

• 科学前沿 •

基于彭罗斯三个世界理论视角的形式化证明

Formal Proof from the Perspective of Penrose's Theory of the Three Worlds

杨帆 / YANG Fan

(大连海事大学公共管理与人文学院, 辽宁大连, 116026)
(College of Public Administration and Humanities, Dalian Maritime University, Dalian, Liaoning, 116026)

摘要: 在信息化时代, 人们能够使用计算机实现对数学定理的形式化证明, 由此改变了数学实践的范式。鉴于形式化证明中涉及到人脑证明、机器证明和数学形式三种不同类别的存在, 有必要从本体论层面对其进行考察。彭罗斯的三个世界理论发展自波普尔, 并以其将数学形式作为柏拉图世界的独立存在而适用于形式化证明的本体论构建。为了使彭罗斯的三个世界理论能够作为形式化证明的本体论, 还需要对其理论中世界间的关系进行改造。改造后的形式化证明的本体论为“数学是发现还是发明”这一著名问题提供了进一步的反思, 彰显出三个世界理论带来的崭新视角。

关键词: 形式化证明 三个世界 本体论 数学哲学

Abstract: In the information age, people are able to use computers to accomplish formal proof of mathematical theorems, thus shifting the paradigm of mathematical practice. In view of different existences involved in formal proof: human brain proofs, machine proofs, and mathematical forms, it is necessary to examine them at the ontological level. Penrose's theory of the three worlds develops from Popper and applies to the ontological construction of formal proof with its inclusion of mathematical forms as independent beings in the Platonic world. In order for Penrose's theory to serve as the ontology for formal proof, the relations between worlds in the theory also need to be reformed. The reformed ontology of formal proof provides a further reflection on the famous question "Is mathematics a discovery or an invention?" and reveals the new perspective brought by the theory of the three worlds.

Key Words: Formal proof; Theory of the three worlds; Ontology; Philosophy of mathematics

中图分类号: B813; N031 DOI: 10.15994/j.1000-0763.2025.07.014 CSTR: 32281.14.jdn.2025.07.014

形式化证明是以符号语言和推理规则为基础所写出的数学证明。相较于大部分数学教材和论文中常见的自然语言证明, 形式化证明具有严格和机器可验证的特点。正是由于这些特点, 人们可以通过计算机实现数学定理的形式化证明, 由此形成了机器定理证明, 具体分为

自动定理证明和交互式定理证明两个方向: 前者是由计算机自动完成数学定理的证明, 后者是通过人机交互证明数学定理。20世纪中期, 伴随着计算机科学的兴起, 机器定理证明成为了人工智能领域中符号主义派的研究方向之一。1956年, 人工智能领域中的第一个程序 Logic

基金项目: 国家社会科学基金重大项目“面向自然语言理解的逻辑构建和符号接地问题的哲学、心理学研究”(项目编号: 18ZDA032); 中央高校基本科研业务费专项资金资助“面向机器定理证明的模糊逻辑研究”(项目编号: 3132022308)。

收稿日期: 2024年12月13日

作者简介: 杨帆(1993-)男, 辽宁营口人, 大连海事大学公共管理与人文学院讲师, 研究方向为数理逻辑与数学哲学。
Email: fanyang@dlmu.edu.cn

Theorist被研发用于证明《数学原理》(*Principia Mathematica*)中的定理。^[1]此后半个多世纪的时间,形式化证明驱动的机器定理证明逐渐在数学定理的证明和验证中得到广泛使用,现今已实现对许多数学定理的辅助验证,并规模化应用于计算机软硬件的验证中。对于形式化证明,可以从狭义和广义两个层面去认识:狭义形式化证明是开篇给出的定义;广义形式化证明则可以泛指形式化证明驱动的机器定理证明等应用领域。我们将在广义的层面上对形式化证明展开讨论。

