

关于心灵、机器与真理的模态论证

薄谋

摘要: 卢卡斯和彭罗斯认为, 人们从哥德尔不完全性定理出发可以推断“心灵不能被机械化”的论题。这是一个强论断。弱论断是由哥德尔提出的。他认为从自己的不完全性定理出发, 人们只能推断或者“心灵不能被机械化”成立或者“存在不可判定的数学命题”成立。该析取式论断等价于“如果理想化人类心灵是理想化有限机器, 那么人们无法判定某些数学命题的真或假”。哥德尔建议人们从解决内涵悖论入手, 证明反机械主义论题成立。科尔纳在机械主义论题上取得的成功, 激发斯特恩尝试从可替代真理论出发, 证明反机械主义论题成立。科尔纳表明斯特恩的论证无效。斯特恩尝试改进他的论证。我们也表明斯特恩的改进方案无效。斯特恩两次尝试均无法取得成功, 这实际上源于人们无法认清理想化人类心灵概念即绝对可证明性概念。

关键词: 绝对可证明性; 免类型真理论; 数学真理; 数学认知

中图分类号: B81

文献标识码: A

哥德尔 (G. Gödel) 认为从他的不完全性定理出发, 人们可以推断或者“心灵不能被机械化”成立, 或者“存在不可判定的数学命题”成立。([7]) 霍斯滕 (L. Hosten) 等人把该结果称为“哥德尔析取式”。([9]) 彭罗斯 (R. Penrose) 认为从哥德尔不完全性定理出发, 人们可以推断“心灵不能被机械化”成立。([18, 19]) 科尔纳 (P. Koellner) 认为不管是“心灵不能被机械化”还是“存在不可判定的数学命题”都是不可判定的。([10, 12, 13]) 斯特恩 (J. Stern) 选择支持彭罗斯, 但他要接受科尔纳的挑战。我们把斯特恩的论证分为第一次论证和第二次论证。([28, 29]) 不管是第一次论证还是第二次论证, 斯特恩的结论都支持彭罗斯的论断。科尔纳表明斯特恩第一次论证不成立。([11]) 在科尔纳的启发下, 我们表明斯特恩第二次论证不成立。反思科尔纳和斯特恩两人的论证, 我们的结论是在无法建立起意识科学的情况下, 人们无法捕捉到绝对可证明性的真正含义。

收稿日期: 2022-07-31; 修订日期: 2022-09-23

作者信息: 薄谋 兰州大学哲学社会学院
bomou@lzu.edu.cn

基金项目: 国家社科基金一般项目“算术与集合的实在论解释”(22BZX128)。

1 斯特恩第一次论证反机械主义论题

斯特恩既支持哥德尔的观点，又支持彭罗斯等人的观点。他采取的策略如下：首先，形式化机械主义；其次，解决语义和内涵悖论。在此基础上，斯特恩尝试表明心灵不是机器的论题。斯特恩把哥德尔析取式论题转化为条件句：如果心灵是机器，那么存在人类心灵既不能证明为真又不能证明为假的语句。换句话说，如果人类心灵是机器，那么存在绝对不可判定的语句。科尔纳表明从哥德尔不完全性定理出发，人们确实能得到哥德尔析取式论题，但无论如何得不到第一个析取支。彭罗斯特别重视绝对可证明性概念。科尔纳把可证明性分为相对可证明性与绝对可证明性。绝对可证明性表达的正是人类心灵产生数学定理的过程。

由此，斯特恩把心灵是否是机器的问题，转化为绝对可证明语句能否由能行算法生成的问题。彭罗斯的新论证似乎找到代表绝对可证明性的可行性假设，不过科尔纳表明这些假设是联合不协调的（jointly inconsistent），它们会导致内涵悖论。斯特恩的内涵悖论解决方案的特点在于，首先把内涵悖论同语义悖论联系起来，其次假设真悖论是所有这些悖论的来源。由此，斯特恩认为选择哪种真理论至关重要。他选择的真理论是克里普克-费弗曼（Kripke-Feferman）真理论，也被称为费弗曼-莫德林（Feferman-Maudlin）真理论。（[3, 15]）斯特恩认为，如果克里普克-费弗曼真理论是真悖论的解决方案，那么在形式化心灵不是机器的论题的前提下，人们就能证明该论题成立。（[29]）

