On a Relationship between Gödel's Second Incompleteness
Theorem and Hilbert's Program

Ryota Akiyoshi*

Abstract

We can find in several places an assertion that Gödel's second incompleteness theorem defeated Hilbert's program. But, (as M. Detlefsen argued in his book [3]) in order to establish this assertion, we need to address additional issues.

First we formulate Hilbert's program. Second we reconstruct a standard argument for the claim that Gödel's second incompleteness theorem defeated Hilbert's program. In doing so, we formulate a critical, and problematic assumption which we call "DCT" (Derivability Conditions Thesis). Finally we examine three arguments whose aims are to justify DCT. We show that the first and the second argument are not valid, and discuss the third argument, which is based on Kreisel's idea. We identify a difficulty in this argument as well. After examining the difficulty, we conclude that we cannot claim that Gödel's second incompleteness theorem defeats Hilbert's program. Moreover we clarify what is essentially needed for such an argument to succeed.

Key words: Hilbert's program, Gödel's second incompleteness theorem, Derivability conditions

1. Introduction

We can find in several places an assertion that Gödel's second incompleteness theorem defeated Hilbert's program. But, (as M. Detlefsen argued in his book [3]) in order to establish this assertion, we need to address additional issues.

* Keio University
E-mail: georg_logic@s7.dion.ne.jp
This paper is part of my master thesis [1]. I read this paper at the Conference of Japan Association for Philosophy of Science 2007 held at Tottori University. I thank Ken Shiotani, Kengo Okamoto, and Masanao Ozawa for their helpful comments and discussions. Discussions with them at the conference made my thought much clearer than before. I want to express my gratitude to Michael Detlefsen for reading an earlier version of this paper, and commenting on it. Finally, I would like to express my deep appreciation toward my adviser Takashi Iida for his successive encouragement, comments, and suggestions.
Notice that Detlefsen's work deserves much credit for our argument. However, it is not our aim to examine his arguments. In his book [3], Detlefsen interpreted Hilbert's pro-
First we must select a workable formulation of Hilbert’s program. Several formulations exist, but we shall follow that proposed by G. Kreisel (Section 1.1)\(^2\). Second we must reconstruct a standard argument for the claim that Gödel’s second incompleteness theorem defeats Hilbert’s program, and also identify those assumptions which are essential to the standard argument (Section 2). In doing so, we formulate a critical, and problematic assumption which, we call “Derivability Conditions Thesis” (DCT). Finally, we examine three arguments whose aims are to justify DCT (Section 3). As to the first and second of these arguments, we present counter arguments and conclude that these arguments are not valid. Then we discuss the third argument, which is based on Kreisel’s idea. We identify a difficulty in this argument as well. After careful consideration of this difficulty, we conclude that we cannot say that Gödel’s second incompleteness theorem, even when supplemented by Kreisel’s argument, defeats Hilbert’s program. Furthermore, we clarify what is essentially needed for such an argument to succeed.

This paper is part of my master thesis [1].

1.1. Brief Description of Hilbert’s Program

Hilbert’s program was a foundational program whose final aim was a justifica-

---

1. Ideal proofs of real theorems which are more complex than any real proof of the same theorem do not yield an instrumental advantage.

2. Ideal proofs which are too long or complex to be comprehended by humans, hence never play a role in actual mathematical reasoning, are also of no instrumental value.

Then, in order for the success of Hilbert’s program, we need a finitary consistency proof for only part of ideal theory which is indeed useful. Detlefsen suggests that the theory to be considered in the context of strict instrumentalism may be a subsystem of a whole ideal theory. Such a theory may be too weak for the proof of Gödel’s second incompleteness theorem because, as he suggested, the theory may be finitary subsystem of a whole ideal theory. If Detlefsen’s suggestion is correct, then the theory may be too weak for the representation theorem, which is necessary for the proof of the diagonalization lemma, and Hilbert’s program under the strict instrumentalism might escape from the argument by Gödel’s second incompleteness theorem which will be dealt with in this paper.

From our description it is clear that his interpretation of Hilbert’s program is considerably different from the usual interpretation of the program because Hilbert’s original intention had been to justify mathematics as a whole. Our aim can be summarized as examining a relationship between Gödel’s second incompleteness theorem and Hilbert’s program under the standard interpretation.

---

\(^2\) There is a very important philosophical problem in the formulation of Hilbert’s program (especially formulation of Hilbert’s program via formalization using reflection principle). We will discuss this issue at another time.
tion of mathematics as a whole. According to Hilbert, mathematics can be divided into two parts. The first part is finitistic mathematics, which contains real elements that do not require any foundation\(^3\). Note that although it is possible to formalize it in some way, finitistic mathematics does not need to be a formal theory. Further, finitistic mathematics consists of combinatorial operations on real elements and mathematical induction on properties of real elements. Therefore it is clear that (formalized) finitistic mathematics is a subsystem of Peano Arithmetic (\(PA\))\(^4\). The second part is non-finitistic mathematics, which contains also ideal elements. Of course, most of mathematics does not fall within finitistic mathematics. But Hilbert intended to justify mathematics as a whole which contains ideal elements by providing an instrumentalist interpretation\(^5\). According to an instrumentalist interpretation, ideal statements are useful instruments to derive real statements. Thus ideal elements do not have meanings that are based on intuition as real elements are assumed to. Rather ideal elements have meanings that are based on their utility. To justify mathematics in this way, that is, to show that non-finitistic mathematics is safe as an instrument, it must be shown that the use of ideal elements cannot lead to any false statements in finitistic mathematics. At this point formalization of mathematics plays a crucial role because, through such formalization, we can develop new mathematics (Proof Theory) about our mathematical activity. This was one of Hilbert’s great discoveries. Notice that the metamathematical properties of a formal theory can be stated meaningfully in finitistic standpoint.

