Gödel's Theorem and Its Philosophical Interpretations: From mechanism to postmodernism

Stanisław Krajewski[1]

This is an English summary of my Polish book *Twierdzenie Gödla i jego interpretacje filozoficzne: od mechanicyzmu do postmodernizmu*, Wydawnictwo IFiSPAN, Warszawa 2003.

This book is a study of Gödel's Incompleteness Theorem. The focus here is, first, on the consequences and interpretations of it in the philosophy of mathematics, philosophy of mind, and logic, and second, on a discussion of attempts to apply the theorem in areas of the humanities, such as literary criticism, social studies, and the theory of law. Considerable space is also devoted to the philosophical views and logical achievements of Gödel, widely seen as "the greatest logician since Aristotle".

Chapter I describes the background of Gödel's work in the study of the foundations of mathematics generally and Hilbert's program in the early 20th Century. Since this book is not a mathematical textbook, summaries, rather than complete technical presentations, are provided. In addition to standard topics some new developments are mentioned. The reception of Gödel's work is seen as falling into three periods: from 1931, when his celebrated paper on incompleteness appeared, to the mid-1950s, by which time it became an accepted part of graduate courses in logic; from the mid-1950s, when the book by Nagel and Newman, 1958 appeared and mathematicians, scientists, and philosophers became more generally aware of it; and finally, from 1979, when the bestseller *Gödel*, *Escher*, *Bach* by Douglas Hofstadter appeared and the general educated public became aware of Gödel's work, and references to it began to appear in work far removed from mathematical logic. In this last period it seems that an effort to avoid misleading oversimplifications in discussions of Gödel's work is needed. This will be especially important when we enter the fourth period, when Gödel's result will be taught in schools.

Chapter II gives a new and more detailed analysis than previously available of an argument by Lucas, 1961, which was an attempt to show that Gödel's Theorem disproves mechanism, that is, that the mind is not mechanical or equivalent to a machine. Then new versions of Lucas' argument given by Penrose, 1989 and 1994, are described and also found defective. As well, Gödel's own views are presented. It was he who first suggested that his results do not exclude that "there may exist (and even be empirically discoverable) a theorem-proving machine which

[1] I am grateful to Richard L. Epstein for help in preparing the English version of this article.

Bulletin of Advanced Reasoning and Knowledge 2 (2003) 103–108.

in fact is equivalent to mathematical intuition, but cannot be proved to be so, nor even be proved to yield only correct theorems on finitary number theory." This chapter and perhaps even the entire text can be understood as an exegesis of this quotation.

In addition to the analysis of Gödelian anti-mechanist arguments, a general theorem is proved, extending work of others, that shows that every kind of argument in the style of Lucas must be either circular or lead to an inconsistency. Such arguments assume that a theorem-proving machine is equivalent to human mathematical powers, even in the realm of elementary arithmetic. Then, the argument goes, the totality of the theorems provable by the machine is either inconsistent, and then not equivalent to us, or consistent, and then we produce the Gödel sentence for the machine that we can see to be true, but it can't be among the theorems proved by the machine. I argue that the argument must respond to every machine that is equivalent to a specifiable Turing machine, or rather, at least, to every consistent machine; furthermore, we require the response to consist in the presentation of an arithmetical sentence not "provable" by the machine, and we assume that this response is effectively determined (otherwise, we would, circularly, assume non-mechanical abilities of the mind). Using Church's Thesis we obtain a partial recursive function F defined for at least consistent machines (i.e., machines whose arithmetical output is consistent) and such that for the n^{th} machine for which $F(n)$ is defined, $F(n)$ is an arithmetical sentence outside the output of the n^{th} machine. The Theorem on Inconsistency states that under those assumptions the set of values of F is inconsistent. (Note that we assume neither that $F(n)$ is produced using Gödel's technique nor that $F(n)$ is true.) A variant related to Penrose's work, the Theorem on Unsoundness, states that if F is defined for (at least) sound machines (ones that prove only true sentences) then the set of values of F is unsound. Loosely put, the former result shows that Lucas is inconsistent, and the latter that Penrose is unsound.

