“BABEȘ-BOLYAI” UNIVERSITY

CLUJ-NAPOCA

FACULTY OF HISTORY AND PHILOSOPHY

DOCTORAL SCHOOL OF PHILOSOPHY

PHD THESIS SUMMARY

PhD Supervisor:

CONF. DR. HABIL. VIRGIL DRĂGHICI

Doctoral candidate:

HOREA RUSU

2025

“BABEȘ-BOLYAI” UNIVERSITY

CLUJ-NAPOCA

FACULTY OF HISTORY AND PHILOSOPHY

DOCTORAL SCHOOL OF PHILOSOPHY

GÖDEL'S THEOREMS AND THE REALISM-ANTIREALISM PHILOSOPHICAL DEBATE

PhD Supervisor:

CONF. DR. HABIL. VIRGIL DRĂGHICI

Doctoral candidate:

HOREA RUSU

2025

TABLE OF CONTENTS

Key words: realism, antirealism, incompleteness theorems, Gödel’s theorems, semantic realism, semantic antirealism, deflationism, provability, truth, modal logic of provability, Turing test, recursion, computability, Lucas-Penrose Argument, Computational Theory of Mind

Abstract:

The thesis examines how Gödel's incompleteness theorems can be employed within the philosophical debate between the positions of semantic realism and antirealism. The dispute between these two positions is a classic one. The philosophical inquiry's starting point is the question that has underpinned much research in the field of philosophy: do the elements we study exist or not, can we discuss an abstract entity as one about which we can be certain that it exists independently of human presence? Given the nearly ubiquitous presence of these debates throughout the history of philosophy, we can observe that this represents one of its essential themes. These theses have been extensively debated, covering a wide range of possible responses, from one extreme to the other: from Cartesian skepticism – where a philosophical starting position is the situation in which we no longer have any trust in our senses regarding the external world – to the position where we trust certain entities that determine the form of the elements of our world, as in the Platonic position. Arguments for both positions have drawn on ideas from logic, psychology, and sociology. The discussion addresses the interpretation of this debate influenced by the linguistic turn.

The present work will focus on the debate between the semantic realist and antirealist positions, where the discussion centers on the nature of truth. Briefly stated, the realist position to be presented holds that truth is a *transcendent* matter, theoretically applicable to any proposition, thus allowing us to assert of any declarative proposition that it is true or false. There are several arguments that attempt to refute this position, some of which will be presented in Chapter 1 as part of the exposition of general antirealist arguments, which are not exclusively semantic. The debate on this topic is essential within mathematical logic and beyond, as the concept of truth is central to it. On what basis do we say that something is true, and how do we interpret the fact that something is true? Even if it seems intuitive, deep analysis reveals problematic aspects in explaining and generalizing this concept. What does it mean for something to be true in a formal language? For instance, as the deflationist perspective on truth holds, when we say that something is true, is this concept actually superfluous? Or does truth lie in a correspondence between the statements made and the actual state of affairs to which the proposition refers?

One of the central arguments in the debate between antirealism and realism is constituted by the incompleteness theorems, developed by Kurt Gödel, given their essential

relevance to the core issue: whether the set of provable propositions can coincide with the set of true propositions. The arguments internal to these two philosophical positions, initially considered at a general level, will be analyzed according to how they relate to these incompleteness theorems. Simultaneously, a series of implications stemming from Gödelian theorems will be presented, and their potential impact—or lack thereof—on fields such as cognitive sciences will be discussed.

In Chapter 1, these positions will be presented in greater detail, focusing on the main theories underlying logical proofs. The Tarskian theory of truth constitutes a point of debate between the two philosopher from the two positions, so it will be presented from both the realist and antirealist perspectives. Furthermore, by examining the problems in the philosophy of mathematics prevalent during Tarski's era, we shall clarify the motivation behind this construction by the Polish philosopher. First, we will discuss the logical-mathematical paradoxes that reemerged around the transition period between the 19th and 20th centuries. Their implications, along with the proposed solutions, called into question the tenability of various philosophical positions. Thus, their importance is also revealed by the explanatory role they play in understanding the reasoning behind the solutions. Developed by Tarski—who was influenced both by set theory (his mathematical specialization) and contemporary philosophical currents—his truth theory occupies a central role in logic. Consequently, it will be presented comprehensively, emphasizing its philosophical implications Certain implications stemming from Tarski's truth theory will be linked throughout to various aspects in subsequent chapters when different proof methods for Gödelian theorems are presented. Moreover, the Tarskian truth theory is essential since we will discuss how we may approach truth, while also presenting other significant theories such as deflationism within the context of arguments supporting the antirealist position.