如果想要深入探究形式化证明为数学等领域带来的转变,那么本体论层面的考察是必不可少的。分析形式化证明我们发现,在这一过程中实际上隐含着两种不同层面的证明,分别是人脑证明和机器证明,二者都是试图对数学定理进行证明的活动。一方面,人脑证明虽然是依托于神经元等物质得以实现的过程,但是它已经属于精神层面的心智活动;另一方面,机器证明是一个机械过程,必须依托计算机硬件才能够实现,因此仍然归属于物质层面。此外,从数学柏拉图主义的观点来看,数学概念和定理是独立存在的抽象实体。简言之,人脑证明、机器证明以及二者共同指向的数学定理似乎暗示了三种不同层面的存在:精神、物质与理念。这种本体论层面的暗示与彭罗斯(R. Penrose)提出的三个世界理论颇为契合。彭罗斯的三个世界理论是英国哲学家波普尔的三个世界理论的变种,二者都是一种多元论立场。不同于肇始于笛卡尔的心物二元论认为只存在精神世界与物理世界,三个世界理论认为还存在着第三个世界,只不过波普尔将其称作文化世界(一个使用更广泛的名字是“世界3”),而彭罗斯将其称作柏拉图世界,显然波普尔和彭罗斯的理论存在着一定的区别。

本文试图借助彭罗斯的三个世界理论,从本体论层面考察形式化证明,其目的是加深对形式化证明的理解,为后续探索形式化证明中的相关哲学问题提供必要的理论框架与规范。本文具体安排如下:首先,我们将在第一节中对波普尔和彭罗斯的三个世界理论间的区别进行

厘清;其次,我们将在第二节中基于彭罗斯的三个世界理论,从本体论层面讨论形式化证明的运行机制;最后,我们将在第三节中借助对“数学是发现还是发明”这个问题的反思,阐释三个世界理论为形式化证明带来的崭新视角。

一、三个世界理论的演化: 从波普尔到彭罗斯

1. 波普尔的三个世界理论

在“没有认识主体的认识论”中,波普尔首次提出了三个世界理论。他认为可以区分出下列三个不同的世界:物理客体或物理状态的世界(世界1)、意识状态或精神状态的世界(世界2)和思想的客观内容的世界(世界3)。([2], p.109)后来在“关于客观心灵的理论”中,波普尔又将世界3进一步解释为“客观意义上的观念的世界——它是可能的思想客体的世界:自在的理论及其逻辑关系、自在的论证、自在的问题情境等的世界。”([2], p.159)

第三个世界的概念并非波普尔首创,古希腊时期柏拉图提出的理念世界就是既非物理世界又非精神世界的第三个世界。但是,波普尔确是第一位完整地提出三个世界理论并持续为之辩护的哲学家。此外,波普尔的世界3与柏拉图的理念世界有很大不同:理念世界是不变的,而世界3是人工产物的世界,因此是可变的;理念世界只包括真理论,而世界3还可以包括假理论甚至是猜想和反驳。([2], p.125)

三个世界只是基础的本体论划分,更重要的是我们如何理解三个世界间的关系。波普尔将三个世界间的关系阐释为“第一世界与第三世界之间以第二世界为中介。”(图1)([2], p.159)世界2可以与世界1或世界3发生相互作用,但是世界1和世界3不能相互作用,除非通过世界2的干预。

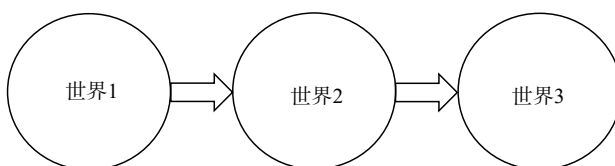


图1 波普尔的三个世界

然而，波普尔始终没有对三个世界各自的范围进行严格的划分，甚至认可三个世界间的互相重叠。例如，作为人类创造物的计算机从属于世界3，但它自身的物质属性又使得它可以从属于世界1。在拉多万 (M. Radovan) 看来，“这种不精确性给他带来了各种困难。”^[3] 拉多万进一步规定：“三个世界应该被观察（和定义）为三个互不相交的现象类别：它们不重叠，也不在本体论上互相还原。”^[3] 相较于波普尔的三个世界理论，彭罗斯的三个世界理论在范围上是更加清晰的。

2. 彭罗斯的三个世界理论

在《心灵之影》(*Shadows of the Mind*) 中，彭罗斯对波普尔的三个世界理论进行了初步的改造，主要目的是佐证其数学柏拉图主义的立场。他保留了波普尔的世界1和世界2，但是将世界3进行了一定的修改。彭罗斯的第三世界不再是波普尔的增加思想的客观内容的世界3，而是只包含数学形式的柏拉图世界。显然，彭罗斯的柏拉图世界的范围要小于波普尔的世界3的范围，但也正是由于柏拉图世界聚焦于数学形式，因此它的复杂程度要远小于世界3。

