arXiv:2303.18099v1 [math.HO] 31 Mar 2023

讲授哥¨德尔不完备性定理

Gilles Dowek*

1 引言

逻辑的基本概念——谓词逻辑、皮亚诺算术、不完备性定理等——长期以来一直是高级课题。近几十年来，这些内容在哲学、数学和计算机科学系中得到更广泛的讲授，授课对象包括研究生和本科生。如今，许多教科书都介绍这些概念，尤其是不完备性定理。

经过数十年的讲授，我们如今可以退一步，分析设计这类课程时面临的选择。在本文中，我们尝试分析讲授不完备性定理时所面临的选择。具体而言，我们试图论证以下几点。

- 不完备性定理内容过于丰富，无法仅用一门课程讲授。对于从未接触过谓词逻辑基础的学生，不可能在短短几周内讲到第二不完备性定理和洛布定理；这就如同不可能在短短几周内向从未接触过函数和复数概念的学生讲授解析函数的概念一样。

  因此，不完备性定理必须在不同层次上多次讲授，例如先在大学三年级的基础课程中讲授，随后在四年级的高级课程中讲授。这些课程的目标和重点各不相同。

- 当不完备性定理被孤立地讲授时，它们往往被视为一个“应许之地”，而引入可计算性、表示、归约、对角化等概念，则是为了证明这些定理。

  相反，我们主张应当分别引入并讲授这些概念，必要时甚至可以安排在不同的课程中。这样一来，关于不完备性定理的章节就能保持精简，并聚焦于不完备性的具体内容。

*Inria 与巴黎萨克雷高等师范学院，61，普雷西登特·威尔逊大道，94235 卡尚 Cedex，法国，gilles.dowek@ens-paris-saclay.fr。

- 例如，若为了让学生能够亲自掌握编号这一概念而将证明讲得过于具体，甚至明确给出编号，往往会让证明充斥着无关且琐碎的细节。

  证明需要足够抽象，让学生能够专注于有意义的要点。现有一些工具可以让证明足够抽象，例如抽象语法、衔接、一般不动点定理、可证性逻辑记法等，均可加以运用。

2 哪些理论?

不完备性定理并不适用于所有理论。有些理论过于薄弱，因此不适用，例如包含一个二元谓词符号 $=$和一条公理 $\forall x \forall y\ (x = y)$ 的理论；该理论是完备的。有些理论则过于强大，因此不适用，例如不一致理论；它们同样是完备的。

2.1 足够强

尽管哥德尔最初的证明适用于《数学原理》，但要证明不完备性定理，一个自然的选择是皮亚诺算术，它已经能够表示可计算函数。

然而，对于第一不完备定理，传统上会考虑一种弱得多的理论：罗宾逊算术。它本质上是皮亚诺算术去掉归纳公理，再加上归纳法的若干推论。这种推广确实很有意思，但在讲授不完备性定理时，尤其是在基础课程中，不应将其视为必不可少，因为它可能会使学生偏离该定理证明中的关键要点。

此外，证明希尔伯特–伯奈斯引理，进而证明第二不完备性定理，似乎需要一个比罗宾逊算术更强的理论，例如皮亚诺算术。

2.2 足够弱

学生们发现算术的不完备性时，一个自然而然但错误的想法是：遗漏了某个公理，而加入这个公理就能使理论完备。

因此，必须指出，加入有限个公理不可能使算术完备。正确的表述方式是证明本质不完备性，也就是说，任何满足某些性质的算术扩张都是不完备的。第一个条件可以表述为：公理集合是可判定的、证明检查是可判定的，或者定理集合是半可判定的这一事实。第二个条件是理论的一致性。

然而，仅假设一致性就需要证明 G¨odel-Rosser 定理，从而引入了一些不必要的复杂性。另一种选择是假设 ω-一致性，

也就是说，每当形如 $\exists x\ A$ 的命题可证时，都存在一个自然数 n，使得 $\neg(\underline{n}/x)$ A 不可证。但这一性质同样引入了一些不必要的复杂性。

一个更简单的选择是假设一个更强的前提：该理论具有标准模型。这样选择的一个优点是，它迫使我们引入标准模型这一概念，为不完备性推论做好准备：非标准模型的存在性。另一个优点是，拥有一个标准模型 $\mathcal{M}$ 使我们能够如将要看到的那样，轻松证明弱表示定理。

正如许多定理一样，在基础课程中，我们不应试图采用最弱的前提，而可以将对该定理最一般形式的讨论留给高级课程。

3 哪种语言?

语言的选择是一个棘手的问题。为了能够为每个自然数 n 关联一个项 $\underline{n}$，我们需要假设，表达该理论所使用的语言包含足够多的函数符号，以表示自然数。这使得不完备性定理无法应用于集合论等没有函数符号的理论。

另一种方案是使用这样一种理论：在其中可以构造带有自由变量的命题，用来刻画自然数 0、后继关系、加法、乘法等，并且某些命题（本质上是罗宾逊算术的公理）是可证明的。这样，对于每个自然数 p，都可以构造一个命题 N_p 来刻画数字 p，并写成 $\forall x\ (N_p[x] \Rightarrow A[x])$ 或 $\exists x\ (N_p[x] \wedge A[x])$，而不是 $A[\underline{p}]$。

这是在 [4] 中作出的选择。事后看来，这有些繁重，其实可以留作练习。

4 编号

不完备性定理证明的第一步通常是引入编号这一概念。对学生而言，这一概念往往既令人惊讶又显得简单。令人惊讶，是因为理论、算法、计算机等会操作各种数据类型，却并不为它们编号。说它简单，是因为如今人人都知道，文本、图像、声音等最终都会被编码为数字，也就是数。

