limitations in our current scientific theories' ability to represent the world adequately. This chapter shifts the focus slightly, addressing a related but distinct question: Are there inherent limits not just to specific scientific theories, but to our very systems of knowledge, logic, and formal reasoning themselves? Can any sufficiently complex system achieve complete self-description or prove its own foundations? This inquiry delves into the fascinating territory of **Formal Systems**, **Self-Reference**, and the profound limitations revealed by mathematical logic, most notably through **Kurt Gödel's Incompleteness Theorems** and related paradoxes like **Russell's Paradox**. We will define what constitutes a formal system amenable to such analysis and explore the rigorous arguments demonstrating inherent boundaries to proof, consistency verification, and computation. This exploration argues that just as physics encounters limits suggesting its frameworks are inadequate or incomplete representations of reality, our most rigorous systems of thought—mathematics and logic formalized—also face fundamental, demonstrable limits when they attempt comprehensive self-understanding or complete axiomatization. Understanding these limits is crucial for a realistic epistemology and for appreciating the boundaries of certainty achievable through formal methods.

11.2 The Ideal of Formalization: Axioms, Rules, and Hilbert's Dream

The early 20th century witnessed a powerful drive towards formalization in mathematics and logic, fueled by the discovery of paradoxes in naive set theory (like Russell's paradox, discussed later) and a desire to place mathematics on absolutely secure foundations. The goal was to capture mathematical reasoning within **Formal Systems**—precisely defined structures free from ambiguity and reliance on intuition. A formal system, in the sense relevant here, consists of several key components. First, a **Formal Language** with a finite alphabet of symbols and explicit syntactic rules for constructing well-formed formulas (WFFs), the meaningful statements of the system. Second, a set of **Axioms**, specific WFFs taken as unproven starting points. Crucially, for the systems Gödel studied, this set of axioms must be **effectively axiomatized**, meaning there exists an algorithm or mechanical procedure that can decide in a finite time whether any given WFF is an axiom. Third, a set of **Rules of Inference**, which are purely syntactic rules (like Modus Ponens) specifying how new WFFs (theorems) can be derived from axioms or previously derived theorems. These rules must also be effective, allowing mechanical checking of their correct application. Finally, a **Formal Proof** is a finite sequence of WFFs where each step is either an axiom or follows from previous steps via an inference rule; a **Theorem** is the final statement in such a proof. The requirement of effective axiomatization and effective inference rules ensures that the validity of any purported proof can be checked algorithmically.

This ideal of reducing mathematics to manipulation within such formal systems reached its zenith in **Hilbert's Program**. David Hilbert sought to find a single, consistent, and complete formal axiomatic system encompassing all of mathematics. Furthermore, he aimed to prove the **consistency** of this system (i.e., that it could never derive a contradiction like P and not- P) using only **finitary methods**—basic, uncontroversial combinatorial reasoning about symbols considered absolutely secure. If successful, this program would provide an unshakeable foundation for mathematics, demonstrating its internal coherence and its capacity to resolve all mathematical questions through formal proof. It was precisely this ambitious program that Gödel's work would famously undermine.

11.3 The Limits Within: Gödel's Incompleteness Theorems

In his groundbreaking 1931 paper, Kurt Gödel demonstrated fundamental limitations inherent in any formal system powerful enough to capture basic arithmetic. His two incompleteness theorems revealed an unavoidable gap between what can be proven within such systems and what is true (at least, true according to the standard interpretation).

Gödel's First Incompleteness Theorem (G1T) states that any consistent, effectively axiomatized formal system F that is sufficiently strong to express elementary arithmetic is necessarily **incomplete**. This means there exists at least one sentence, G , formulable in the language of F , such that F can neither prove G nor disprove G (i.e., prove not- G). Gödel achieved this by ingeniously showing how such a system could talk about itself. Using **Gödel Numbering**, he assigned unique numbers to symbols, formulas, and proofs, allowing syntactic properties to be represented arithmetically. He then used a **Diagonalization** technique (similar to the Liar Paradox) to construct a specific arithmetic sentence, the **Gödel Sentence (G)**, which effectively encodes the statement “This sentence is not provable in system F ”. Assuming F is consistent, Gödel argued that G must be unprovable within F (if it were provable, F would be inconsistent) and its negation, not- G , must also be unprovable (requiring a slightly stronger condition, ω -consistency, in Gödel's original proof, later removed by Rosser's refinement using a different sentence). Thus, G is formally undecidable within F .

This theorem dramatically highlights the distinction between **syntactic provability** (derivability from axioms via rules within F) and **semantic truth** (truth in an intended interpretation, like the standard model of natural numbers). Since G asserts its own unprovability, and we have meta-theoretically reasoned that G is indeed unprovable in F (assuming F is consistent), G appears to be *true* in the standard interpretation of arithmetic. This “true but unprovable” aspect underscores that no single formal system meeting the criteria can capture all arithmetic truths. It also reveals the **inexhaustibility of mathematics**: we can always strengthen F by adding G (or its negation) as a new axiom, creating a new system F' , but F' will itself be incomplete, having its own undecidable Gödel sentence G' .