彭罗斯的三个世界理论是一种回归柏拉图的尝试，不过他的改造绝不仅是复古那么简单。在彭罗斯看来，数学柏拉图主义并未过时，真正的困境在于如何看待柏拉图世界对物理世界的影响。([4], p.413) 因此，相较于争论如何规定三个世界各自的内涵与外延，彭罗斯更加关心的是三个世界间的相互生成关系，其动机很可能是源自对波普尔在三个世界间关系上阐释的不满。彭罗斯将这种生成关系称作奥秘，这也是其三个世界理论的核心概念。他认为，物理世界、精神世界和柏拉图世界间两两存在着单向的奥秘（如图2中的箭头所示）：从柏拉图世界到物理世界是奥秘1，从物理世界到精神世界是奥秘2，从精神世界到柏拉图世界是奥秘3。同时，三个奥秘还相应地显现出单向的扩张形态（如图2中的箭头外侧线段所示），这些扩张形态进一步地刻画了三个世界间的神秘联系。

奥秘1体现为物理世界对于柏拉图世界中

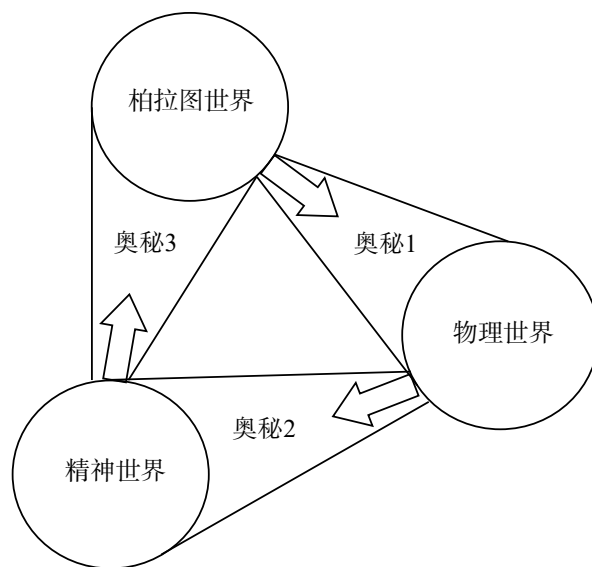


图2 彭罗斯的三个世界

数学形式的严格遵守，即物理世界符合建立在物理学之上的数学模型。彭罗斯认为，“在我们关于基本物理过程的理论中，始终存在一个潜在的数学结构，它不仅被证明非常准确，而且在数学上也很复杂。”([4], p.415) 奥秘1的扩张形态表示，整个物理世界都可以由数学形式所刻画。反过来，并非所有数学形式都是对物理世界的刻画，例如数学中无穷的概念就不是对物理世界的刻画，我们无法在物理世界中找到无穷，却可以在数学（尤其是集合论等领域）中谈论和表示无穷。

奥秘2体现为精神世界源自物理世界，这是心灵哲学讨论的焦点，例如意识如何从物质中产生这样的心身问题。类似于奥秘1的扩张形态，奥秘2的扩张形态表示，所有精神世界中的心灵都是以物理世界中的部分实在为基础的，例如人类的心灵建立在构成大脑神经元的物质基础上。反过来，并非所有物质都具有心灵，例如桌子和椅子就不具有心灵。

奥秘3并不意味着数学是人类心灵的产物，而是说心智活动可以通达数学形式。奥秘3的扩张形态表示，整个柏拉图世界中的数学形式都可以通过精神世界中的部分心智活动得到认识。反过来，并非所有心智活动都是对数学形式的认识，例如喜悦也是一种心智活动，但它与数学形式毫无关联。