通常，命题、证明、程序、图灵机等彼此独立地进行编号。这会导致引入许多定义。我们主张，应将编号这一概念定义得足够一般，使得命题、证明等的编号都只是这一一般定义的具体实例。

另一方面，编号这一概念也不能定义得过于宽泛，因为将一个编号与一个不可计算函数复合，会得到另一个函数，而这个函数不应被视为编号 [2]。

在 [4], 中，我们提议将编号限制于可分节集合，采用语言学中常见的 [6], 衔接概念。如果一个集合是有限的，就称其为 0-可分节的。如果一个集合是有限树的集合，就称其为（$n+1$）-可分节的

这些有限树的标记来自一个 n-可分节集合。因此，包含某个理论的连接词、量词、函数符号和谓词符号的集合是 0-可分节的；包含这些符号和变量的集合是 1-可分节的；项和命题的集合是 2-可分节的；相继式的集合是 3-可分节的；证明的集合是 4-可分节的，依此类推。更一般地说，我们需要进行编号的所有对象都是某个可分节集合的元素。

使用从 $\mathbb{N}^2$ 到 $\mathbb{N} \setminus \{0\}$的康托尔双射；

$$n; p = (n + p)(n + p + 1)/2 + n + 1$$

我们可以将任意一个由（$n+1$）衔接集合构成的树 f（t_1, ..., t_p）编号，其中根节点标记为 f，其直接子树为 t_1、...、 t_p：

$$\ulcorner f(t_1, ..., t_p)\urcorner = \ulcorner f\urcorner; (\ulcorner t_1\urcorner; (...; (\ulcorner t_p\urcorner; 0)...))$$

其中，第一个 $\ulcorner.\urcorner$是包含标签 f 的 n衔接集合的编号，其余各项则是通过归纳法定义的当前函数。这样，编号只取决于我们最初所使用的 0-衔接集合中各元素的编号。并且很容易证明，由于该集合是有限的，改变这一编号并不会改变从一个衔接集合到另一个衔接集合的可计算函数集合。

这是抽象机制的一个例子，它可以避免任意定义和繁琐的重复。当然，正如所有一般性概念一样，它必须通过具体例子加以说明；但在定义和证明中，这些具体例子不应取代一般概念。

使用衔接这一概念也迫使我们将所有希望进行编号的对象——命题、证明、程序等——定义为树。这意味着，这些对象的抽象语法，而不是它们的具体语法，得到了强调。我们认为这是件好事；括号、唯一可读性、前缀、中缀和后缀运算符、优先级等概念并非逻辑语言所特有，因此不应在逻辑课程中讲授，而应在语言理论课程中讨论。

最后，当一个命题 A 被编号为$\ulcorner A\urcorner$时，这个编号 n 通常用于构造项 $S^n(0)$，记作 $\underline{n}$。因此，证明中的许多表达式都具有 $\underline{\ulcorner A\urcorner}$的形式。尽管函数 $\ulcorner.\urcorner$ 和 . 有时会单独使用，但为 $\underline{\ulcorner A\urcorner}$ 引入记号“A”可以使证明更加清晰。事实上，这个由函数 $\ulcorner.\urcorner$和 . 复合而成的映射，将算术命题映射为算术项，正是齐性反映这一概念。

5 哥德尔的 β 函数与可计算函数的定义

不完备性定理证明中的第二步通常是将每个可计算函数与一个命题对应起来。

事实上，我们并不是为每个可计算函数都关联一个命题，而是为可计算函数的每种构造关联一个命题。例如，二元零函数与一元零函数和二元投影的复合在外延上相等，但与这些构造关联的是不同的命题。

因此，我们可以首先引入可计算函数的构造这一概念，使得 Z^2 和 $\circ_1^2(Z^1, \pi_1^2)$ 是同一个函数的不同构造。这种构造是一棵树，其中标记着符号 Z^n ，表示元数为n的零函数；Succ 表示后继函数； μ^n 表示对元数为$n+1$的函数进行最小化；等等。它可以被称为“程序”，因为它是表达可计算函数的句法对象。事实上，它就是用规则名称 [4]标记的推导树，这些规则名称与可计算函数的归纳定义相关联，而该推导树则是函数可计算性的证明。

可计算函数的集合通常定义为包含以下函数的最小集合：

- 投影函数、
- 空函数、
- 以及后继函数

并由.....闭合

- 复合、
- 归纳定义，
- 以及最小化，

从而得到语言 π_i^n、 Z^n、后继函数、$\circ_i^n$、递归^n、 μ^n。

为每个程序指定一个命题是很直接的——$y = 0$ 对于程序 Z^n， $y = S(x_1)$ 对于程序 $Succ$，等等——但归纳定义除外，因为归纳定义需要使用 G¨odel 的 β 函数和中国剩余定理。

如果我们采用可计算函数集的另一种定义，增加三个函数：加法、乘法和序关系的特征函数，即函数 $\chi_{\leq}$，使得当 $n \leq p$ 时，$\chi_{\leq}(n, p) = 1$，否则 $\chi_{\leq}(n, p) = 0$，并省略归纳定义，就可以避免这一困难。