Gödel's Second Incompleteness Theorem (G2T) strikes even more directly at Hilbert's program. It states that for any consistent, effectively axiomatized formal system F capable of formalizing basic proof theory (like Peano Arithmetic or ZFC set theory), the system F cannot prove its own consistency. The statement “ F is consistent,” denoted **Con(F)**, can itself be encoded as an arithmetic sentence within F . Gödel showed that F proves the implication “If F is consistent, then G is true” (formally, $F \vdash \text{Con}(F) \rightarrow G$). Since G is unprovable in F (by G1T), it follows that **Con(F)** must also be unprovable in F . This means that sufficiently powerful formal systems lack the internal resources to rigorously demonstrate their own freedom from contradiction. Establishing the consistency of such systems requires stepping outside them and using potentially stronger meta-theoretic assumptions or methods (like Gentzen's proof of PA's consistency using transfinite induction). This demolished Hilbert's hope for a self-contained, finitary proof of consistency for all of mathematics.

It is crucial to emphasize the **conditions for applicability**: the theorems apply only to formal systems that are **consistent**, **effectively axiomatized**, and possess **sufficient arithmetic power** (at least Robinson Arithmetic Q for G1T, often Primitive Recursive Arithmetic PRA for G2T). Systems lacking any of these properties (e.g., pure propositional logic, inconsistent systems, systems not algorithmically specifiable, or “True Arithmetic” which isn’t effectively axiomatized) fall outside the theorems’ direct scope.

11.4 Echoes of Paradox: Self-Reference and Foundational Crises

Gödel’s proof technique, relying on self-reference achieved through Gödel numbering and diagonalization, resonates deeply with classic logical and semantic paradoxes that also reveal limits in formal systems or language.

The **Liar Paradox** (“This statement is false”) exemplifies the dangers of semantic self-reference. If the statement is true, then what it says must be the case, meaning it must be false. If it is false, then what it says is not the case, meaning it is not false, hence true. This creates an inescapable contradiction, demonstrating that naive assumptions about truth and reference can lead to inconsistency within natural language. Tarski’s work on the undefinability of truth showed that a formal language cannot contain its own truth predicate without risking paradox, necessitating a hierarchy of languages (object language, metalanguage, etc.). Gödel cleverly bypassed direct semantic paradox by focusing on provability (“This statement is unprovable *in system F*”), a syntactic concept definable *within* the arithmetic system, leading to incompleteness rather than inconsistency.

Russell’s Paradox exposed a fundamental contradiction in naive set theory, which implicitly assumed an unrestricted **Comprehension Principle**: for any property, there exists a set of all things having that property. Bertrand Russell considered the property “is a set that does not contain itself as a member.” He asked: Does the set R of all such sets contain itself? If R contains itself, then by definition it must *not* contain itself. If R does not contain itself, then it satisfies the property defining membership in R, so it *must* contain itself. This contradiction demonstrated the inconsistency of naive set theory. The resolution came through developing **axiomatic set theories**, like Zermelo-Fraenkel set theory (ZFC), which replace unrestricted comprehension with more restricted axioms (like Specification and Replacement) designed to block the formation of such paradoxical “sets.” Russell’s paradox thus revealed limits on naive collection-forming principles and motivated the move towards rigorous axiomatization as a foundation for mathematics.

These paradoxes, like Gödel’s theorems, arise from systems (whether linguistic, conceptual, or formal) attempting a certain kind of closure or self-application—defining truth within the language, forming sets based on any property, proving consistency within the system. They collectively suggest that such complete self-containment or unrestricted self-reference often leads to contradiction or incompleteness, indicating inherent limitations in formal systems attempting full self-description or justification.

11.5 Computational Limits: The Halting Problem

The limitations revealed by Gödel find a direct parallel and deep connection in the theory of computation, particularly through **Alan Turing’s Halting Problem**. Turing formalized the intuitive notion of computation with the **Turing machine**, an abstract model of a general-purpose computer. He then asked whether there exists an algorithm (a Turing machine) that can determine, for any arbitrary program (another Turing machine M) and any input I, whether program M will eventually halt (stop) when run on input I, or run forever.

Turing rigorously proved that **no such general algorithm exists**; the Halting Problem is **undecidable**. His proof, like Gödel’s, used a diagonalization argument. He imagined a hypothetical halting-solver machine H and constructed a paradoxical machine P that takes another machine’s description as input, uses H to see if that machine

would halt when given its own description as input, and then deliberately does the opposite (loops if it would halt, halts if it would loop). Feeding P its own description leads to a contradiction, proving H cannot exist.