根据麦希尔（J. Myhill）和蒙塔古（R. Montague）的结果，如果人们把模态概念和内涵概念当作谓词处理，那么就会出现悖论。（[16, 17]）如果人们把绝对可证明性概念当作谓词处理，那么也会出现悖论。如果人们用语句谓词 K 形式化绝对可证明性的两个构成性原则，那么它们是联合不协调的。它们分别是隐真原则和必然原则。前者说的是如果 $K \ulcorner \phi \urcorner$ ，那么 ϕ 。后者说的是如果 ϕ 是定理，那么 $K \ulcorner \phi \urcorner$ 是定理。导致悖论的原因在于两条原则隐性地假设素朴真谓词。为解决悖论人们需要显性地引入真谓词。人们需要用非素朴真谓词替换素朴真谓词，以便 ϕ 和 $T \ulcorner \phi \urcorner$ 不再互换。¹

对两条原则改造的结果就是两条新原则：首先是显真原则，说的是对所有语句而言，如果某个语句是绝对可证明的，那么它是真的；其次是真必然原则，说的是如果 $T \ulcorner \phi \urcorner$ 是定理，那么 $K \ulcorner \phi \urcorner$ 也是定理。悖论是否出现仅仅取决于人们采用何种真理论。斯特恩认为，如果人们采用的真理论是协调的，那么绝对证真理论也是协调的，不会再出现内涵悖论。我们把由此产生的策略称为内涵真策略：第一种策略是科尔纳策略；第二种策略是斯特恩策略。他们的共同点在于通过把内涵概念与非素朴真概念联系起来，从而解决内涵悖论。

¹真原则说的就是，如果某个哥德尔语句是绝对可证明的，那么它是真的。

科尔纳把真理论分为类型化 (typed) 真理论与免类型 (type-free) 真理论, 与之对应的是, 斯特恩把真理论分为素朴真理论与非素朴真理论。科尔纳使用内涵真策略有两个目的: 首先解决内涵悖论; 其次判定反机械主义论题是否有效。科尔纳表明, 如果人们采用费弗曼的确定真理论, 那么反机械主义论题将不会取得成功。这是因为机械主义论题与确定真理论是协调的。科尔纳实际上表明机械主义论题是独立的, 说的是人们既不能证明它为真, 又不能证明它为假。

在科尔纳的启发下, 斯特恩使用克里普克-费弗曼真理论替代费弗曼的确定真理论, 以检验反机械主义论题是否成立。斯特恩采用克里普克-费弗曼真理论中的两条原则: 首先是真谓词析取分配原则; 其次是真消除原则。斯特恩把通过显真原则、真谓词析取分配原则和真消除原则扩张皮亚诺算术的理论, 称为绝对证真理论。² 也就是说, 绝对证真理论是由皮亚诺算术、显真原则、真谓词析取分配原则和真消除原则构成的。绝对证真理论的优势在于它是协调的。人们能从克里普克-费弗曼真理论的协调性, 推出绝对证真理论的协调性。([29], 第 84 页)

我们来看机械主义论题经历的演变。首先, 在卢卡斯-彭罗斯论证框架下, 机械主义论题说的是, 理想化人类心灵能证明的数学定理是由能行程序生成的。其次, 如果人们用“图灵机”替换“机器”, 那么机械主义论题说的是, 理想化人类心灵能产生的定理是图灵机的输出。再次, 如果人们启用绝对可证明性概念, 机械主义论题说的是人们能递归枚举绝对可证明语句集。换句话说, 存在能证明所有绝对可证明语句的显性公理与规则系统。反机械主义论题说的是, 不存在能证明所有绝对可证明语句的显性公理与规则系统。