程序表示得到了简化。

- 我们将命题 $y = x_i$ 与 π_i^n 相关联。
- 我们将命题 $y = 0$ 与 Z^n 相关联。
- 我们将命题 $y = S(x_1)$ 与后继函数相关联。
- 我们将命题 $y = x_1 + x_2$ 与 $+$ 相关联。
- 我们将命题 $y = x_1 \times x_2$ 与 $\times$ 相关联。
- 我们将命题与 $\chi_{\leq}$相关联

$$(x_1 \leq x_2 \wedge y = 1) \vee (x_2 < x_1 \wedge y = 0)$$

其中，命题 $x \leq y$ 简写为 $\exists z(z + x = y)$，而 $x < y$ 简写为$S(x) \leq y$。

- 对于 $\circ_m^n(h, g_1, \ldots, g_m)$，我们关联命题$\exists w_1 \ldots \exists w_m$ ($B_1[x_1, \ldots, x_n, w_1] \wedge \ldots \wedge B_m[x_1, \ldots, x_n, w_m] \wedge C[w_1, \ldots, w_m, y]$)，其中 B_1、...、B_m 和 C分别表示程序 g_1、...、 g_m 和 h。

- 对于 $\mu^n(g)$，我们关联命题

$$\forall z \ (z < y \Rightarrow \exists w \ (B[x_1, \ldots, x_n, z, S(w)])) \wedge B[x_1, \ldots, x_n, y, 0]$$

其中， B 表示程序 g。

从根本上说，使这种表示概念变得简单的，是用于构造算术项的符号——变量、0、 S、 $+$和$\times$——与定义可计算函数集的八个子句中的五个之间的相似性：投影函数、空函数、后继函数、加法和乘法。

但是，当然，我们需要证明这两个可计算函数集定义之间的等价关系，而这需要使用 G¨odel 的 β 函数和中国剩余定理。证明这种等价关系甚至比直接使用 G¨odel 的 β 函数和中国剩余定理将程序表示为命题稍微困难一些，但它使证明更具模块化。当证明这种等价关系时，只使用从自然数映射到自然数的函数，并不提及命题的概念。这种等价关系也可以在讨论不完备性定理之前很久就得到证明。其动机也可以来自不完备性定理以外的其他目标：例如，它简化了其他定理的证明，例如将可计算函数表示为重写系统、 λ-演算的项、 [5], 图灵机等的表示定理。当可计算性课程与逻辑课程分开讲授时，这种等价关系当然应当在可计算性课程中讲授，而不应在逻辑课程中讲授。

甚至可以用这八条规则来定义可计算函数，而 G¨odel 的β 函数和中国剩余定理届时只需用于证明：可计算函数集在归纳定义下是封闭的。

6 表示定理的形式

6.1 弱表示定理

设 $\mathcal{T}$是一个理论，它的公理集可判定，是罗宾逊算术的扩展，并且具有一个标准模型 $\mathcal{M}$。

我们要证明：如果命题 A 表示程序 f，那么命题 $A[\underline{p_1}, ..., \underline{p_n}, \underline{q}]$ 表达的是程序 f在 p_1 处终止，....., p_n并返回 q；也就是说，当且仅当 $f(p_1, ..., p_n) = q$时，该命题在 $\mathcal{T}$中可证。

对程序的结构进行简单的归纳法可以证明：如果$f(p_1, ..., p_n) = q$，那么命题 $A[\underline{p_1}, ..., \underline{p_n}, \underline{q}]$ 可在 $\mathcal{T}$中证明。完备性定理表明，如果该命题可在 $\mathcal{T}$中证明，那么它在 $\mathcal{M}$中有效。并且，对程序 f 的结构进行简单的归纳法可以证明：如果该命题在 $\mathcal{M}$中有效，那么 $q = f(p_1, ..., p_n)$。

[1]原证明存在一些疏漏；修正后的证明（法语）可在线获取：http://www.lsv.fr/~dowek/Books/Lc/prop317.pdf

6.2 可证明性的不可判定性

同样地，我们可以证明，命题 $\exists y\ A[\underline{p_1}, ..., \underline{p_n}, y]$ 表达了程序 f 在 p_1、...、 p_n处终止，也就是说，当且仅当 f在 p_1、...、 p_n处终止时，该命题可在$\mathcal{T}$中证明。

将可计算函数 F把 f 和 p_1，...， p_n 映射到命题 $\exists y\ A[\underline{p_1}, ..., \underline{p_n}, y]$，从而将停机问题归约为在 $\mathcal{T}$ 中的可证明性。

如果 G是这样一个函数：当命题 A 可证明时将其映射为 1，否则映射为 0；那么函数 $G \circ F$ 将 f 和 p_1，...， p_n 映射为 1，当且仅当 f 在 p_1，...， p_n 处终止，否则映射为 0。利用可计算函数关于复合的闭包性质的逆否命题，由于 $G \circ F$ 不可计算而 F 可计算，所以 G 不可计算。因此， $\mathcal{T}$ 中可证明性的不可判定性，仅仅是停机问题的不可判定性和这个表示定理的一个直接推论。

6.3 更强表示定理

在不完备性定理的一些证明中，我们需要一个更强的定理来表达：如果 f 在 p_1，...， p_n 处终止，那么命题

$$\forall y\ (A[\underline{p_1}, ..., \underline{p_n}, y] \Leftrightarrow y = \underline{f(p_1, ..., p_n)})$$

是可证明的。对于这个定理，我们不需要理论 $\mathcal{T}$ 有标准模型，甚至不需要它是一致的。我们只需要它具有可判定的公理集，并且是罗宾逊算术的扩展。

请注意，这个表示定理比第 6.1 节的定理既更强，也更弱。它更强，是因为与 f 的等价关系被内化了，即通过语言中的一个等价关系来表达，并且该关系对所有 y都成立。但它也更弱，因为当 f 在 p_1、.....、 p_n处不终止时，它没有说明任何情况，而第 6.1 节的定理表明，在这种情况下，$A[\underline{p_1}, ..., \underline{p_n}, \underline{q}]$ 不可证明。