The undecidability of the Halting Problem is deeply equivalent to Gödel's First Incompleteness Theorem. One can derive G1T from the Halting Problem: if a complete formal system F for arithmetic existed, it could be used to solve the Halting Problem (by proving whether "M halts on I" or "M does not halt on I"), contradicting Turing's result. This connection underscores that Gödel's theorems are fundamentally about the limits of **algorithmic processes** and **formal computability**. They demonstrate that there are well-defined mathematical questions (like the truth of G or the halting of certain programs) that cannot be answered by any fixed, rule-based, computational procedure.

11.6 Philosophical Implications and Misinterpretations: Mind, Physics, and Beyond

While Gödel's theorems have precise mathematical content, their implications have fueled widespread philosophical discussion, speculation, and frequent misinterpretation.

Valid philosophical implications include the recognition of the **limits of formalism** (no single system captures all mathematical truth), the **inexhaustibility of mathematics**, and the fundamental distinction between **truth and provability-within-a-system**. They challenge purely formalist philosophies of mathematics and suggest that mathematical understanding may involve elements beyond axiomatic deduction.

However, applications beyond mathematics and computation are often problematic. The **Lucas-Penrose argument**, claiming the theorems prove human minds are non-algorithmic because humans can "see" the truth of Gödel sentences that machines cannot prove, is widely considered flawed. It relies on questionable assumptions about human consistency, conflates informal insight with formal proof, and ignores the complexity and potential unknowability of any formal system that might model the mind.

Similarly, invoking Gödel's theorems to argue for the **impossibility of a physical "Theory of Everything" (ToE)** is speculative and controversial. Physics is empirical, not purely formal; physical theories are not typically presented as fully formalized axiomatic systems meeting Gödel's criteria; a ToE might not require the full power of arithmetic; and undecidable statements might lack physical relevance. While the theorems might suggest limits on *predicting* the ultimate behavior of sufficiently complex physical systems if they are computational, a direct preclusion of a workable ToE is not a rigorous consequence.

More egregious **misapplications** occur when the theorems are invoked in domains like theology, ethics, law, or social theory. These domains typically lack the precise formal language, effective axiomatization, and arithmetic content necessary for the theorems to apply logically. Equating any perceived limitation, paradox, or incompleteness in these informal systems with Gödelian incompleteness is an unwarranted analogy that ignores the theorems' specific technical requirements. It is crucial to distinguish between the theorems' rigorous demonstration of limits within specific formal systems and their looser, metaphorical resonance with broader themes of self-reference and systemic limitation.

11.7 Synthesis: Knowing the Boundaries of Formal Knowledge

Gödel's Incompleteness Theorems, alongside related paradoxes (Liar, Russell's) and computational limits (Halting Problem), establish profound and inherent limitations within **formal systems** of sufficient complexity. Any consistent system powerful enough to formalize basic arithmetic (and thus computation and its own syntax) cannot prove all true statements expressible within it, nor can it internally guarantee its own consistency. This demonstrates fundamental boundaries to what can be achieved through purely formal, axiomatic, algorithmic methods.

These results have deep epistemological significance. They refute the dream of achieving absolute certainty and completeness through purely formal methods, at least for domains encompassing basic arithmetic. They reveal an inherent limitation in the ability of complex systems to achieve full self-description or self-validation through their own internal rules.

Relating this back to the broader themes of this work, these formal limitations resonate with the **representational inadequacies** identified in fundamental physical theories. Just as QM and GR seem unable to provide a complete and coherent picture of physical reality, our most rigorous systems of mathematical and logical thought are demonstrably unable to capture all truths within their scope or guarantee their own foundations internally. This suggests a potential parallel: perhaps the limits encountered in physics are not merely due to missing data or incomplete theories within a fundamentally knowable reality, but might reflect deeper, Gödel-like limitations inherent in any sufficiently complex system (whether physical, cognitive, or formal) attempting to represent or encompass itself.

Understanding these formal limits is crucial for a mature philosophy of science and a realistic epistemology. It cautions against unrealistic expectations of finality, absolute certainty, or complete formalization in our scientific endeavors. It highlights the potential gap between our formal descriptions (theories, models) and the reality they attempt to capture, suggesting that truth may always transcend proof within any given system. It underscores the importance of meta-theoretic reasoning, empirical grounding (in physics), and potentially non-formalizable intuition or insight in advancing knowledge beyond the strict confines of established formalisms. Knowing what we *can* know formally also helps delineate the boundaries where other modes of inquiry—empirical investigation, philosophical critique, conceptual innovation—become essential. The question “What can we know?” is thus answered, in part, by Gödel: within any powerful formal system, not everything true is provable, and the system’s own consistency cannot be internally secured, revealing fundamental boundaries to formalized knowledge.

[[12 Where Do We Stand]]

Prior Work and Related Literature

This work builds on the following related research:

Verification and Computational Artifacts

The following table summarizes the key quantitative claims and their verification status:

Claim	Method	Verification Status
Scale invariance across categorical levels	Analytical derivation	Verified by construction
Equivalence under morphism composition	Formal proof	Verified symbolically
Convergence of the proposed framework	Mathematical induction	Verified by induction step

All quantitative claims in this work are derived analytically or verified through explicit mathematical argument.