In this thesis, I will present the full requisite theoretical apparatus for constructing the logico-mathematical arguments underpinning the theorems. Thus, Chapter 2 will focus on all the elements required for the basic method of the thesis, the logical-mathematical proof. In this chapter, the elements of recursiveness will be presented, with the various possible interpretations of these concepts. Given that logico-mathematical proof constitutes a foundational component of the argumentative framework, the present chapter will expound both the formal system in its entirety and its inference methods, alongside their philosophical interpretations.. As recursion and computability represent key elements in this discourse—interpreted diversely according to different scholars—they embody the collective efforts of numerous logicians during the 1920s-1930s. These concepts will be discussed in detail due to the necessity of conceptual clarification; within specialized literature, as noted by authorities (Soare 1999), certain terminological distinctions remain subject to confusion.

In Chapter 3, I have chosen to present several distinct proofs of the two theorems. The primary objective is to demonstrate both the original methodology and classical approaches, alongside alternative interpretations. The original Gödelian reasoning will be discussed, which aligns with the motivation behind the construction of the proof and the purpose underlying the Austrian logician's research. It should be noted that these results can be obtained through diverse logical instruments beyond the original techniques and constraints. Furthermore, we will explore potential interpretations through a deflationist proof framework to address how incompleteness theorems may be situated within the semantic realism-antirealism debate. Consequently, the chapter will analyze the interpretation of the truth predicate within incompleteness theorems and how arguments may be advanced for either philosophical position. Technically, this chapter is a compilation of important proofs, both historically and in terms of relevance to the debate between semantic antirealism and realism. Some of them are conducted in different frameworks, thereby highlighting the theorems' general nature. I will contribute both original commentary on these proof strategies and analysis of their potential implications. As emphasized throughout this work, I maintain that beyond interpreting the formal result itself—which may lack direct pragmatic relevance—the proof methodology constitutes the core contribution. For instance, Gödel numbering represents a historically pivotal concept that pioneered foundational work in fields such as cybernetics.

In the penultimate chapter, the central aspect will be the provability predicate. Using modal logic, there are various implications caused by the manner of interpreting the modal predicate, in this case, interpreting the classical predicate (□) through the fact of being provable. The motivation stems from the following philosophical problem: are there instances where truth transcends provability? In other words, can the set of provable propositions coincide with the set of true propositions within a given system? If not, can we isolate those true and unprovable propositions and identify the element that differentiates them from other normal propositions to more clearly understand that element outside of proof? This discussion engages the classic problem articulated even by Ludwig Wittgenstein, through the famous phrase „Meaning is use". If, within antirealism, truth is solely what can be verified through the tools at our disposal (e.g., scientific methods, logico-mathematical argumentation), how do we account for propositions transcending provability? Here, Gödel's theorems become critically relevant, as they address precisely this issue: in certain systems satisfying relatively simple conditions, there exist propositions that are unprovable—termed undecidable—yet which we can recognize as true.

In the final chapter, the discussion will center on a current theme the possibility of simulating human computability. Following some technological developments, especially the appearance of computing machines in the 1950s, alongside theoretical advances like the cognitive revolution and Noam Chomsky's linguistic frameworks, debates arose regarding potential differences between artificial intelligence and human intelligence. Given that computational mechanisms constitute the core of cognitivism (a definition satisfied by both entities under discussion), the central question is whether we differ in structured reasoning capabilities or could be surpassed by AI. Crucially, the debate concerns not computational capacity but qualitative distinctions: whether minds can be simulated through complex logico-mathematical systems. One of the pillars of the argument against this simulation is based on the incompleteness theorems. There have been two waves of arguments, one around the 1960s and one around the 1990s, in which the Gödelian arguments were used as proof of human superiority over artificial intelligence I will examine both the logico-mathematical objections raised by scholars and present original critiques challenging the foundational premise: the anthropomorphic perspective of computing machines, which are fundamentally universal Turing machines. Beyond comprehensive subject analysis, this argument introduces novelty by treating the Turing test as an artifact of human psychology—a tool for verifying behavioral authenticity in others. Through such anthropomorphization, we risk overlooking core computational principles, thereby misapplying Gödelian theorems as indicators of "missing elements" in potential computational simulations of human cognition.