但是，假设理论 $\mathcal{T}$ 还一致，那么当函数在 p_1、.....、 p_n处终止时，弱表示定理就是强表示定理的推论，因为

$$A[\underline{p_1}, ..., \underline{p_n}, \underline{q}]$$

是可证明的，当且仅当 $\underline{q} = f(p_1, ..., p_n)$ 是可证明的，也就是说，当且仅当 $q = f(p_1, ..., p_n)$。

这个强表示定理的表述比弱表示定理的表述略显不自然，因为它在程序的参数与值之间引入了不对称性。参数是语言外部量化的封闭项 $\underline{p_1}$、.....、 $\underline{p_n}$ ，而值则是语言本身内部量化的变量 y 。

因此，如果强表示在八种情形中的七种情形下都可以直接证明，那么对于最小化 $\mu^n(g)$，证明过程就略显间接，在这种情况下，函数 g 的一个参数会成为函数 $\mu^n(g)$ 的值。我们需要证明

$$\forall y\ ((\forall z\ (z < y \Rightarrow \exists w\ B[\underline{p_1}, \ldots, \underline{p_n}, z, S(w)]) \wedge B[\underline{p_1}, \ldots, \underline{p_n}, y, 0]) \Leftrightarrow y = \underline{r})$$

假设 $\mu^n(g)$ 终止，并在 p_1、.....、 p_n 处取值 r 。证明

$$\forall y\ (y = \underline{r} \Rightarrow (\forall z\ (z < y \Rightarrow \exists w\ B[\underline{p_1}, \ldots, \underline{p_n}, z, S(w)]) \wedge B[\underline{p_1}, \ldots, \underline{p_n}, y, 0]))$$

这等价于

$$\forall z\ (z < \underline{r} \Rightarrow \exists w\ B[\underline{p_1}, \ldots, \underline{p_n}, z, S(w)]) \wedge B[\underline{p_1}, \ldots, \underline{p_n}, \underline{r}, 0]$$

这很容易，因为有界量化可以化约为有限合取。但要证明逆命题

$$\forall y\ ((\forall z\ (z < y \Rightarrow \exists w\ B[\underline{p_1}, \ldots, \underline{p_n}, z, S(w)]) \wedge B[\underline{p_1}, \ldots, \underline{p_n}, y, 0]) \Rightarrow y = \underline{r})$$

我们需要利用这样一个事实：由假设 $\forall z(z < y \Rightarrow \exists w\ B[\underline{p_1}, \ldots, \underline{p_n}, z, S(w)])$ 可以推出 $\underline{r} < y \Rightarrow \exists w\ B[\underline{p_1}, \ldots, \underline{p_n}, \underline{r}, S(w)]$ 因此 $\underline{r} < y \Rightarrow \exists w\ 0 = S(w)$ 并且$y \leq \underline{r}$，从而证明只需证明命题

$$\forall y\ (y \leq \underline{r} \Rightarrow (B[\underline{p_1}, \ldots, \underline{p_n}, y, 0] \Rightarrow y = \underline{r}))$$

化约为有限合取的命题。

正如我们将看到的，弱表示定理足以用于第一不完备定理的若干证明；在初级课程中，我们可以只讲这一条定理，而高级课程则需要用到强表示定理。

7 第一不完备定理的各种证明

第一不完备定理的证明分为两大类。

7.1 计算机科学家的证明

在第一类证明中，理论的不完备性被视为该理论中可证明性的不可判定性的结果。设 $\mathcal{T}$ 是一个具有可判定的公理集、是罗宾逊算术的扩展，并且具有标准模型 $\mathcal{M}$ 的理论。如我们所见， $\mathcal{T}$ 中的可证明性是不可判定的。

如果该理论是完备的，那么将 $\ulcorner A \urcorner$ 映射到满足 $\mathsf{proof}(x, \ulcorner A \urcorner) = 1$ 或 $\mathsf{proof}(x, \ulcorner \neg A \urcorner) = 1$ 的最小 x 的可计算函数将是全函数，并且当且仅当 A 是可证明的时，该函数会返回 A 的证明。其中，可计算函数 proof 在 $n = \ulcorner \pi \urcorner$、$p = \ulcorner A \urcorner$ 且 π 是 A 的证明时，将 n 和 p 映射为 1，否则映射为 0。因此，它将能够构造一个判定 $\mathcal{T}$ 中可证明性的算法。

这个定理甚至可以进一步抽象为如下事实：一个半可判定集，如果其补集也是半可判定的，那么它就是可判定的。

这个证明运用了理论计算机科学中的许多概念：可计算函数、归约和证明搜索。将 $\ulcorner A \urcorner$ 映射到满足 $\mathsf{proof}(x, \ulcorner A \urcorner) = 1$ 的最小 x 的函数，是一种生成并检验式证明搜索算法；而将 $\ulcorner A \urcorner$ 映射到满足 $\mathsf{proof}(x, \ulcorner A \urcorner) = 1$ 或 $\mathsf{proof}(x, \ulcorner \neg A \urcorner) = 1$ 的最小 x 的映射，则是一种类似的算法，可同时搜索 A 和 $\neg A$ 的证明。因此，这个证明可以称为计算机科学家的证明。