Several different formulations of Hilbert’s program exist. Perhaps, usually, the most common formulation relies upon a formalized version of finitistic mathematics. If we had followed this formulation in the present article, the description below would be shorter. However, we do not adhere to this tradition here. Let us explain our choice. First, as previously mentioned, it is not required that finitistic mathematics is a formal system. Second, if the standard argument described below were correct, then this argument would establish the failure of Hilbert’s program, namely the informal unprovability of the consistency of a formal system by Gödel’s second incompleteness theorem, a theorem which asserts the formal unprovability of its consistency (if it is consistent). Therefore the distinction between informal mathematics

---

\(^3\) As to the problem to what extent mathematics (especially arithmetic) we can admit as finitistic, we do not discuss this because our argument is neutral in this respect. For this problem, see [18, 19, 21].

\(^4\) In this paper we consider Hilbert’s program only in the original sense. Therefore we exclude the possibility that finitistic mathematics includes transfinite induction (restricted to a quantifier free formula) up to \(\epsilon_0\) in the case of Gentzen, or something like that.

\(^5\) The instrumentalist interpretation of Hilbert’s program has become popular since Kreisel’s papers [11, 12, 13, 14], but there is an objection to this. For this issue see [8, 9].
and formal mathematics is very important for the standard argument. Third, we will formulate our key assumption, which we call “Derivability Condition Thesis”, as a necessary component of the standard argument. This assumption fills a gap between informal unprovability and formal unprovability. Therefore we may say that the distinction between informal mathematics and formal mathematics will play a great role in our argument.

We assume the usual arithmetization of metamathematics in what follows. Let $T$ be a formal theory which contains ideal elements (typically $PA^6$), and $\text{Prf}_T(y, x)$ be a relation which means “$y$ is a proof of a formula $x$ in $T$”. This relation is meaningful in finitistic standpoint. Let $P(x)$ be a meaningful statement in finitistic standpoint, and $A$ be a class of such statements. Let $f$ be a function that maps from the class $A$ to a class of formulas of $T$, and let $\pi$ be a function that maps from a class of informal proofs of statements in $A$ to a class of formal proofs in $T^8$. One might argue that use of the latter two functions is unparsimonious. But it will become clear that they are necessary in bridging a gap between informal finitistic mathematics and formal mathematics, which Kreisel expressed as “the abstract theory”, which contains also ideal elements.

Using these formalisms, following Kreisel, we next formulate Hilbert’s program$^9$. The goal of Hilbert’s program is to prove the following HP using only resources of finitistic (meta) mathematics.

**HP**: Let $a$ be any natural number, and $\bar{n}$ be any finite sequence of natural numbers.

Then

$$\text{If } \text{Prf}_T(a, \pi f(P(x))) \text{ then } P(\bar{n}).$$

In a similar way the consistency of $T$ can be formulated as $\neg\text{Prf}_T(a, \pi f(0 = 1))$. Note that HP and the consistency of $T$ can be defined in finitistic (meta) mathematics. Hilbert’s great discovery$^{10}$ was that (if $T$ satisfies some conditions$^{11}$) HP is

---

$^6$ The typical case is $PA$, but arguments in this paper go through in the cases of $PA^2$, or $ZF$, etc...

$^7$ $\bar{x}$ represents a finite sequence of free variables $x_1, \ldots, x_n$.

$^8$ Kreisel did not use the function $f$, but the notation $P(\bar{x})$ which corresponds to $f(P(\bar{x}))$ in our notation, and he explained the function $\pi$ as “(elementary) functions which essentially describe how the intended computation of $P(\bar{n})$ can be mimicked in the abstract theory,” in [12, p.152]. But Kreisel used the function $f$ from $A$ to a class of formulas of $T$, and formulated HP in a more abbreviated way in [13]. Our formulation here is a combination of Kreisel’s two formulations.

$^9$ cf. [12, p.152], [13, p.322], and [14, p.233].

$^{10}$ As pointed out by M. Detlefsen, it would be more accurate to say Hilbert-Bernays’s discovery.

$^{11}$ These can be stated as follows ([12, p.152]).

(1) If $P(\bar{n})$ then $\text{Prf}_T(\pi(P(\bar{n})), \pi f(P(\bar{n})))^\pi).$
equivalent to the consistency of $T^{12}$.

Therefore to prove the consistency of $T$ in finitistic standpoint is sufficient for the success of Hilbert’s program.

2. DC Thesis

We have several ways in which Hilbert’s program seems to be defeated by incompleteness theorems. By the first incompleteness theorem, (if $T$ is consistent and contains at least Robinson arithmetic $Q^{13}$) there is an unprovable but true sentence in it. As is well known, it is impossible to formalize the whole mathematics. Then Hilbert’s program fails in the first step. But in this paper we consider a more direct, and standard argument using the second incompleteness theorem.

Let $S$ be a (formalized) finitistic mathematics, and $T$ be a formal theory which contains ideal elements (typically $PA$). Then, as previously indicated, $S$ is a subsystem of $PA$. Further it is plausible that every (informal) argument in finitistic (meta) mathematics can be formalized in $S$. Then, in order to establish the failure of Hilbert’s program, it is sufficient to show, for any formula $\text{Con}_T$ which expresses the consistency of $T$, $\not \vdash_S \text{Con}_T^{14}$. Moreover, for any formula $\text{Con}_T$ which expresses the consistency of $T$, $\text{Con}_T$ is of the form $\text{Con}_T \equiv \neg \phi(\overline{0} = \overline{1})$ (where $\phi(x)$ is a formula equivalent in $T$ to some provability predicate $\text{Prf}_T(x)$ expressing the provability in $T$). We will see later that we can establish the failure of Hilbert’s program using

\[
(2) \text{If } \neg P(\overline{n}) \text{ then } \text{Prf}_T(\pi(\neg P(\overline{n})), \forall f(\neg P(\overline{n})))'.
\]

These conditions state that the formal theory $T$ contains elementary arithmetic.