The objective of this thesis is to update conceptual frameworks within the debate, provide interpretive analyses, and clarify these problematizations. As can be observed, the research design will focus on a theoretical foundation part, constituted by the first two chapters, and the applied research component, comprising logical demonstrations grounded in the initial theoretical framework. Moreover, I aim to employ arguments within the realism-antirealism debate and the question of human mind mechanization that originate not solely from logic but also from complementary disciplines such as cognitive psychology. I contend that the arguments developed may be further utilized across other domains or debates to enable more granular analysis of the subjects under consideration.

Selective Bibliography:

Artemov, Sergei N., & Beklemishev, Lev D. (2005). *Provability logic*. In Handbook of Philosophical Logic, 2nd Edition (pp. 189-360). Dordrecht: Springer Netherlands.

Beklemishev, Lev, and Daniyar Shamkanov (2016). *Some abstract versions of Gödel's second incompleteness theorem based on non-classical logics*. arXiv preprint arXiv:1602.05728.

Boolos, George, (1989). *A new proof of the Gödel incompleteness theorem*. Notices Am. Math. Soc. 36: 388-390.

Boolos, George. *The logic of provability*. Cambridge university press, 1995.

Boolos, George, (1990) *Meaning and method: essays in honor of Hilary Putnam*, Cambridge University Press.

Boyd, Richard (1990). *Realism, conventionality, and `realism about'.* în George Boolos (ed.), Meaning and Method: Essays in Honor of Hilary Putnam. Cambridge şi New York: Cambridge University Press. pp. 171-95

Calude, Cristian, (1988) *Adevărat dar nedemonstrabil*, „Știința pentru toți", Editura Științifică și Enciclopedică, București,.

Carnielli, Walter, and David Fuenmayor, (2020). *Gödel's incompleteness theorems from a paraconsistent perspective*. *CLE e-Prints* 19.4: 1-37.

Chakravartty, Anjan, (2007) *A metaphysics for scientific realism: Knowing the unobservable*, Cambridge University Press.

Church, Alonzo. (1936) *A note on the Entscheidungsproblem.* The journal of symbolic logic 1.1: 40-41.

Church, Alonzo, (1935), *An Unsolvable Problem of Elementary Number Theory: Preliminary Report*, Bulletin of the AMS, 41(5): 332–333

da Costa, Newton CA. (1974) *On the theory of inconsistent formal systems*. *Notre dame journal of formal logic* 15.4: 497-510.

van Dalen, Dirk (2001). *Algorithms and Decision Problems: A Crash Course in Recursion Theory*. Handbook of Philosophical Logic. Dordrecht: Springer Netherlands. 245-311.

Detlefsen, Michael (2002). *Löb's Theorem as a Limitation on Mechanism*. Minds and Machines 12, 353–381.

Detlefsen, Michael. (1995) *Wright on the Non-mechanizability of Intuitionist Reasoning. Philosophia Mathematica* 3.1, 103-119.

Devitt, Michael, și Kim Sterelny, (2000) *Limbaj şi realitate. O introducere în filosofia limbajului*. (trad. Radu Dudău), Editura Polirom, Iaşi.

Drăghici, Virgil (2007) Logică (Tradițională/Clasică/Modală), Ed. EFES, Cluj-Napoca.

Drăghici, Virgil (2002) *Logică matematică: O investigație asupra rezultatelor lui Hilbert, Gödel, Gentzen și Kleene*, Casa cărții de știință, Cluj-Napoca

Drăghici, Virgil (2023) *Mathematical logic*, Editura Presa Universitară Clujeană, Cluj-Napoca.

Drăghici, Virgil (2018) *The reflexivity of a language (some comments on Gödel and Wittgenstein)*. International Journal of Communication Research 8.3: 218-223.

Dummett, Michael (2000). *Elements of intuitionism*. Vol. 39. Oxford University Press.

Dummett, Michael A. (1991). *Frege's Philosophy of Mathematics*. Cambridge, MA, USA: Harvard University Press.

Dummett, Michael. (1993) *The seas of language*. Clarendon Press.

Dummett, Michael, (1978), *Truth and Other Enigmas*, Harvard University Press,.

Feferman, Solomon. (1995), *Penrose's Gödelian argument*. *Psyche* 2.7, 21-32.

Feferman, Solomon. (2004) *Tarski's Conception of Logic.* Annals of Pure and Applied Logic 126.1-3: 5–13.