Chapter III surveys Gödel's work in mathematics, emphasizing the conceptual novelties he introduced. As well, an outline of his philosophical views is given, including his Platonism, his Leibniz-style monadology, and admiration for Husserl, along with his theological views. These discussions follow the lines of earlier work by Hao Wang, 1987 and 1996, and the commentary in the edited versions of Gödel's collected works. Questions are raised why Gödel said nothing about the work of Alfred Tarski on formal semantics and definitions of truth, seeking to see if there is an explanation, as Feferman, 1984 has urged, in terms of Gödel's philosophy and Weltanschauung. After reviewing the contacts of Gödel and Tarski, along with all mention of Tarski in Gödel's papers and manuscripts, the following possible explanations are suggested: (i) Gödel's idea of truth may have been something like an intuitive provability "in general"; (ii) he may have seen truth as an inexhaustible idea in the sense of Kant; (iii) even though Gödel was the first to understand well the methods of model theory, he seems to prefer the idea of logic as a universal language, in the sense developed by Hintikka, 1997, as opposed to the view of logic as a calculus that can be always reinterpreted. The chapter concludes with a discussion of Gödel's mental instability and possible connections of his paranoia to the nature of his philosophy, as first raised by Dawson, 1997. Specifically, his instability is related to the special role of logic and mathematics in his vision of a scientific philosophy.

Chapter IV discusses the uses and abuses of Gödel's results. In mathematics Gödel's Incompleteness Theorem is usually taken to mean that the consistency of a sufficiently strong theory cannot be proved inside the theory itself. But we now know that it matters how the assertion of consistency is expressed. This "intensional" aspect is seen by Kadvany, 1989, p. 177 as an example of the presence of an historical dimension in mathematics.

Another standard consequence of the Incompleteness Theorem is expressed by saying that there is no universal, effectively presented, mathematical theory. Yet a subtlety discovered by Gödel implies that while there is no universal theory that can be known by us, it is not excluded by his discoveries that there could be an effectively presented theory capturing "subjective" mathematics, that is, the totality of mathematical theorems accessible to the human mind. Still, it seems that there is no finite description of the natural numbers that we could formulate and give to a computer to make it behave as if it understood our notion of number.

Gödel's Theorem has been applied in philosophy to the issue of mechanism (discussed above) and to argue against the view that arithmetic is composed of only analytic truths, and more generally against the view that *a priori* truths must be analytic. It is suggested that no conclusive argument is possible for the latter issue. Another philosophical application of Gödel's results is to relate them to the classical, Pyrrhonist, tradition of skepticism, since the consistency of fundamental theories is not provable without circularity. In a separate section, Wittgenstein's intentional disregard of Gödel's work is argued as part of his fight against the Platonic view of mathematics and his opposition to the view that something like Hilbert's program is needed in the foundations of mathematics.

Another section is devoted to mistaken applications of Gödel's work by post-modernist authors. Examples are given from Sokal and Bricmont, 1997 and other sources. "Gödelity is a widespread disease", admitted one author (Debray, 1996, p. 6) about the tendency to use mathematical incompleteness in the humanities after he himself had been criticized for just that. On the other hand, it is argued that a "post-modernist" approach does make sense in the foundations of mathematics because of incompleteness: There are different, mutually inconsistent useful set-theories, and the choice of one rather than another is arbitrary. However, Gödel himself was a Platonist who believed that we are capable of discovering the one true set-theory. While Gödel's Theorem can inspire valuable work in various fields, attempts to use it in sociology, the theory of law, psychology, medicine, and in natural science are found to be largely misguided. The theorem has also become, not without good reasons, one of the symbols of a scientific paradigm shift or of limitations in the development of the Euclidean and Cartesian views: In the Twentieth Century we gradually abandoned the idea of the possibility of a complete, certain, absolute description of the world.

In order to avoid oversimplifications and errors in the application of Gödel's work, a *Guide to Popularization* is proposed. Among the following formulations of Gödel's Theorem some, although not quite rigorous, are nonetheless correct if properly understood.

G1 Every consistent mathematical theory that is effectively presented and which includes elementary arithmetic is incomplete (i.e., leaves some arithmetical sentences undecided).

G2 No computer that produces only mathematical truths can produce all such truths.

G3 Mathematics is inexhaustible; it is undefinable by an algorithm.

G4 Mathematical truth is not reducible to provability in any given system.

G5 In every strict description of mathematics something is missing.

G6 There is no effectively presented theory of the whole of mathematics or even arithmetic.

The above limitations refer to objective mathematics. Gödel did not prove that subjective mathematics, that is mathematics that is potentially accessible to the human mind, could not be expressible as one theory or one algorithm. If such a theory existed we could not understand it.