---

12 For one direction, note that $(0 = 1)$ is a false meaningful statement in finitistic mathematics. For another direction, suppose that $T$ is consistent, that is, $\neg \text{Prf}_T(b, \forall f(0 = 1))$ for any given natural number $b$, $\text{Prf}_T(a, \forall f(P(\overline{x})))$, and $P(\overline{n})$ is false. Then $\neg P(\overline{n})$ is true. From the condition stated before, it holds that $\text{Prf}_T(\pi(\neg P(\overline{n})), \forall f(\neg P(\overline{n})))$. From the assumption $\text{Prf}_T(a, \forall f(P(\overline{n})))$ holds. Therefore $\text{Prf}_T(\pi, \forall f(P(\overline{x})) \land f(\neg P(\overline{n})))$ for some $c$. Now, since $T$ contains propositional logic, $\text{Prf}_T(\pi, \forall f(0 = 1))$ for some $d$. But this contradicts the assumption $\neg \text{Prf}_T(d, \forall f(0 = 1))$. Therefore $P(\overline{n})$ is true for any given $\overline{n}$. Note that this argument itself is meaningful from finitistic standpoint (cf. [14, pp.235–236]), and that we needed only the condition (2). Kreisel mentioned both in [12, p.152], but, only the second condition in [14, p.236].

13 Roughly speaking, $Q$ is arithmetic without mathematical induction.

14 One might raise the question of why we need to consider any formula $\text{Con}_T$ which expresses the consistency of $T$. In fact, we consider usually a formula $\text{Con}_T$ which expresses the consistency of $T$ in the context of Hilbert’s program. But we can reply to this question as follows. It seems to us that Hilbert’s program succeeds if and only if $\vdash_S \text{Con}_T$ for a formula $\text{Con}_T$ which expresses the consistency of $T$. Then Hilbert’s program does not succeed if and only if $\not \vdash_S \text{Con}_T$ for any formula $\text{Con}_T$ which expresses the consistency of $T$. Therefore we need to consider any formula expressing the consistency (or the provability) in $T$. 

— 17 —

NII-Electronic Library Service
these propositions, the DCT assumption, and the second incompleteness theorem. DCT plays a role in filling a gap between the informal unprovability like the failure of Hilbert’s program and the formal unprovability which the second incompleteness theorem asserts.

Derivability conditions are conditions for a provability predicate which are necessary for the second incompleteness theorem. In order to formulate DCT, we have to identify derivability conditions because DCT varies depending on types of derivability conditions. Although there are several types of derivability conditions, we rely upon the most familiar derivability conditions. This is compatible with our aim of considering the most standard of the arguments designed to defeat Hilbert’s program by using the second incompleteness theorem. It is quite possible that the standard argument, which many people have in their mind, depends on the usual proof of the second incompleteness theorem.

Let \( \phi, \psi \) be any formulas of \( T \).

(DC1) If \( \vdash T \phi \), then \( \vdash T \Pr_T(\overline{\phi}) \).

(DC2) \( \vdash T \Pr_T(\overline{\phi}) \rightarrow \Pr_T(\overline{\Pr_T(\overline{\phi})}) \).

(DC3) \( \vdash T (\Pr_T(\overline{\phi}) \wedge \Pr_T(\overline{\phi \rightarrow \psi})) \rightarrow \Pr_T(\overline{\psi}) \).

Now we can formulate DCT as follows.

\textbf{DCT}: If, for any provability predicate \( \Pr_T(x) \), \( \Pr_T(x) \) means the provability of \( T \), then \( \Pr_T(x) \) satisfies the derivability conditions\(^{15}\).

Let us reconstruct the standard argument using the assumptions described above, DCT, and the second incompleteness theorem. As noted, it is sufficient to show, for any formula \( \text{Con}_T \) which expresses the consistency of \( T \), \( \forall S \text{ Con}_T \) in order to establish the failure of Hilbert’s program. Now take any formula \( \text{Con}_T \) which expresses the consistency of \( T \). Then (by the assumption above) \( \text{Con}_T \) is of the form \( \text{Con}_T \equiv \lnot \phi(\overline{0 = 1}) \) (where \( \phi(x) \) is a formula equivalent to some provability predicate \( \Pr_T(x) \) expressing the provability in \( T \)). By DCT, this \( \Pr_T(x) \) satisfies the derivability conditions. Therefore \( \forall T \lnot \Pr_T(\overline{0 = 1}) \) by the second incompleteness theorem\(^{16}\). Assume that \( \vdash T \text{ Con}_T \). Because the formula \( \Pr_T(x) \) is equivalent in \( T \) to \( \phi(x) \), it must be that \( \vdash T \Pr_T(\overline{\psi}) \leftrightarrow \phi(\overline{\psi}) \) for any formula \( \psi \) of \( T \). Then \( \vdash T \lnot \Pr_T(\overline{0 = 1}) \), which is a contradiction. Therefore \( \forall T \lnot \text{ Con}_T \). From this, we

\(^{15}\) This formulation is meaningless for a finitist because the formula \( \Pr_T(x) \) contains an unbounded quantifier. But we should not be worried about this because our purpose is not to formulate the standard argument in the framework of the finitistic standpoint.