考虑一种理论：它具有可判定的公理集，是罗宾逊算术或皮亚诺算术的扩展，并且具有标准模型；证明

该理论的弱表示定理，推导出该理论中可证明性的不可判定性，并进一步将其不完备性作为推论得出，对于一门入门课程而言，或许就是介绍不完备性的充分目标。

7.2 Thequick 数学家的证明

第二类证明有效地构造了一个命题 G，使得G 和 $\neg G$都不可证明。

它使可计算函数这一概念所处的地位不那么核心。通常，只需表示少数几个函数，例如证明函数、替换函数（当 $n = \ulcorner A \urcorner$ 和 $m = \ulcorner (\underline{p}/x)A \urcorner$ 时，将 n 和 p 映射到 m），以及否定函数（当 $n = \ulcorner A \urcorner$ 和 $m = \ulcorner \neg A \urcorner$ 时，将 n 映射到 m），就足够了。

这个证明的一种简单表述，即使在初等课程中也可以作为练习，是只考虑一个函数，也就是证明函数和替换函数的混合体：当 $n = \ulcorner \pi \urcorner$、$p = \ulcorner A \urcorner$ 和 π 是 $(\underline{q}/x)A$ 的证明时，将 n、p 和 q 映射为 1。

设 $\mathcal{T}$ 是一个理论，它具有可判定的公理集，是罗宾逊算术的扩展，并且具有一个标准模型 $\mathcal{M}$。

称 F 为表示函数 f 的命题，T 为对角命题，$\forall x\ \neg F[x, w, w, \underline{1}]$，并称 G 为命题 $T[\text{`}T\text{'}],$。容易证明，在一个公理集可判定、包含罗宾逊算术且具有标准模型的理论中，G 和 $\neg G$ 都不可证。

如果 G 可证，那么 $\forall x\ \neg F[x, \text{`}T\text{'}, \text{`}T\text{'}, \underline{1}]$ 可证，因此对所有 n，$f(n, \ulcorner T \urcorner, \ulcorner T \urcorner) = 0$，因此 $T[\text{`}T\text{'}],$，也就是 G，不可证。矛盾。如果 $\neg G$ 可证，那么存在一个自然数 n，使得 $f(n, \ulcorner T \urcorner, \ulcorner T \urcorner) = 1$，因此 G 可证，所以 $\bot$ 可证。矛盾。

7.3 严谨数学家的证明

这个简短的证明仍然有些神秘，因为证明函数和替换函数混合在这个函数 f 中，而命题 G的含义——它与说谎者悖论、对角论证以及自指的关系——并不明确。

一个更严谨的证明，同时也能更好地为第二个不完备定理的证明做准备，是使用证明函数和替换函数 s 来分解这个函数 f。

然后，我们可以引入一个可证明性命题

$$Bew = \exists z\ Proof[z, x, \underline{1}]$$

并引入受可证性逻辑启发的记号 $\Box A$，表示 Bew[‘A’]。

很容易证明必然化引理：如果 A 是可证明的，那么 $\Box A$ 是可证明的。事实上，如果 A 有一个证明 π，那么根据表示定理，Proof[‘π’, ‘A’, $\underline{1}$]。因此，$\exists z$ Proof[z, ‘A’, $\underline{1}$],，也就是 Bew[‘A’],，也就是 $\Box A$，是可证明的。

这个引理的逆命题通常并不成立：一个理论可以证明 $\Box \bot$，却无法证明 $\bot$；但如果该理论是 ω一致的，则这个逆命题成立。事实上，由$\Box A$，即 $\exists z$ 证明$[z, \text{`}A\text{'}, \underline{1}]$，我们可以推导出存在一个 n，使得

¬证明[$\underline{n}$,‘A’, $\underline{1}$] 不可证，因此根据强表示定理 ¬$\underline{1}$ =证明(n,⌜A⌝) 不可证，所以证明(n,⌜A⌝) $\neq 0$，证明(n,⌜A⌝) $= 1$，并且 A 可证。

设 $\mathcal{T}$ 是一个公理集可判定、扩张了罗宾逊算术且 ω-一致的理论。

利用强表示定理，我们现在可以构造说谎者命题 G，使得 $G \Leftrightarrow \neg\Box G$ 可证。为此，我们引入一个命题 D，它表达了$A[\underline{p}]$ 不可证，

$$\exists z\ (\neg Bew[z] \wedge S[x_1, x_2, z])$$

一个表达了 A[‘A’] 不可证的命题

$$D[w, w]$$

并且 G

$$E[\text{‘}E\text{’}]$$

也就是说

$$\exists z\ (\neg Bew[z] \wedge S[\text{‘}E\text{’}, \text{‘}E\text{’}, z])$$

那么 G 可证地等价于

$$\exists z\ (\neg Bew[z] \wedge z = \text{‘}E[\text{‘}E\text{’}]\text{’})$$

也就是说

$$\neg Bew[\text{‘}G\text{’}]$$

因此，命题 G是命题 ¬Bew$[x]$ 的不动点定理。这个定理可以推广如下 [7]。我们考虑任意一个包含变量 x的命题 C，并证明存在一个命题 G，使得 $G \Leftrightarrow C[\text{‘}G\text{’}]$ 可证。我们不将 D 定义为 $\exists z$ (¬Bew$[z] \wedge S[x_1, x_2, z]$)，而是将其定义为

$$\exists z\ (C[z] \wedge S[x_1, x_2, z])$$

E as

$$D[w, w]$$

并将 G 定义为

$$E[\text{‘}E\text{’}]$$

也就是说

$$\exists z\ (C[z] \wedge S[\text{‘}E\text{’}, \text{‘}E\text{’}, z])$$

命题 G在可证意义上等价于

$$\exists z\ (C[z] \wedge z = \text{‘}E[\text{‘}E\text{’}]\text{’})$$

因此等价于

$$C[\text{‘}G\text{’}]$$

取命题 $C =$ ¬Bew$[x]$，我们得到上面的说谎者命题。

取命题 $C = \text{Bew}[x]$，我们得到 Henkin真值说谎者命题H，使得 $H \Leftrightarrow \Box H$ 是可证明的。

取命题 $C = \text{Bew}[x] \Rightarrow P$，我们得到 L¨ob 命题，使得$L \Leftrightarrow (\Box L \Rightarrow P)$ 可证明。