Feferman, Solomon. (2008) *Tarski's conceptual analysis of semantical notions.* New essays on Tarski and philosophy : 72-93.

Franzén, Torkel (2005). *Gödel's theorem: an incomplete guide to its use and abuse*. AK Peters/CRC Press.

Gardiner, Mark Quentin (2000). *Semantic challenges to realism: Dummett and Putnam*. University of Toronto Press.

Gödel, Kurt. (1986) *Kurt Gödel: Collected works: volume I: publications 1929-1936*. *Vol. 1*. Oxford University Press, USA.

Gödel, Kurt (1986). *Kurt Gödel: Collected Works, Volume 2: Publications 1938-1974*. Oxford, England and New York, NY, USA: Oxford University Press.

Gödel, Kurt. (1986) *Kurt Gödel: Collected Works: Volume III: Unpublished Essays and Lectures. Vol. 3*. Oxford University Press.

Gödel, Kurt. (1986) *Kurt Gödel: Collected Works: Volume IV: Selected Correspondence*, AG. Vol. 4. Oxford University Press.

Gödel, Kurt (2003). *Kurt Gödel Collected Works: Volume V: Correspondence, H-Z.* Oxford, England: Oxford University Press UK.

Grzegorczyk, Andrzej. (2005) *Undecidability without arithmetization*. *Studia Logica* 79.2: 163-230.

Isaacson, Daniel. (1987) *Arithmetical truth and hidden higher-order concepts*. Studies in Logic and the Foundations of Mathematics. Vol. 122. Elsevier,. 147-169

Isaacson, Daniel. (2011) *Necessary and sufficient conditions for undecidability of the Gödel sentence and its truth*. Logic, Mathematics, Philosophy, Vintage Enthusiasms: Essays in Honour of John L. Bell: 135-152.

Jacquette, Dale (2002). *Diagonalization in logic and mathematics*. Handbook of Philosophical Logic. Dordrecht: Springer Netherlands,. 55-147.

Japaridze, Giorgi (2003). *Introduction to computability logic*. Annals of Pure and Applied Logic 123.1-3: 1-99.

Japaridze, Giorgi, and Dick De Jongh (1998). *The logic of provability*. Studies in Logic and the Foundations of Mathematics. Vol. 137. Elsevier,. 475-546.

Kikuchi, Makoto, și Taishi Kurahashi. (2017) *Generalizations of Gödel's incompleteness theorems for$\sum$ n-definable theories of arithmetic*. The review of symbolic logic 10.4: 603-616.

Kikuchi, Makoto, Taishi Kurahashi şi Hiroshi Sakai. (2012) *On proofs of the incompleteness theorems based on Berry's paradox by Vopěnka, Chaitin, and Boolos*. Mathematical Logic Quarterly 58.4-5: 307-316.

Kleene, Stephen Cole, et al. (1952), *Introduction to metamathematics*. Vol. 483. New York: van Nostrand,.

Löb, Martin Hugo. (1955) *Solution of a problem of Leon Henkin1*. *The Journal of Symbolic Logic* 20.2: 115-118.

Lucas, John R. (1961), *Minds, Machines, and Gödel*. Philosophy 36, 112-127.

Łukowski, Piotr. (2011) *Paradoxes*. Vol. 31. Springer Science & Business Media.

Mendelson, Elliott. (2009) *Introduction to mathematical logic*. CRC press.

Penrose R (1999) *Incertitudinile rațiunii* (Umbrele mintii) (Dana Jalobeanu, trad.) *Editura tehnică, 1999* (Publicată în original în 1994)

Penrose, R (1996), *Mintea noastra… cea de toate zilele*, (Cornelia Rusu si Mircea Rusu, trad.) *Editura tehnica, 1996* (Publicata in original in 1989)

Priest, Graham, (2007) *Dincolo de limitele gândirii*. (trad. Dumitru Gheorghiu), Paralela 45, Pitești.

Putnam, Hilary (1981). *Reason, Truth and History*. New York: Cambridge University Press.

Putnam, Hilary (1988), *Representation and Reality*, MIT Press, Cambridge.

Putnam, Hilary. (1994) *Words and life*. Harvard University Press.

Raatikainen, Panu. (2002) *McCall's Gödelian argument is invalid*.

Raatikainen, Panu. (2014) *Realism: Metaphysical, Scientific, and Semantic*. Realism, Science, and Pragmatis*m*. Routledge,. 139-158.