\(^{16}\) We may assume the consistency of \( T \). We can admit the consistency of \( S \) because \( S \) is a system of very elementary arithmetic, and we could prove the consistency of \( S \) if finitistic mathematics were formalized as \( S \) in some appropriate way. Thus we obtain \( \forall S \text{ Con}_T \) even if \( T \) is inconsistent.
obtain \( \forall_S \text{Con}_T \) because \( S \) is a subsystem of \( T \).

We will investigate some arguments in Section 3 whose aims are a justification of DCT in order to examine the standard argument.

3. Justification of DCT

3.1. The First Argument

The first argument is very naive and simple. The underlying idea can be summarized as follows. The derivability conditions express the truths about the provability in \( T \), and an advocate of the first argument asserts that the provability predicate \( \Pr_T(x) \) which means the provability in \( T \) must register every truth concerning the provability in \( T \) as a theorem of \( T^{17} \). For example, following him, DC3 expresses the truth that the provability in \( T \) is closed under modus ponens. Then \( \Pr_T(x) \) which means the provability in \( T \) must satisfy the derivability conditions. If this argument were justifiable, the standard argument could survive.

Nevertheless, it is possible to identify two defects. If a truth concerning the unprovability in \( T \) is admitted as one about the provability in \( T \), then following this argument, the provability predicate \( \Pr_T(x) \), which means the provability in \( T \), must register every truth concerning the unprovability in \( T \) as a theorem of \( T \). But we can show that such a provability predicate \( \Pr_T(x) \) does not exist by the following theorem.

**Theorem** Let \( T \) be a formal theory which is a consistent recursive extension of \( Q \), and \( \phi \) be any formula of \( T \). Then, there does not exist a provability predicate \( \Pr_T(x) \) satisfying the conditions below.

\[
\begin{align*}
\text{If } & \vdash_T \phi \quad \text{then } \vdash_T \Pr_T(\overline{\phi^2}), \\
\text{If } & \not\vdash_T \phi \quad \text{then } \vdash_T \neg \Pr_T(\overline{\phi^2}).
\end{align*}
\]

**Proof.** Assume that there exists such a predicate \( \Pr_T(x) \). By the diagonalization lemma there exists a sentence \( \psi \) such that

\[
(1) \quad \vdash_T \psi \leftrightarrow \neg \Pr_T(\overline{\psi^2}).
\]

Applying the conditions to \( \psi \), we obtain

\[
\begin{align*}
\text{(2) If } & \vdash_T \psi \quad \text{then } \vdash_T \Pr_T(\overline{\psi^2}), \\
\text{(3) If } & \not\vdash_T \psi \quad \text{then } \vdash_T \neg \Pr_T(\overline{\psi^2}).
\end{align*}
\]

\[^{17}\text{Probably the notion of truth does not need to be formal one in this context for an advocate of the first argument (if there is). In other words, it does not need to be characterized by any formal concepts as, typically, the provability in } T.\]
Assume \( \vdash_T \psi \). From (2) and (1), \( \vdash_T \Pr_T(\overline{\psi^\perp}) \) and \( \vdash_T \neg \Pr_T(\overline{\psi^\perp}) \). But this contradicts the assumption that \( T \) is consistent. Thus \( \not\vdash_T \psi \). By (3) we obtain \( \vdash_T \neg \Pr_T(\overline{\psi^\perp}) \). But then from (1), \( \vdash_T \psi \) holds. This contradicts \( \not\vdash_T \psi \). Therefore a predicate such as \( \Pr_T(x) \) does not exist. \( \Box \)

In the usual proof of the second incompleteness theorem, we construct a provability predicate \( \Pr_T(x) \) parallel to the definition of syntax. Such a predicate \( \Pr_T(x) \) can be admitted as expressing the provability in \( T \) without doubt. Following this argument, the provability predicate \( \Pr_T(x) \) must register every truth concerning the provability in \( T \) as a theorem of \( T \). Is this possible? As shown above, it is impossible if the truth of the unprovability in \( T \) can be included in the set of the truths about the provability in \( T \). Therefore if the first argument were right, this \( \Pr_T(x) \) would not express the provability in \( T \). But this contradicts our belief that such a predicate \( \Pr_T(x) \) expresses the provability in \( T \). Even if our argument were not valid, the first argument would be very doubtful unless an advocate of the first argument could show the fact that this provability predicate \( \Pr_T(x) \) could register every truth concerning the provability in \( T \) as a theorem of \( T \).

### 3.2. The Second Argument

The second argument is due to A. Mostowski\(^{18} \). In his survey [17], after reviewing the works of Gödel ([7]), Hilbert and Bernays ([10]), and Feferman ([5]), he presented an argument which depended on the following criterion.

**Mostowski's Criterion:** The best representation of a metamathematical notion \( M \) of \( T \) is provided by a formula which registers as many as possible of the intuitive truths concerning \( M \) as theorems of \( T \).

First we remark that Mostowski's intention will be that any formula which expresses a metamathematical notion \( M \) of \( T \) is the best presentation of it. Following this criterion, it seems to us *prima facie* that a provability predicate \( \Pr_T(x) \) satisfying the derivability conditions can register a greater number of the intuitive truths\(^{19} \) concerning \( M \) as theorems of \( T \) than a provability predicate not satisfying the derivability conditions does. If this argument were valid, a provability predicate \( \Pr_T(x) \) satisfying the derivability conditions would be better than a provability predicate not satisfying the derivability conditions. As we noted, it is plausible that a provability predicate which means the provability in \( T \) must be the best representation of the provability in \( T \). Then we can conclude *prima facie* that a provability predicate

---

\(^{18}\) For a background of Mostowski’s argument, see [3, pp.101–103]. We are greatly indebted to his book in reconstructing Mostowski’s, and Kreisel’s arguments.