在彭罗斯看来，三个奥秘的方向是他看待

三个世界间联系的一种方式,也有一些哲学家的观点是与他的奥秘方向相反的,例如贝克莱就将物理世界看作是精神世界的产物。([4], p.417)似乎自相矛盾之处在于,每个世界都从前一个世界的一小部分中显现,最后却能够形成回环的联系。彭罗斯对此表达了非常开放的态度,他并未真正说出三个奥秘的具体过程,“除非所有这些世界的相互关联的特征都被视为发挥作用,否则不会有明确的答案出现。”([4], p.420)

3. 波普尔与彭罗斯的理论间的区别

无论是波普尔还是彭罗斯,他们对三个世界理论的辩护都暗示了其多元论的倾向。毕竟,如果世界3中的实在与世界1和世界2中的实在没有什么区别的话,那么世界3的设定就是冗余的;而如果柏拉图世界中的数学形式可以被视为是物质或精神,那么柏拉图世界也将不过是物理世界或精神世界的子集。如此都将使得波普尔和彭罗斯沦为笛卡尔式的二元论者。可以说,彭罗斯的三个世界理论发展自波普尔,但是鉴于其对数学和物理学的侧重,他的三个世界理论仍与波普尔的三个世界理论有很大的区别,这主要体现在以下三个方面:

第一,二者对于第三世界的规定不同。波普尔的世界3是动态的、持续扩张的,随着人类创造物的不断涌现,世界3的内容也不断被充盈。在波普尔看来,他的世界3是人造的,同时又是超乎人类的;([2], p.163)而彭罗斯的柏拉图世界是永恒的、不变的、先于人类存在的,并且逐渐为人类所揭开面纱的。由于波普尔的世界3包含了并不属于柏拉图世界的实在,因此彭罗斯认为波普尔的世界3不能够作为物理世界的生成框架。([4], p.421)甚至在彭罗斯看来,数学形式之外的理念(例如“善”和“美”的理念)能否归于柏拉图世界并不重要。([4], p.416)

第二,二者对于三个世界间关系的规定不同。结合图1和图2可以看出,波普尔的三个世界间的关系是线性的;而彭罗斯的三个世界间的关系是回环的。这体现在三个世界通过三个奥秘及其扩张形态互相作为基础,最后形成一

幅回环状的图景。三个奥秘是彭罗斯对波普尔的三个世界理论最大的改进,这使得三个世界得以真正地互相联系和互相影响。

第三,二者通过三个世界理论所要探讨的问题不同。波普尔是通过论证世界3的自主性来支持他的知识客观性观点,从而呈现出他在知识论方面持有的达尔文主义进化论立场。而彭罗斯默认了三个世界各自的实在性,他更关心的是三个世界间神秘的生成关系。对于彭罗斯来说,柏拉图世界的自主性是毋庸置疑的,重要的是我们如何去看待和理解柏拉图世界对于物理世界的影响。

总之,我们无法武断地认定波普尔和彭罗斯的三个世界理论哪个更好、更正确,只能说二者通过三个世界理论所要讨论的问题具有不同的侧重点。既然如此,那么我们为何要选取彭罗斯的三个世界理论作为视角来考察形式化证明?回顾英国数学家哈代(G. H. Hardy)对数学实在与精神间关系的困惑:“关于数学实在的本质,数学家或哲学家还没有取得一致意见。一些人认为它是‘精神的’产物,在某种意义上我们构造了它。另一些人认为它在我们之外,独立于我们而存在。一个人如果能给出关于数学实在性的有说服力的描述,他就能解决许多最困难的形而上学问题。如果他的描述能同时包括物理实在性,他就解决了全部的形而上学问题。”^[5]彭罗斯的三个世界理论就同时包括了对数学实在、物理实在与精神实在的描述。虽然我们不敢轻易断言彭罗斯真的如哈代所言有可能“解决全部的形而上学问题”,但是既然我们想要为涉及到数学定理的形式化证明寻求一种可靠的本体论,那么显然选取彭罗斯的三个世界理论是更加合适的。

二、形式化证明的本体论

1. 人脑证明与机器证明

本小节对形式化证明中涉及到的两种证明进行区分。本文开篇曾简要指出:人脑证明是属于精神世界的心智活动,而机器证明是属于物理世界的机械过程。虽然心智活动建立在物

理世界中组成人脑的神经元基础上，但是不可否认它对精神世界的从属地位。计算机的运行过程是，先通过输入设备获取指令，中央处理器根据指令读取数据进行运算，并将运算结果存入存储器中，所有指令执行完毕后通过输出设备输出运算结果。因此，机器证明显然是从属于物理世界的机械过程，它只是执行人类输入的指令，而不是精神世界的心智活动。