取命题

$$C = \forall y\ (Proof[y, x] \Rightarrow \exists z\ (z \leq y \wedge (\exists w\ (Neg[x, w] \wedge Proof[z, w]))))$$

我们得到 Rosser命题，使得

$$R \Leftrightarrow \forall y\ (Proof[y, \text{`}R\text{'}] \Rightarrow \exists z\ (z \leq y \wedge (\exists w\ (Neg[\text{`}R\text{'}, w] \wedge Proof[z, w]))))$$

可证明，也就是说

$$R \Leftrightarrow \forall y\ (Proof[y, \text{`}R\text{'}] \Rightarrow \exists z\ (z \leq y \wedge (Proof[z, \text{`}\neg R\text{'}])))$$

可证明。

这个不动点定理是另一个例子，说明抽象化——考虑任意命题——能够简化证明，避免冗余。

证明 G 和 $\neg G$ 在 $\mathcal{T}$中都不可证明，使用 $G \Leftrightarrow \neg\Box G$，即必然化规则及其逆命题，就相当直接了：如果 G可证明，那么根据 $\Box G$ （必然化规则），$\neg\Box G$ （等价关系）也可证明。因此， $\bot$ 可证明。矛盾。如果 $\neg G$ 可证明，那么根据 $\Box G$ （等价关系）， G 也可证明，进而根据 $\bot$ （必然化规则的逆命题），也可证明。因此，可证明。矛盾。

由于该证明使用了必然化规则的逆命题，因此要求理论 $\mathcal{T}$ 是ω-一致的。

8 一致性与 ω-一致性

上面讨论的三个证明使用了比一致性更强的假设： ω-一致性或存在标准模型。G¨odel-Rosser 定理允许将这一假设弱化为一致性。因此，我们考虑一个理论 $\mathcal{T}$，它具有可判定的公理集，是罗宾逊算术的扩展，并且是一致的。

该证明使用了“小必然化规则逆命题”：如果该命题

$$\exists x\ (x \leq \underline{n} \wedge Proof[x, \text{`}A\text{'}])$$

可证明，那么 A 也是可证明的。事实上，由

$$\exists x\ (x \leq \underline{n} \wedge Proof[x, \text{`}A\text{'}])$$

我们可以推导出

$$Proof[\underline{0}, \text{`}A\text{'}] \vee Proof[\underline{1}, \text{`}A\text{'}] \vee ... \vee Proof[\underline{n}, \text{`}A\text{'}]$$

并且如果 A 不可证明，那么 ¬证明[$\underline{0}$,‘A’]、 ¬证明[$\underline{1}$,‘A’], ...、 ¬证明[$\underline{n}$,‘A’] 将是可证明的。于是 $\bot$ 将是可证明的， A 也同样如此。一个矛盾。

那么， R 不可 $\neg R$ 证明这一点的证明相当直接。如果 R 有一个证明 π ，那么命题

$$Proof[\text{`}\pi\text{'}, \text{`}R\text{'}]$$

是可证明的，而命题

$$\forall y\ (Proof[y, `R'] \Rightarrow \exists z\ (z \leq y \wedge Proof[z, `\neg R']))$$

也。因此，命题

$$\exists z\ (z \leq `\pi' \wedge Proof[z, `\neg R'])$$

是可证明的，并且根据小必然化规则逆命题，$\neg R$ 也。因此 $\bot$ 是可证明的，这与该理论的一致性矛盾。

如果 $\neg R$ 有一个证明 π，那么命题

$$Proof[`\pi', `\neg R']$$

是可证明的，而命题

$$\exists y\ (Proof[y, `R'] \wedge \forall z\ (\neg z \leq y \vee \neg Proof[z, `\neg R']))$$

也。因此，命题

$$\exists y\ (Proof[y, `R'] \wedge (\neg `\pi' \leq y \vee \neg Proof[`\pi', `\neg R']))$$

是可证明的。因此命题

$$\exists y\ (y < `\pi' \wedge Proof[y, `R'])$$

是可证明的，而且根据小必然化规则逆命题，R 也成立。因此 $\bot$ 是可证明的，这与理论的一致性相矛盾。

引入这个小必然化规则逆命题，可以使这个证明更具模块性。

然而，由于这个证明不仅用到了可证明性的概念，也用到了证明的概念，因此无法用记号 $\Box A$抽象地表述。

9 第二不完备性定理

第二不完备性定理表明，皮亚诺算术的一致扩张无法证明其自身的一致性。

9.1 希尔伯特–伯奈斯引理

除了必然化引理：如果 A 可证明，那么 $\Box A$也可证明，第二不完备性定理的证明还需要另外两个引理：肯定前件式的内化

$$\Box(A \Rightarrow B) \Rightarrow \Box A \Rightarrow \Box B$$

是可证明的，以及必然化内化

$$\Box A \Rightarrow \Box\Box A$$

是可证明的。

这两个引理的证明使用了归纳法，无法在罗宾逊算术中得到证明。因此，我们考虑一个理论 T，它具有可判定的公理集，是皮亚诺算术的扩展，并且是一致的。

许多书籍只说，这两个引理的证明冗长而乏味。有些书则强调，非形式化表述

$$\text{If } A \Rightarrow B \text{ and } A \text{ are provable, then } B \text{ is provable.}$$

and

$$\text{If } A \text{ is provable, then } \Box A \text{ is provable.}$$

可以在算术中表达为上面的两个命题，并且它们的证明可以在算术中形式化——而将这些证明形式化的过程冗长而乏味。

正如 Miquel [7], 所指出的，这些证明取决于命题 Proof 的选择，也就是表达函数 proof 的证明检查程序的选择。我们使用的程序将这样的树进行变换：树的叶子标记为证明或符号 0 和 1，所有内部节点均标记为符号 and。