\(^{19}\) Here, again, the notion of truth does not need to be formal one (for example the provability in \( T \)) for an advocate of the second argument (if there is).
Pr$_T(x)$ which expresses the provability in $T$ must satisfy the derivability conditions. We remark that this second argument can be regarded as a modified version of the first argument. The underlying idea is that, because a provability predicate Pr$_T(x)$ satisfying DC1 cannot register every truth concerning $M$ as we have seen, we must modify the idea underlying the first argument. Then we might modify the idea of the first argument by replacing the phrase “every truth” by “the maximal number of the intuitive truths”.

We present a crucial criticism\(^{20}\) against this argument. As previously mentioned, a provability predicate satisfying the derivability conditions cannot register even the most basic intuitive truth regarding the unprovability in $T$. Moreover a nonstandard provability predicate\(^{21}\) can register a portion of intuitive truths concerning the unprovability in $T$ (for example the consistency of $T$) as theorems of $T$\(^{22}\).

Let Pr$_T(x)$ be the provability predicate defined parallel to the definition of the syntax of $T$. This Pr$_T(x)$ is of the form $\exists y$Pr$_T(y, x)$ where Pr$_T(y, x)$ strongly represents the recursive predicate Pr$_T(y, x)$. Then we define Pr$_T^M(y, x)$ as below.

$$\text{Pr}_T^M(y, x) = \text{Pr}_T(y, x) \land \neg\text{Pr}_T(y, \neg \neg 0 = 1).$$

It can be shown that Pr$_T^M(y, x)$ strongly represents the recursive predicate Pr$_T(y, x)$. Thus we can prove the first incompleteness theorem using this Pr$_T^M(y, x)$. This Pr$_T^M(y, x)$ present prima facie a counterexample to the second incompleteness theorem. Using Pr$_T^M(y, x)$, we define Pr$_T^M(x)$ as $\exists y$Pr$_T^M(y, x)$. We define Con$_T^M$ as

$$\text{Con}_T^M \equiv \neg\text{Pr}_T^M(\neg \neg 0 = 1).$$

Then it is easy see that Con$_T^M$ is equivalent to

$$\forall y (\text{Pr}_T(y, \neg \neg 0 = 1) \land \neg\text{Pr}_T(y, \neg \neg 0 = 1)).$$

Therefore Con$_T^M$ is provable in $T$ even if $T$ is consistent, and then Pr$_T^M(x)$ does not satisfy at least one of the derivability conditions by the second incompleteness theorem. Since Pr$_T^M(y, x)$ strongly represents the recursive predicate Pr$_T(y, x)$, Pr$_T^M(x)$ can register a countably infinite number of many truths about the provability in $T$ as Pr$_T(x)$ does, which clearly expresses the provability in $T$.

If Mostowski’s criterion were true, then it would hold that a provability predicate satisfying the derivability conditions can register a greater number of the intuitive

\(^{20}\) Detlefsen presented his counterargument in [3, pp.103–113]. But it seems to us that our counterargument is persuasive enough to defeat the second argument by itself.

\(^{21}\) Mostowski constructed a nonstandard provability predicate in [17]. There are several types of nonstandard provability predicates. For other types, see [5, 12].

\(^{22}\) Note that we can present this argument based on other types of nonstandard provability predicates, but we choose Mostowski’s one for the sake of simplicity.
truths concerning $M$ as theorems of $T$ than a provability predicate not satisfying the derivability conditions does. But it is clear from the explanation above that the number of properties which a provability predicate satisfying the derivability conditions can register and the number of properties which a provability predicate not satisfying the derivability conditions can register are same, namely $\mathfrak{N}_0$.

In addition to this criticism, we point out a similar deficiency which we did in 3.1. The second argument is very doubtful unless an advocate of the second argument can show that the provability predicate $\text{Pr}_T(x)$ (defined parallel to the definition of the syntax of $T$) must register a greater number of the intuitive truths concerning $M$ as theorems of $T$ than a provability predicate not satisfying the derivability conditions does. But we have sufficient evidence which defeats their hope. Thus to justify DCT using Mostowski’s criterion seems to be impossible.

### 3.3. The Third Argument

The third argument is based on a criterion which traces back to Kreisel. According to [3, pp.113–116], Kreisel proposed the following criterion in [15, p.118], and [16, pp.34–35]\(^{23}\).

**Kreisel’s Criterion**: If a provability predicate $\text{Pr}_T(x)$ which expresses the provability in a formal system of $T$ which can be regarded as capturing a main part of our mathematical practice (or activity), then $\text{Pr}_T(x)$ must satisfy DC1, the formalized $\Sigma^0_1$ completeness theorem, and DC3.

One feature of this argument is that its main aim lies in justifying the formalized $\Sigma^0_1$ completeness theorem. There are two motivations behind this criterion: (1) (at least in the usual proof of the second incompleteness theorem) DC2 is a consequence of the formalized $\Sigma^0_1$ completeness theorem, and (2) in contrast to DC2, it is easy to find a counterpart of the formalized $\Sigma^0_1$ completeness theorem in our informal mathematical practice.

This criterion can be divided into two parts. The first part is called a substantive element, and the second part is called a procedural element. Let us see why $\text{Pr}_T(x)$ which expresses the provability in $T$ must satisfy the formalized $\Sigma^0_1$ completeness theorem by this criterion. The formalized $\Sigma^0_1$ completeness theorem is stated as follows\(^{24}\).

(FSC) For any $\Sigma^0_1$ sentence $\phi$ of $T$, $\vdash_T \phi \rightarrow \text{Pr}_T(\overline{\phi})$ holds.