除了所属世界的严格区分，二者还有一个不容忽视的差异：人脑证明依赖的是数学直觉和逻辑推理，而机器证明依赖的仅是逻辑推理。人脑在证明数学定理时，数学直觉先于逻辑推理，也就是先通过直觉获取对数学定理的认知。直觉是一个极具争议的概念。菲茨拜因(E. Fischbein)将“直觉”的多重含义总结为三种：^[6]其一，直觉是特定知识的基本来源，例如笛卡尔用“直观”来表示“纯净而专注的心灵的构想”；^[7]其二，直觉是把握真理的方法，例如柏格森(H. Bergson)将直觉作为一种朝向生命的本能，“直觉能使我们看到智慧材料的不足，能使我们隐约看到补充智慧材料的方法”；^[8]其三，直觉是一种特殊类型的认知，例如彭加勒(J. H. Poincaré)将数学研究分为逻辑和直觉两部分，并认为逻辑是证明的工具，直觉是发明的工具。^[9]美国数学家怀尔德(R. L. Wilder)则认为，直觉在数学概念的演变中起到了重要的作用。^[10]但是，数学直觉常常存在着严重的误导性，这就不得不使用逻辑推理来进行澄清。例如，当被问及所有自然数和所有偶数的个数哪一个更多时，有人依靠直觉会给出所有自然数更多的错误答案。然而，通过逻辑推理我们会发现，自然数集和偶数集是等势的，也就是所有自然数和所有偶数是一样多的。因此，人脑证明的实际进程是，先由数学直觉获取对数学定理的认知，再由逻辑推理完成对数学定理的有效证明。

机器证明则是计算机严格执行逻辑推理指令的过程。它当然也会以数学定理为其证明的目标，但是它并不会获得对数学定理的认知，证明数学定理是由人类作为一项任务指派给它的。这种指派过程与我们使用工具的过程相类

似。例如，当我们使用菜刀切菜时，菜刀的目标就是切割蔬菜，我们通过指派给菜刀一个向下的力以使得它完成这个进程，但显然菜刀并不会获得对切割蔬菜的认知。也就是说，计算机作为物理世界中的存在，并不具备数学直觉，却可以实现逻辑推理。那么是什么使得逻辑推理可以在其中实现呢？作为计算模型的图灵机给出了答案。^[11]图灵机是一个抽象模型，它包含：(1)一条无穷长的纸带，纸带被划分为多个格子用于记录符号；(2)一个读写头，可以左右移动，也可以读取或改变当前格子的符号；(3)控制器，用于制定状态转移规则和控制读写头的操作。现代计算机的计算能力等价于图灵机模型。图灵认为：“实际上，人们可以用任何语言与这些机器进行交流，只要它是一种精确的语言，即原则上人们应该能够用任何符号逻辑与机器交流，前提是给机器提供指令表，使其能够解释该逻辑系统。”([12], p.103)他甚至设想：“可能会有一些让机器对数学公式进行实际操作的尝试。”([12], p.103)而机器证明就验证了这一设想。

2. 形式化证明：三个世界间的再联系

下面我们在彭罗斯的三个世界理论视角下考察形式化证明的运作机制。我们将形式化证明中的具体过程用关系A、B和C表示(图3)，简洁起见，我们省略了图2中表示三个奥秘的箭头。同时需要注意的是，关系A、B和C与奥