Gödel's Second Incompleteness Theorem can be presented as follows, assuming that the formal assertion of consistency is made properly:

G7 The consistency of a consistent, effectively presented mathematical theory that includes elementary arithmetic is not provable within the theory.

G8 No computer producing only mathematical truths can produce the truth that it is consistent.

G9 The consistency of mathematics must be assumed, taken as a matter of faith.

One formulation by Gödel himself is noteworthy:

G10 "Any systematic procedure for solving problems of all kinds must be nonmechanical."

For the general public, the following aspects of Gödel's Theorem seem to be most attractive, even though they are vague and can be seriously misleading:

A1 The human mind can do something no computer can.

A2 There are unprovable truths.

A3 Consistency and completeness are incompatible.

A4 There are sentences that are provably neither provable nor refutable.

A5 A formal approach is not enough; intuition or faith is indispensable.

A6 Given any system, it is necessary to take into account a higher level, a metasystem.

A7 The method of arithmetization.

A8 Self-reference can be described rigorously.

While all these aspects can be used fruitfully, they can lead to misunderstandings and mistakes. The following, corresponding to the points above, must be kept in mind:

B1 It is not proven that there is no theory or program that is equivalent to our mathematical powers.

B2 The truth of Gödel's self-referential sentence (a formal counterpart of "I am not provable") is a normal mathematical truth, and not a special new one.

B3 In Gödel's work, consistency and completeness are strict technical concepts of mathematical logic, which must not be thoughtlessly extrapolated to other theories.

B4 The unprovability of Gödel's sentence is not absolute; it can be proved in other reasonable theories.

B5 Gödel's Theorem is provable in a weak theory, unlike Gödel's sentence.

B6 Stepping outside a theory to a higher level is not sufficient for the proof of Gödel's sentence, and it is not necessary for the proof of Gödel's Theorem.

B7 The very possibility of arithmetization means that despite incompleteness, elementary arithmetic is a strong theory.

B8 Self-reference is important, but some proofs of incompleteness don't use it. Nonetheless, some form of diagonal argument seems unavoidable.

In addition, the following remark summarizes the criticism of misguided applications of Gödel's work in various fields.

B9 Applying Gödel's Theorem to theories outside mathematics must be preceded by a formalization of such theories. The independent sentences defined with the methods specific to the field will most probably be much more illuminating than the hypothetical counterpart of Gödel's sentence.

Bibliography

Dawson, John
1997 *Logical Dilemmas*: *The Life and Work of Kurt Gödel*
A. K. Peters.

Debray, Regis
1996 L'incomplétude, logique du religieux?
Bulletin de la Société française de philosophie, vol. 90, pp. 1–35.
Feferman, Solomon
1984 Kurt Gödel: Conviction and Caution
In *Philosophy of Science–History of Science*: *A Selection of Contributed Papers of the 7th International Congress of Logic*, *Methodology and Philosophy of Science*, *Salzburg 1983*, P. Weingartner and Ch. Puhringer ed.s, Anton Hain, Meisenheim/Glan. Also in *In the Light of Logci*, by S. Feferman, Oxford University Press, 1998.
Gödel, Kurt
1986– *Collected Works*
1995 Solomon Feferman et al. ed.s, Volume I, 1986; Volume II, 1990; Volume III, 1995; Oxford University Press.
Hintikka, Jaakko
1997 *Lingua Universalis vs. Calculus Ratiocinator*: *An Ultimate Presuppostion of Twentieth-Century Philosophy*
Kluwer.
Hofstadter, Douglas R.
1979 *Gödel, Escher, Bach, an Eternal Golden Braid*
Basic Books.
Kadvany, John
1989 Reflections on the legacy of Kurt Gödel: Mathematics, skepticism, postmodernism
The Philosophical Forum, vol. 20, pp. 161–181.
Lucas, John
1961 Minds, machines, and Gödel
Philosophy, vol. 36, 1961, pp. 112–127.
Nagel, Ernest and James R. Newman
1958 *Gödel's proof*
New York University Press.
Penrose, Roger
1989 *The Emperor's New Mind*
Oxford University Press.
1994 *Shadows of the Mind*
Oxford University Press.
Sokal, Alan and Jean Bricmont
1997 *Impostures intellectuelles*
Editions Odile Jacob, Paris.
Wang, Hao
1987 *Reflections on Kurt Gödel*
MIT Press.
1996 *A Logical Journey: From Gödel to Philosophy*
MIT Press.