在这样的树中，我们考虑两类可约表达式。

- 一个标记为 and 的内部节点，如果它的两个子节点都标记为 0 或 1，则是一个可约表达式。当两个子节点都标记为 1 时，它约化为 1；否则约化为 0。

- 一个标记为证明的叶子 π 是一个可约表达式。如果该证明的根标记为命题 A，且直接子证明为 π_1、.....、 π_n，我们将其写作 $A(\pi_1$、.....、 $\pi_n)$。令 π_1、.....、 π_n的根所标记的命题为 B_1、.....、 B_n。如果存在一条推导规则，可以从 B_1、.....、 B_n推出 A，那么它约化为树 (...(π_1 and π_2)... and π_n)，并且当 $n = 0$时约化为 1。否则，它约化为 0。

我们首先定义一个程序 step，它会将树中最左侧的可归约表达式进行归约。然后，我们定义一个程序 check，反复执行程序 step，直到得到一棵不可归约的树；该树只包含一个节点，其标签为 1 或 0，这就是算法的结果。

应用于 π 和 A 的程序 proof，会对仅包含一个标签为 π的节点的树执行程序 check。如果结果为 1 且 π 的根为 A，则返回 1；否则返回 0。

很容易通过对 n 进行归纳法证明：如果 step$^n(\ulcorner\pi_1\urcorner) = \ulcorner\pi_1'\urcorner$，那么 step$^n(\ulcorner\pi_1$ 和 $\pi_2\urcorner) = \ulcorner\pi_1'$ 和 $\pi_2\urcorner$；并且如果 step$^n(\ulcorner\pi_2\urcorner) = \ulcorner\pi_2'\urcorner$，那么 step$^n(\ulcorner 1$ 和 $\pi_2\urcorner = \ulcorner 1$ 和 $\pi_2'\urcorner$。

因此，如果 proof$(\ulcorner\pi_1\urcorner, \ulcorner A \Rightarrow B\urcorner) = 1$ 且 proof$(\ulcorner\pi_2\urcorner, \ulcorner A\urcorner) = 1$，那么存在自然数 n 和 p，使得 step$^n(\ulcorner\pi_1\urcorner) = \ulcorner 1\urcorner$ 且 step$^p(\ulcorner\pi_2\urcorner) = \ulcorner 1\urcorner$。因此，step$^{n+p}(\ulcorner\pi_1$ 和 $\pi_2\urcorner) = \ulcorner 1$ 和 $1\urcorner$，且 step$^{n+p+1}(\ulcorner\pi_1$ 和 $\pi_2\urcorner) = \ulcorner 1\urcorner$。由于 step$(\ulcorner B(\pi_1, \pi_2)\urcorner = \ulcorner\pi_1$ 和 $\pi_2\urcorner$，step$^{n+p+2}(\ulcorner B(\pi_1, \pi_2)\urcorner) = \ulcorner 1\urcorner$。因此，$proof(\ulcorner B(\pi_1, \pi_2)\urcorner, \ulcorner B\urcorner) = 1$。

这个仅使用归纳法的证明可以用皮亚诺算术表达，并且它是以下命题的证明：

$$\forall x_1 \forall x_2 \forall y\, (Proof[x_1, \text{`}A \Rightarrow B\text{'}, \underline{1}] \Rightarrow Proof[x_2, \text{`}A\text{'}, \underline{1}] \Rightarrow M[\text{`}B\text{'}, x_1, x_2, y] \Rightarrow Proof[y, \text{`}B\text{'}, \underline{1}])$$

其中， M 是表示将「B」、「π_1」和「π_2」映射到「B（π_1、π_2）」的函数的命题。

由这个命题以及将「B」、「π_1」和「π_2」映射到「B（π_1、π_2）」的函数的总体性，我们得到一个关于下式的证明

$$(\exists x_1 Proof[x_1, \text{`}A \Rightarrow B\text{'}, \underline{1}]) \Rightarrow (\exists x_2 Proof[x_2, \text{`}A\text{'}, \underline{1}]) \Rightarrow \exists y\ Proof[y, \text{`}B\text{'}, \underline{1}]$$

也就是下式的证明

$$\Box(A \Rightarrow B) \Rightarrow \Box A \Rightarrow \Box B$$

这就是肯定前件式的内化。

对于其他所有演绎规则，也可以给出类似的论证。这使我们能够对必然化引理给出一个新的证明：如果 A 有一个证明，那么 $\Box A$ 也有一个证明，即通过对 A 的证明进行归纳，逐步构造 $\Box A$ 的证明。

这个证明也可以在算术中形式化为下式的证明

$$\Box A \Rightarrow \Box\Box A$$

9.2 第二不完备性定理

第二不完备性定理的证明相当直接：理论 $\mathcal{T}$ 不会证明自身的一致性 $\neg\Box\bot$。

令 G 为这样的命题： $G \Leftrightarrow \neg\Box G$ 可证明。

命题 $G \Rightarrow \Box G \Rightarrow \bot$ 可证明，因此利用必然化引理和肯定前件式内化，命题 $\Box(G \Rightarrow \Box G \Rightarrow \bot)$ 和 $\Box G \Rightarrow \Box\Box G \Rightarrow \Box\bot$ 也可证明。于是，利用必然化内化，命题 $\Box G \Rightarrow \Box\bot$ 也可证明。