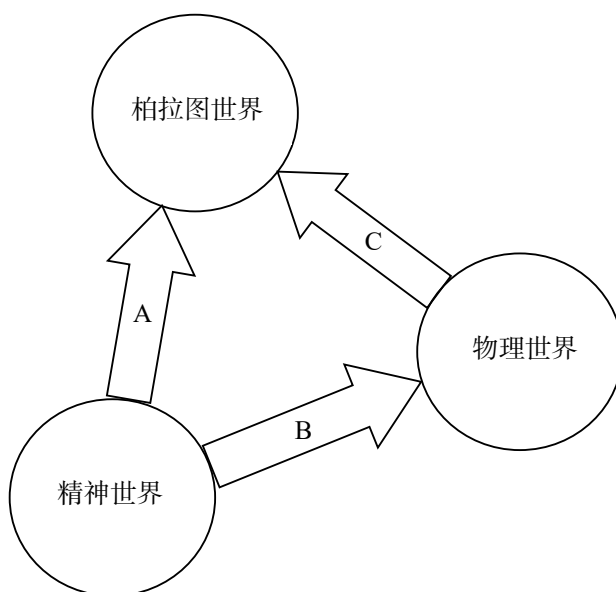


图3 三个世界间的再联系示意图

秘1、2和3的不同之处在于，它们不具有扩张形态。

形式化证明的运作机制将在三个世界间的再联系中得到阐释：

关系A是精神世界中心智活动对柏拉图世界中数学定理的认识，也就是所谓的人脑证明，这一过程依赖于数学直觉和逻辑推理。虽然关系A与彭罗斯的奥秘3具有相同的方向，但是二者之间存在严格的区别：奥秘3包含了心智活动对数学形式（包括数学概念和数学定理）的认识，而关系A只包括对数学定理的认识。这是因为心智活动对有些数学概念的认识不需要逻辑推理的参与，例如我们认识自然数1的心智活动就仅依赖于数学直觉，而不需要逻辑推理。因此，关系A是被容纳在奥秘3之中的。

关系B是精神世界中关于数学定理证明的心智活动在物理世界中计算机程序上的模拟，这一过程依赖于人机交互，是交互式定理证明（也称计算机辅助证明）的一部分。人类通过将关于数学定理证明的心智活动转化为程序语言，输入到既定的形式化证明程序中，从而由计算机完成对证明活动的模拟。

关系C是物理世界中计算机对柏拉图世界中数学定理的证明，这一过程依赖于逻辑推理。机器证明的过程是关系B和关系C的合成，人类将待证明的数学定理输入到计算机程序中，并由计算机完成定理的证明和验证工作。在执行数学定理证明的逻辑推理方面，关系B和关系C的合成等价于关系A。不过，在发现新的数学定理方面，前者与后者并不是等价的。这也说明，为什么至今为止计算机始终不能自发地找到尚未被人类所发现的数学定理并给出证明，因为机器证明的过程缺乏关键性的数学直觉。从这一点来看，形式化证明也就不能被视为人工智能。彭罗斯认为，计算机不能够模拟意识活动。^[13]这是否会对我们的阐释造成影响呢？正如我们前面所说的，形式化证明只是将人脑证明中的逻辑推理部分转移到计算机中进行，而没有真正模拟与意识密切相关的数学直觉。因此，我们的解释与彭罗斯的观点没有产生冲突。

三、重思“数学是发现还是发明？”

数学是发现还是发明？这是一个古老的数学哲学问题，长期以来人们争论不休却不能得到一个令人满意的回答。英国数学家、1998年菲尔兹奖得主高尔斯（W. T. Gowers）认为，这个问题提出的原因是人们希望用它来支持自己的哲学观点：发现说为数学柏拉图主义提供支持，发明说则为反实在论提供论据。（[14]，p.1）但是，这个问题不免使人在回答时无意间陷入二元论的选择之中，即数学要么是发现，要么是发明，不可有折中选择。实际上，发现和发明同时存在于数学中。这体现在对数学形式及其表达的区分之间：

数学形式是存在于柏拉图世界中的永恒实在，人类心智活动对它的认识是一种“发现”。彭罗斯以芒德布罗集（Mandelbrot set）为例对此进行了辩护。^[15]芒德布罗集的图像具有非常复杂的细节，将其图像进一步放大我们会看到更加深层次的、重复的复杂细节，而这些仅是始于我们对复二次多项式的迭代计算。我们无法在静态的印刷上展现这个极富美感的进程，感兴趣的读者可以在网络上搜索查看相关展示。在彭罗斯看来，芒德布罗集的复杂结构不是任何人的发明，也不是精神的一部分，而是本身具有着实在性，它“就在那里”。对“就在那里”的一个更精确的表述是，“它的存在依赖于这些概念深刻、永恒和普遍的本质，以及它们的规律独立于发现它们的人这一事实。”（[4]，p.413）