First a substantive element of the formalized $\Sigma^0_1$ completeness theorem claims

\(^{23}\) As we noted earlier, we owe a lot to Detlefsen’s book in reconstructing Kreisel’s arguments.

\(^{24}\) This type of a justification of why $\text{Pr}_T(x)$ must satisfy the formalized $\Sigma^0_1$ completeness theorem can be found in Kreisel’s papers. In particular see [16, p.34].
that $T$ must capture informal elementary arithmetic: the unformalized $\Sigma^0_1$ completeness theorem, which states that any true $\Sigma^0_1$ sentence $\phi$ is provable in $T$. Second, a procedural element of this theorem claims that (1) we must know that $T$ captures indeed informal elementary arithmetic, and (2) the fact that we know the substantive element is established by a metamathematical proof of the substantive element element. Therefore $\Pr_T(x)$ which expresses the provability in $T$ must satisfy the formalized $\Sigma^0_1$ completeness theorem by this criterion.

A justification of DC3 is similar. The substantive element of DC3 claims that the formal system $T$ must capture a main part of our logical practice, especially modus ponens. This is because $T$ must capture a main part of our mathematical practice and modus ponens is frequently used in our practice. The procedural element of DC3 claims that we must know the substantive element, and the fact that we know the substantive element is established by proving a metamathematical theorem regarding the substantive element, that is, by proving DC3.

It seems to us that there is no problem in the substantive element of this argument, that is, $T$ must capture a main part of our informal mathematical practice. One might raise the question of why this fact must be established by proving a metamathematical theorem. Kreisel does not indicate an answer to this question. Probably a ‘metamathematical theorem’ means a (pure) proof-theoretical or computational theorem in this context. Recall that we need the notion of the standard model in order to prove the $\Sigma^0_1$ completeness theorem, or for even stating this theorem. Thus if there is a priority in proving a metamathematical theorem by a proof-theoretical or computational method, then to prove a substantive element by such a method provides a computational (or epistemological) basis\(^{25}\) for the substantive element.

There is a gap in this argument. According to it, a justification of the derivability conditions proceeds in the following way.

1. Take any provability predicate $\Pr_T(x)$ which expresses the provability in $T$.
2. By Kreisel’s criterion, $\Pr_T(x)$ satisfies DC1, DC3, and the formalized $\Sigma^0_1$ completeness theorem.
3. Since we can prove DC2 from the formalized $\Sigma^0_1$ completeness theorem, $\Pr_T(x)$ satisfies DC2.
4. Then this provability predicate $\Pr_T(x)$ satisfies the derivability conditions.
5. Thus if a provability predicate $\Pr_T(x)$ expresses the provability in $T$, this $\Pr_T(x)$ satisfies the derivability conditions.

However, we cannot go to (3) from (2). We need an important assumption to go to (3) from (2). To bring out this assumption, let us examine a proof of DC2 from the formalized $\Sigma^0_1$ completeness theorem. Assume that a provability predicate $\Pr_T(x)$

\(^{25}\) This situation may be analogous to Hilbert’s program.
satisfies the formalized $\Sigma_1^0$ completeness theorem. Therefore, for any $\Sigma_1^0$ sentence $\phi$ of $T$,

$$\vdash_T \phi \rightarrow \Pr_T([\phi]).$$

To derive DC2 from this, we take $\Pr_T([\psi])$ as $\phi$. Then,

$$\vdash_T \Pr_T([\psi]) \rightarrow \Pr_T([\Pr_T([\psi])]).$$

In this step we used the fact $\Pr_T([\psi])$ is a $\Sigma_1^0$ sentence. This requires that $\Pr_T(x)$ is a $\Sigma_1^0$ formula. Since $\Pr_T(x)$ is any formula which expresses the provability in $T$, the assumption can be stated as follows.

**Assumption:** For any provability predicate $\Pr_T(x)$ which expresses provability in $T$, this $\Pr_T(x)$ must be a $\Sigma_1^0$ formula.

If it is impossible to justify this assumption, we cannot pass from (2) to (3). Of course the provability predicate $\Pr_T(x)$ which is used in the usual proof of the second incompleteness theorem is a $\Sigma_1^0$ formula. But the assumption requires much more. The fact that the provability predicate $\Pr_T(x)$ which is used in the usual proof of the second incompleteness theorem is a $\Sigma_1^0$ formula does not show that any provability predicate which expresses the provability in $T$ is a $\Sigma_1^0$ formula. Thus we cannot conclude that any provability predicate $\Pr_T(x)$ which expresses the provability in $T$ must be a $\Sigma_1^0$ formula. This is because it might be the case that there is some provability predicate which expresses the provability in $T$ but which is, nonetheless, constructed in some way different from the way in which the provability predicate used in the usual proof of the second incompleteness theorem is constructed, and is not a $\Sigma_1^0$ formula.

In what follows, we examine an argument which *prima facie* fills the gap. This argument is very plausible, and it appeals to the conception of formalization. Nevertheless it remains to be insufficient as we see below. Recall that we have considered the relationship between Gödel’s second incompleteness theorem and Hilbert’s program in the standard setting. At least in the context of Hilbert’s program in the standard setting, a formal theory, say $T$, satisfies the condition that the relation $\Pr_T(y, x)$ is recursive.\(^{27}\)

Of course this fact does not exclude the possibility that there is a formal theory $T$ which deserves to be considered in the context of the usual Hilbert’s program, and the relation $\Pr_T(y, x)$ is not recursive. But it seems to us that such a theory does not deserve to be called a formal theory for the following reasons. Recall the Church-Turing thesis, which asserts that every computable function is a recursive function and vice versa. An $n$-place predicate is recursive (decidable) if its characteristic function is

\(^{26}\) Of course, $\psi$ is any formula of $T$.