最后, 系统是由逻辑定律的地位决定的。斯特恩把自己的立场称为正统逻辑, 说的是如果人们采用经典逻辑, 那么所有经典逻辑真在选定系统中都是可证明的, 不管它是何种系统。斯特恩把科尔纳的立场称为非正统逻辑, 说的是逻辑没有特殊地位, 人们不能假设某一门特殊逻辑的逻辑真在每个形式系统中是可证明的。([29], 第 84-85 页) 斯特恩选用正统立场且把经典逻辑当作他的逻辑。由此, 机械主义论题说的是, 人们能在经典逻辑中从递归语句集出发证明绝对可证明语句。

我们来看斯特恩的句法规定。他使用 K 表示绝对可证明性, 并且使用 Σ 表示理论 $\mathcal{T}$ 的递归公理集, 同时使用 σ 表示 Σ 的自然表示, 而且使用 Pr_σ 表示 $\mathcal{T}$ 的自然可证明性谓词。斯特恩把机械主义论题表示为 $\exists\sigma\forall x(Kx \leftrightarrow Pr_\sigma(x))$, 说的是所有绝对可证明语句都是递归可枚举的。他把反机械主义论题表示为 $\neg\exists\sigma\forall x(Kx \leftrightarrow Pr_\sigma(x))$, 说的是并非所有绝对可证明语句都是递归可枚举的。

斯特恩使用归谬法证明反机械主义论题。他首先假设 $\forall x(Kx \leftrightarrow Pr_\sigma(x))$, 其

²我们把斯特恩的“真与绝对可证明性理论”简称为“绝对证真理论”, 其他人的理论继续称为“真与绝对可证明性理论”。

次从假设出发推出矛盾。³然而,在绝对证真理论中,人们不能表明机械主义论题是假的,也就是说,人们无法表明反机械主义论题是真的。根据克里普克,机械主义论题是无根语句(an ungrounded sentence, [14])。既非机械主义论题,又非反机械主义论题,在绝对证真理论中能被证明为真。这种情况表明,在克里普克-费弗曼真理论中,可证明性与真两者相分离,实际上可证明性远超真。

2 科尔纳表明斯特恩第一次论证无效

科尔纳认为斯特恩的论证无效。他的基本判断是,从哥德尔思考反机械主义论题开始,历经长时期的发展,人们不可能如此轻易地证出难度如此大的问题。事实上科尔纳对斯特恩的反驳有效。科尔纳也承认斯特恩对能否解决该问题持谨慎态度。科尔纳曾经证明在费弗曼的确定真理论下彭罗斯的新论证无效。([10, 13])换句话说,在确定真理论下,机械主义论题有效,而反机械主义论题无效。科尔纳对反机械主义者提出的挑战在于,为人们能证明反机械主义论题,找到可行的真与绝对可证明免类型理论。([11])斯特恩持反机械主义立场,他接受了科尔纳的挑战。斯特恩论证是卢卡斯-彭罗斯论证的延续,我们可以称之为卢卡斯-彭罗斯-斯特恩论证。

为了得到同科尔纳相反的结论,斯特恩需要对科尔纳论证进行改造。他的改造分两步进行:首先,他用克里普克-费弗曼真理论取代确定真理论;其次,他用新反机械主义论题取代科尔纳的旧反机械主义论题。科尔纳认为斯特恩的形式论证是正确的。问题可能出在两个地方。科尔纳的分析如下:首先,他认为用一种免类型真理论取代另一种免类型真理论,不会引起麻烦;其次,他认为斯特恩的问题出在用新反机械主义论题取代旧反机械主义论题。([11])