可以注意到，该命题是第一不完备性定理前半部分的内化：如果 G 可证明，那么 $\bot$ 也可证明。

现在，假设命题 $\neg\Box\bot$ 可证明，我们可以推出命题 $\neg\Box G$ 也可证明。因此，命题 G 也可证明；根据必然化规则， $\Box G$也可证明，进而 $\bot$ 也可证明，这与理论的一致性相矛盾。

9.3 可证性逻辑

第一不完备性定理的证明使用符号 $\Box$ ，但只用到了必然化引理。相比之下，第二定理的证明运用了更多模态逻辑：引理 K 和 4

$$\Box(A \Rightarrow B) \Rightarrow \Box A \Rightarrow \Box B$$

$$\Box A \Rightarrow \Box\Box A$$

这里可以提到模态逻辑，尤其是可证性逻辑；不过，可证性逻辑的记号及其模块化方法——先证明希尔伯特–伯奈斯引理，再在第二不完备性定理的证明中使用这些引理——即使不定义可证性逻辑本身，也可以采用。

9.4 L¨ob 定理

使用可证性逻辑的记号和一般不动点定理，L¨ob 定理是第二不完备性定理的直接推广。L¨ob 定理如下：对于任意命题 P，如果 $\Box P \Rightarrow P$ 可证明，那么 P。

我们使用命题 L 来代替命题 G，其中 $L \Leftrightarrow (\Box L \Rightarrow P)$ 是可证的。

命题 $L \Rightarrow \Box L \Rightarrow P$ 是可证的，因此利用必然化引理和肯定前件式内化，命题 $\Box(L \Rightarrow \Box L \Rightarrow P)$ 和$\Box L \Rightarrow \Box\Box L \Rightarrow \Box P$ 也都是可证的。利用必然化内化，命题 $\Box L \Rightarrow \Box P$ 也是可证的。

现在，如果我们假设命题 $\Box P \Rightarrow P$ 是可证的，就可以推导出命题 $\Box L \Rightarrow P$ 是可证的。因此，命题 L 也是可证的，并且根据必然化规则，$\Box L$ 也是可证的，从而 P 是可证的。

取 $P = \bot$，第二不完备性定理是该定理的推论，因此，该定理也可以在第二不完备性定理之前得到证明。

最后，取 $P = H$，即 Henkin真值说谎者命题，使得 $H \Leftrightarrow \Box H$可证明，另一个推论是，该命题 H 可证明。事实上，既然$\Box H \Rightarrow H$ 可证明， H也可证明。

10 历史与哲学

不完备定理的历史本身就是一个非常丰富的主题。学生可以了解到，可计算函数的概念以及可证明性的不可判定性（因此也包括计算机科学家的证明），是在不完备定理（1931年）之后，由 Church 和 Turing 在其 1936 年的工作中提出的。因此，从历史上看，第一个证明是数学家的证明。学生还可以了解到，第二不完备性定理和不可判定性分别解决了什么问题：Hilbert第二问题和 Hilbert判定问题。

从更偏哲学的角度来看，许多评论者认为，不完备性证明了理论无法谈论自身。恰恰相反，第二不完备性定理之所以存在，正是因为理论的一致性可以在理论内部表述。不完备性也常被描述成一场灾难。即便它揭示了演绎方法的某些局限，也并不意味着演绎方法走到了尽头。

11 结论

正如《引言》中所指出的，不完备性定理内容丰富，无法在一门课程中讲完。我们尝试区分适合在基础课程中讲授的内容——可计算函数的弱表示、可证明性的不可判定性，以及在合理假设下的第一不完备定理——与可以留到进阶课程中讲授的内容：可计算函数的强表示、第一不完备定理的完整证明、希尔伯特–伯奈斯引理、第二不完备定理、L¨ob 定理，以及这些定理所需的最小假设。

我们还主张，应当独立讲授可计算函数这一概念、其两种定义之间的等价关系（借助 G¨odel 的 β 函数和中国剩余定理），以及具体语法，并且应在讲授算术的不可判定性和不完备性之前讲授这些内容。

最后，我们为这些证明辩护，主张采用抽象且模块化的方法，运用抽象语法、衔接、通用编号、可证性逻辑记法、必然化规则的逆命题、小必然化规则逆命题、一般不动点引理以及希尔伯特–伯奈斯引理。这项工作仍需继续。尤其是，G¨odel-Rosser 定理以及希尔伯特–伯奈斯引理的证明还不够抽象。

当然，抽象定义和一般引理应当配以具体示例加以说明，但这些具体示例不应取代它们。

参考文献

[1] P.B. Andrews。《数理逻辑导论：通过证明求真》。Academic Press，1986年。

[2] U. Boker 和 N. Dershowitz。《任意论域上的丘奇–图灵论题》。载于 A. Avron、N. Deshowitz 和 A. Rabinovich 编，《计算机科学的支柱：献给 Boris (Boaz) Trakhtenbrot 的论文集，纪念其 85 岁生日》，第 4800 卷《计算机科学讲义》，第 199–229 页。Springer，2008年。

[3] R. Cori 和 D. Lascar。《数理逻辑：一门附习题的课程》。牛津大学出版社，2000年。

[4] G. Dowek。《证明与算法》。Springer，2011年。

[5] J.-L. Krivine。《λ演算、类型与模型》。Ellis Horwood，1993年。

[6] A. Martinet。《´普通语言学 ´要义》。Colin，1968年。

[7] A. Miquel。《G¨odel不完备定理》。´里昂高等师范学院。可在线获取。