数学形式的表达是人类心智活动的产物，因此是一种“发明”。例如，“6”实际上是对一种数量概念的表达，我们同样也可以用二进制“110”或者中文“六”来对这种数量概念进行表达。如果在侏罗纪时期有六只恐龙相聚，那么关于“6”的数量概念就已潜在于其中了——即便那时还没有人类和相应的数量概念的表达。数学形式的表达本质上是心智活动，无论是“6”、二进制“110”还是“六”都不是柏拉图世界中的实在，而是对柏拉图世界中

的实在的表达。

基于三个世界理论的形式化证明本体论为“数学是发现还是发明？”这个问题提供了怎样的新视角呢？柏拉图世界的自主性已经没有太多新意可讲。我们想要说明的是，以往的数学活动都是从精神世界通达柏拉图世界的，“发现”体现在心智活动对数学形式的发现。现在，我们可以借助计算机从物理世界通达柏拉图世界，因此形式化证明为这样的发现提供了一个新的路径。

然而，在对形式化证明进行深入考察后，我们认为两种发现间存在着一个不容忽视的区别：人类心智活动所发现的数学形式，既包括数学概念，例如自然数和实数，也包括数学定理，例如勾股定理；而计算机通过形式化证明所发现的数学形式，仅仅只是数学定理。计算机不能发现新的数学概念！设想我们只将自然数集教给计算机，那么当输入“2-4”这个运算后，计算机不会得出-2作为答案。在交互式定理证明程序Coq（现已更名为Rocq）中（如图4），当我们使用Compute指令计算“2-4”时，计算机会在右边区域输出0（这已经是它此时能够算到的最小的数了）。虽然我们知道“2-4”的结果是-2，但是计算机不会自动地发现超出自然数之外的负数。这是因为计算机的运算仍然是从属于物理世界中的机械过程，它会严格地执行指令，Coq初始状态中的运算是建立在自然数集的，而我们设定的指令中实际上埋下了一个伏笔：自然数集对于减法不是封闭的。而如果你询问一个能理解简单的自然数加减法但尚未学习负数知识的儿童， $2-4$ 等于多少时，即便他不能给出正确的答案，他也很有可能对这个问题有所迟疑——这就是数学直觉的影响。回顾人类发现负数的历史，从最初记账时表示负债的便利，到对负数概念的正确认识，负数的发现经历了历史和文化的多重影响。因此，数学概念的发现一定程度上具有历史性的

特征，而这种特征深受精神世界中数学直觉的影响，这是物理世界中的计算机所不具备的。

总之，基于三个世界理论的形式化证明本体论由于其数学柏拉图主义的立场，支持了数学形式是发现的观点，并且提供了计算机从物理世界通达柏拉图世界的发现数学形式的新路径，只不过这个新路径只能发现数学形式中的数学定理（甚至还不能发现人类未知的定理），而不能直接发现新的数学概念。英国数学家、1966年菲尔兹奖得主阿蒂亚（M. Atiyah）曾宣称：“数学是人类的活动而不是计算机的程序。”^[16]这是否说明人脑在某种程度上对于柏拉图世界的理解是优于计算机的，或者说数学概念的发现是精神世界中心智活动的特权，而不被物理世界中的存在所享有？似乎只有在计算机真的具有相当于人类心智的能力时，我们才能够重新回到这个问题上来。

结 语

基于彭罗斯的三个世界理论考察形式化证明，有助于我们获得以下三个方面的启示：

首先，从本体论层面阐释形式化证明的运作机制。形式化证明是少见的连通三个世界的进程，它始于精神世界中关于数学定理证明的心智活动，执行于物理世界中的计算机程序，最终通向柏拉图世界中的数学定理。这种阐释也为我们后续探讨形式化证明中的具体话题（例如人机关系等）提供了框架。

其次，厘清形式化证明与人工智能的关系。虽然形式化证明驱动的机器定理证明是符号主义通往人工智能的尝试，但是在三个世界理论视角下来看，形式化证明是从物理世界通达柏拉图世界的进程，计算机在执行过程中并不具备自主的思维，需要人机交互进行引导。因此，目前不能将其视为能思考的人工智能，而只是辅助人类验证数学定理的工具。

```
Compute (minus 2 4).
```

```
Messages / Errors / Jobs /
= 0
: Datatypes.nat
```