\(^{27}\) For example, $PA$ or $PA^2$. 

—24—
recursive (decidable). Thus, by the Church-Turing thesis, every $n$-place computable predicate is a recursive one, and vice versa. Therefore, we identify these notions in what follows.

The aim of formalization which provides a resulting formal system is to systematically codify our knowledge about a certain field. In order to fulfill this aim, such a formal theory consists of basic propositions which does not need any proofs (axioms), and rules from (valid) propositions to (valid) propositions (inference rules). Thus if it is impossible to decide whether any given formula of the formal system is an axiom of the theory, then it is useless. Likewise, it must be decidable whether any given finite sequences of formulas of the formal system is a proof in the theory. From this it follows that any theory which deserves to be called “formal” must satisfy the condition that the relation $Prf_T(y,x)$ is decidable (recursive), and this condition is essential to the conception of a formal theory.

Now, we can present the argument whose aim is to justify the assumption which states that any provability predicate $Prf_T(x)$ expressing the provability in $T$ is a $\Sigma^0_1$ formula as follows:

1. If a formula $\phi$ expresses some relation $R$, then we expect that $\phi$ must capture essential features of $R$ (assumption).
2. If a proof predicate $Prf_T(y,x)$ expresses the proof relation $Prf_T(y,x)$ in $T$, then this predicate must capture essential features of the proof relation $Prf_T(y,x)$ (from 1).
3. The feature that the relation $Prf_T(y,x)$ is decidable is essential for $T$ to be called “formal” (assumption).
4. Thus if a proof predicate $Prf_T(y,x)$ expresses the proof relation $Prf_T(y,x)$ in $T$, then this predicate must capture the feature that $Prf_T(y,x)$ is decidable (from 2 and 3).
5. The fact that a proof predicate $Prf_T(y,x)$ captures the feature that $Prf_T(y,x)$ is decidable must be shown by the fact that $Prf_T(y,x)$ is a $\Delta^0_1$ formula (assumption).
6. Therefore if a proof predicate $Prf_T(y,x)$ expresses the proof relation $Prf_T(y,x)$

---

28 Perhaps a careful reader may notice the distinction between axiomatization and formalization. But, as far as we follow the standard terminology, formalization presuppose axiomatization.

29 We can give examples of theories which violate this condition. Feferman gave an example of the nonstandard provability predicate in [5], and he proved that the formula prima facie expressing the consistency, which is built up from this provability predicate, is provable even if the formal theory is consistent. As analyzed in [1], this provability predicate is defined relative to a given formal system (for example $PA$), the defined theory is not clearly recursive (not even recursively enumerable). Note that formal theories which lie behind Rosser type, and Mostowski type nonstandard provability predicates are recursive.
in $T$, then this predicate is a $\Delta^0_1$ formula (4 and 5).
7. If a provability predicate $\text{Prf}_T(x)$ expresses the provability in $T$, then $\text{Prf}_T(x)$ is of the form $\exists y \text{Prf}_T(y, x)$, and $\text{Prf}_T(y, x)$ expresses the proof relation $\text{Prf}_T(y, x)$ in $T$ (assumption).
8. If a provability predicate $\text{Prf}_T(x)$ expresses the provability in $T$, then $\text{Prf}_T(x)$ is of the form $\exists y \text{Prf}_T(y, x)$, and $\text{Prf}_T(y, x)$ is a $\Delta^0_1$ formula (from 6 and 7).
9. Therefore any provability predicate $\text{Prf}_T(x)$ which expresses the provability in $T$ must be a $\Sigma^0_1$ formula (from 8).

This argument seems to us very plausible. However, it involves the same problem as before. Evidently, 1 and 7 are valid assumptions. By the previous argument, 3 is also valid. Thus the problematic assumption is 5. Of course, assuming the notion of the standard model and the standard interpretation based on it, it is clear that any $\Delta^0_1$ formula is decidable, that is, we can check in finite steps that a given $\Delta^0_1$ formula is true or false. But this is not what we need. We know that for any decidable (recursive) relation there is a $\Delta^0_1$ formula of $T$ which strongly represents it if $T$ contains at least $Q$. But, again, this is much weaker than what we need. Hence, from the fact that for any decidable (recursive) relation there is a $\Delta^0_1$ formula of $T$ which strongly represents it, we cannot conclude that any proof predicate $\text{Prf}_T(y, x)$ which captures the feature that $\text{Prf}_T(y, x)$ is decidable must be a $\Delta^0_1$ formula. Therefore, again we face the same kind of difficulty as previously encountered.

Recall the aim of introducing the assumption stated in step 5. What we need is the proposition asserted in step 6, that is,

6: If a proof predicate $\text{Prf}_T(y, x)$ expresses the proof relation $\text{Prf}_T(y, x)$ in $T$, then this predicate is a $\Delta^0_1$ formula.

As we noted, if we were justified in this proposition, the argument for the claim that Gödel’s second incompleteness theorem defeated Hilbert’s program would be valid.

But, to this point, we have no way of justifying either this proposition directly or the assumption stated in step 5. It seems to us that any proof predicate $\text{Prf}_T(y, x)$ which expresses the provability in $T$ is extensionally equivalent to the proof predicate which is used in the usual proof of the second incompleteness theorem. This is because such a $\text{Prf}_T(y, x)$ at least must strongly represent the proof relation $\text{Prf}_T(y, x)$ in $T$. Consequently one might be tempted to use the extensional identity with a proof predicate $\text{Prf}_T(y, x)$ expressing the proof relation $\text{Prf}_T(y, x)$ in order to justify the proposition directly. Let us reconstruct such an argument in detail. Take a proof predicate $\text{Prf}_T(y, x)$ which expresses the proof relation $\text{Prf}_T(y, x)$ in $T$. Then, at least, $\text{Prf}_T(y, x)$ must strongly represent the proof relation $\text{Prf}_T(y, x)$ in $T$. That is, the following must be hold for any $n, m \in \omega$. 