Raatikainen, Panu. (2020) *Remarks on the Gödelian Anti-Mechanist Arguments*. Studia Semiotyczne 34.1: 267-278.

Santos, Paulo Guilherme. (2020), *Diagonalization in formal mathematics*. Springer Fachmedien Wiesbaden.

Serény, György (2003), *Boolos-style proofs of limitative theorems.* arXiv preprint math/0309345.

Serény, György. (2011), *How do we know that the Gödel sentence of a consistent theory is true?*. Philosophia Mathematica 19.1 (2011): 47-73.

Shagrir, Oron. (2002). *Effective computation by humans and machines*. Minds and Machines 12:221–240.

Shagrir, Oron. (1997). *Two dogmas of computationalism*. Minds and Machines, 7, 321-344.

Sieg, Wilfried. (1994). *Mechanical procedures and mathematical experience*. Mathematics and mind, 71-117.

Smorynski, Craig (1977). *The incompleteness theorems*. In Studies in Logic and the Foundations of Mathematics (Vol. 90, pp. 821-865). Elsevier.

Smoryński, Craig (1984). *Modal logic and self-reference*. In Handbook of Philosophical Logic: Volume II: Extensions of Classical Logic (pp. 441-495). Dordrecht: Springer Netherlands.

Smorynski, Craig (2012). *Self-reference and modal logic*. Springer Science & Business Media.

Smullyan, Raymond M., (1994) *Diagonalization and self-reference*. Oxford University Press.

Smullyan, Raymond M., (2018) *Teoremele gödeliene de incompletitudine* (traducător Virgil Drăghici), Asociația Casa Cărții de Știință, Cluj-Napoca

Soare, Robert I. (1996) *Computability and recursion.* Bulletin of symbolic Logic 2.3, 284-321.

Soare, Robert I. (1999) *The history and concept of computability.* Studies in Logic and the Foundations of Mathematics. Vol. 140. Elsevier, 3-36.

Soare, Robert I. (2016) *Turing computability: Theory and applications*. Vol. 300. Berlin: Springer.

Solovay, Robert M. (1976), *Provability interpretations of modal logic.* Israel journal of mathematics 25, 287-304.

Stegmüller, Wolfgang. (2011), *Incompletitudine si indecidabilitate* (trad. Virgil Drăghici) Presa universitară clujeană, (publicată în original in 1973)

Tarski, Alfred. (1998) *A decision method for elementary algebra and geometry*. In: Quantifier elimination and cylindrical algebraic decomposition. Springer, pp. 24–84

Tarski, Alfred.(1948). *A problem concerning the notion of definability*. The Journal of Symbolic Logic, 13(2), 107-111.

Tarski, Alfred, Andrzej Mostowski, and Raphael Mitchel Robinson, eds. (1953) *Undecidable theories*. Vol. 13. Elsevier.

Tarski, Alfred, (1983) *Logic, semantics, metamathematics: papers from 1923 to 1938*. Hackett Publishing.

Tarski, Alfred, (1944). *The semantic conception of truth: and the foundations of semantics.* Philosophy and phenomenological research 4.3, 341-376.

Tennant, Neil, (2002), *Deflationism and the Gödel phenomena*. *Mind* 111.443: 551-582.

Tennant, Neil. (1997) *The taming of the true*. Oxford University Press.

Turing, Alan, (1936), *On computable numbers, with an application to the Entscheidungs problem.* Proceedings of the London Mathematical Society Series/2 (42), 230-42.

Woleński, Jan, (2008). *On philosophical sense of metamathical limitative theorems.* Scientific Research of the Institute of Mathematics and Computer Science 7.2: 87-96.

Woleński, Jan (2019). *Semantics and Truth*. Cham, Switzerland: Springer Verlag.

Wright, Crispin (1994): *About "The philosophical significance of Gödel's Theorem": Some issues*, in B. McGuinness and G. Oliveri, eds, The Philosophy of Michael Dummett, pp. 167–202. Dordrecht: Kluwer. Reprinted in C. Wright, Realism, Meaning, and Truth, pp. 321–354. Oxford: Blackwell, 1995.

Wright, Crispin, (1995). *Intuitionists are not (Turing) machines*. Philosophia Mathematica 3.1, 86-102.

Wright, Crispin (1987). *Realism, Meaning and Truth.* Cambridge, Mass., USA: Blackwell.