图4 Coq中的指令示意图(左侧是输入,右侧是输出)

最后,进一步理解人脑、计算机与数学形式间的关系。本文的论述表明,数学形式(包括数学概念和数学定理)是发现,而数学形式的表达是发明。在数学形式中,数学定理的发现既可以由人类心智活动实现,也可以由计算机的机械过程实现;数学概念的发现却只能由人类心智活动来完成,物理世界中的计算机(至少现在)不可能自发地发现新的数学概念。

即便我们的本体论构建基于的是彭罗斯的改造版三个世界理论,但是仍不可否认波普尔的三个世界理论是极富开放性和包容性的哲学理论。正如王克迪所言,“一个哲学概念或理论是否有生命力,也要看它对于世界的解释能力、理论与概念的开放程度和对新事物的包容性。”^[17]哲学史的演进一再证明,不可能有一套固定不变的哲学理论能够完整地解释我们生活的现实世界——毕竟,现实世界是一直在变化和发展的。彭罗斯对波普尔三个世界理论的改造,以及我们对彭罗斯三个世界理论的再改造,也只能是对现时的形式化证明提供本体论层面的解释。倘若形式化证明在未来取得进一步的技术突破和更广泛的现实应用,那么我们的解释也将沦为窠臼,不得不与技术突破一道寻求新的演进。

[参考文献]

- [1] Newell, A., Simon, H. A. 'The Logic Theory Machine: A Complex Information Processing System'[J]. *IEEE Transactions on Information Theory*, 1956, 2(3): 61-79.
- [2] 卡尔·波普尔. 客观的知识: 一个进化论的研究[M]. 舒炜光、卓如飞、周柏乔等译, 杭州: 中国美术学院出版社, 2003.
- [3] Radovan, M. 'Computation and the Three Worlds'[J]. *Minds and Machines*, 2000, 10(2): 255-265.
- [4] Penrose, R. *Shadows of the Mind: An Approach to the Missing Science of Consciousness*[M]. Oxford: Oxford University Press, 1994.
- [5] 哈代. 一个数学家的辩白[M]. 李文林、戴宗铎、高嵘译, 大连: 大连理工大学出版社, 2022, 57.
- [6] Fischbein, E. *Intuition in Science and Mathematics: An Educational Approach*[M]. Dordrecht: Reidel, 1987, 3.
- [7] 笛卡尔. 探求真理的指导原则[M]. 管震湖译, 北京: 商务印书馆, 1991, 10.
- [8] 亨利·柏格森. 创造进化论[M]. 姜志辉译, 北京: 商务印书馆, 2004, 149.
- [9] 昂利·彭加勒. 科学的价值[M]. 李醒民译, 北京: 商务印书馆, 2011, 20.
- [10] Wilder, R. L. 'The Role of Intuition'[J]. *Science*, 1967, 156(3775): 605-610.
- [11] Turing, A. M. 'On Computable Numbers, with an Application to the Entscheidungsproblem'[J]. *Proceedings of the London Mathematical Society*, 1936-1937, 2(42): 230-265.
- [12] Turing, A. M. 'Lecture to the London Mathematical Society on 20 February 1947'[A], Ince, D. C. (Ed.) *Mechanical Intelligence*[C], Amsterdam: North-Holland, 1992, 87-105.
- [13] Penrose, R. *The Large, the Small and the Human Mind*[M]. Cambridge: Cambridge University Press, 1997, 101.
- [14] 蒂莫西·高尔斯. 数学是一种发现还是一种发明?[A], 约翰·查尔顿·波金霍尔: 数学的意义[C], 王文浩译, 长沙: 湖南科学技术出版社, 2018, 1-16.
- [15] 罗杰·彭罗斯. 皇帝新脑[M]. 许明贤、吴孟超译, 长沙: 湖南科学技术出版社, 2018, 124.
- [16] 阿蒂亚. 数学的统一性[M]. 袁向东译, 大连: 大连理工大学出版社, 2022, 90.
- [17] 王克迪. 知识-机器互动: 在世界3与世界1之间[M]. 北京: 人民出版社, 2020, 37.

[责任编辑 李斌]