— 26 —

NII-Electronic Library Service
If $\text{Prf}_T(n, m)$, then $\vdash_T \text{Prf}_T(n, m)$.
If $\neg\text{Prf}_T(n, m)$, then $\vdash_T \neg\text{Prf}_T(n, m)$.

Let $\text{Prf}^S_T(y, x)$ be the standard (or usual) proof predicate used in the usual proof of the second incompleteness theorem. Then $\text{Prf}^S_T(y, x)$ strongly represents the proof relation $\text{Prf}_T(y, x)$ in $T$, and $\text{Prf}^S_T(y, x)$ expresses the proof relation $\text{Prf}_T(y, x)$. From the former, it is easily verified that $\text{Prf}_T(n, m)$ is extensionally equivalent (in $T$) to $\text{Prf}^S_T(n, m)$ for any given $n, m \in \omega$. Thus one might be tempted to say that $\text{Prf}_T(y, x)$ is a $\Delta^0_1$ formula.

But it is possible to identify a deficiency in this argument. Remark that we can prove only the numeralwise extensional equivalence, namely we can prove only $\vdash_T \text{Prf}_T(n, m) \iff \text{Prf}^S_T(n, m)$ for any given $n, m \in \omega$. From this we cannot conclude that $\vdash_T \text{Prf}_T(n, m) \iff \text{Prf}^S_T(n, m)$. Therefore we cannot derive that $\text{Prf}_T(y, x)$ is a $\Delta^0_1$ formula immediately. Moreover, this problem is not simple. Recall the previous argument in 3.2. Take $\text{Prf}^M_T(y, x)(\equiv \text{Prf}_T(y, x) \land \neg\text{Prf}_T(y, x) = \overline{1})$. Then it is easy to see $\text{Prf}^M_T(y, x)$ is extensionally equivalent to the standard proof predicate $\text{Prf}^S_T(y, x)$, which is used in the usual proof of Gödel’s second incompleteness theorem, namely we can prove that $\vdash_T \text{Prf}^M_T(n, m) \iff \text{Prf}^S_T(n, m)$ for any $n, m \in \omega$. But, by Gödel’s second incompleteness theorem, we have $\forall T. \text{Prf}^M_T(y, x) \iff \text{Prf}^S_T(y, x)$, equivalently, $\forall T. \forall x(\text{Prf}_T(y, x) \iff \text{Prf}^S_T(y, x))$\textsuperscript{30}. These facts show that the problem is very delicate.

4. Conclusion

Let us make final remarks. First no counterexample to the proposition 6 in the previous section has been found. That is, the existence of a proof predicate $\text{Prf}_T(y, x)$ which expresses the proof relation $\text{Prf}_T(y, x)$ in $T$ but is not a $\Delta^0_1$ formula. Moreover, we do not know any counterexample that will refute the assumption stated in 5, namely the existence of a proof predicate $\text{Prf}_T(y, x)$ which captures the feature that $\text{Prf}_T(y, x)$ is decidable, but is not a $\Delta^0_1$ formula.

Second, probably, a plausible way to justify the proposition asserted in step 6 is one that entails a justification of the assumption stated in step 5. The latter states that any proof predicate $\text{Prf}_T(y, x)$ which captures the feature that $\text{Prf}_T(y, x)$ is decidable is a $\Delta^0_1$. The reason for this is that the feature that the relation $\text{Prf}_T(y, x)$ is decidable is essential for $T$ to be called formal. Moreover, the assumption stated in the step 5 seems to be plausible because a typical $\text{Prf}_T(y, x)$ is defined parallel to the definition of the syntax of $T$, and is a $\Delta^0_1$ formula. Sometimes we are tempted to claim that such a proof predicate $\text{Prf}_T(y, x)$ is “natural” or “standard”, but the problem then becomes one of characterizing naturalness, or standard.

\textsuperscript{30} Of course, $\text{Prf}^M_T(y, x)$ is a $\Delta^0_1$ formula if $\text{Prf}_T(y, x)$ which occurs in it is a $\Delta^0_1$ formula.
Therefore our conclusion is that we cannot say that Gödel’s second incompleteness theorem together with a supplemented Kreisel’s argument defeats Hilbert’s program\(^{31}\). We need a more delicate analysis of a proof predicate which assures that any proof predicate \(\text{Pr}_T(y, x)\) which captures the feature that the \(\text{Pr}_T(y, x)\) is decidable is a \(\Delta^0_1\) formula. Presumably, such an analysis must reflect internal structures (or intensional character) of a proof predicate \(\text{Pr}_T(y, x)\)^{32}, but no attempt to elaborate such an analysis has been made.

References


\(^{31}\) Moreover, we concede that, in order to argue this problem more fully, we must consider Detlefsen’s arguments against a justification of the derivability conditions. See [3, 4]. In particular we must consider Detlefsen’s argument whose aim is to show that the strategy aimed at justification of the derivability conditions is mistaken.

\(^{32}\) Kreisel has suggested such a theory about a provability predicate \(\text{Pr}_T(x)\) expressing the provability relation \(\text{Pr}_T(x)\) [12, p.154], [15, pp.117–118]. But no attempt to elaborate such a theory has been made, and it is open to question whether or not such a theory provides some progress on the present problem.
On a Relationship between Gödel's Second Incompleteness Theorem and Hilbert's Program