关于第一点,科尔纳和斯特恩都注意到人们不能有 $K(\ulcorner \lambda \vee \neg \lambda \urcorner)$, 因为由此人们会推出 $T(\ulcorner \lambda \vee \neg \lambda \urcorner)$, 这说的是不确定语句是真的,会导致矛盾。两人的解决办法都是修改原系统使得 K 仅仅对确定语句成立。反机械主义论题,不管在科尔纳系统,还是在斯特恩系统都是可证明的。这从而说明,反机械主义论题从一个系统转移到另一个系统保持不变。关于第二点,他们得到的结果不同。科尔纳得到的结果是,旧反机械主义论题是独立的,而斯特恩得到的结果是,新反机械主义论题是可证明的。科尔纳和斯特恩达成的共识是绝对可证明性仅仅对确定语句成立,而对不确定语句不成立。问题出在斯特恩的新机械主义论题上,因为它仅

³证明:使用反证法。假设 $\forall x(Kx \leftrightarrow Pr_{\sigma}(x))$ 。根据显真原则,人们有 $\forall y(Pr_{\sigma}(x) \rightarrow T(y))$ 。根据全称例化,人们有 $Pr_{\sigma}(\ulcorner \lambda \vee \neg \lambda \urcorner) \rightarrow T(\ulcorner \lambda \vee \neg \lambda \urcorner)$ 。由于 $\lambda \vee \neg \lambda$ 是重言式,人们有 $Pr_{\sigma}(\ulcorner \lambda \vee \neg \lambda \urcorner)$ 。使用假言推理,人们有 $T(\ulcorner \lambda \vee \neg \lambda \urcorner)$ 。根据真谓词析取分配原则,人们有 $T\ulcorner \lambda \urcorner \vee T\ulcorner \neg \lambda \urcorner$ 。根据说谎者语句,左析取支等价于 $\neg \lambda$ 。根据真消除原则,右析取支也等价于 $\neg \lambda$ 。根据说谎者语句,人们有 $T\ulcorner \lambda \urcorner$ 。根据真消除原则人们有 λ 。与 $\neg \lambda$ 矛盾!

仅考虑能输出所有一阶经典有效式的系统，而不确定语句就属于它们。包含不确定语句在内的系统，当然无法模仿绝对可证明性，因为绝对可证明性对任意的不确定命题都无效。

科尔纳对斯特恩的最终判定有两点：首先，在用克里普克-费弗曼真理论替代确定真理论的前提下，且用新机械主义论题替代旧机械主义论题的前提下，斯特恩得到心灵不是机器的结论，实际上这个结果很弱；其次，由于斯特恩的新机械主义论题包含不确定语句，最终导致他的结论无效。因此，科尔纳认为，人们需要继续寻找能证明反机械主义论题的可行免类型绝对证真理论。([11])

3 斯特恩第二次论证反机械主义论题

斯特恩关注彭罗斯的新论证 (Penrose's New Argument)，而且研究人们在使用真与绝对可证明性理论的情况下，能否以可靠的和连贯的方式形式化彭罗斯的新论证的问题。斯特恩分两步展开他的论证。首先他表明不存在形式化彭罗斯的新论证的协调理论。⁴ 其次他考虑彭罗斯的整体策略且提供合理的绝对证真理论，以表明彭罗斯的新论证是可靠的论证。然而，由于彭罗斯的新论证依赖理论的病态 (pathological) 特征，他的论证在直观上无法实现。⁵ 斯特恩把哥德尔析取式论题重述为：或者绝对可证明语句集不是递归可枚举的，或者存在并非绝对可证明的真数学语句。卢卡斯和彭罗斯认为，人们可以从哥德尔不完全性定理出发，推断哥德尔析取式的第一个析取支成立，也就是说，绝对可证明语句集不是递归可枚举的。

斯特恩把卢卡斯的论证称为第一个论证，而把彭罗斯的新论证称为第二个论证。⁶ 卢卡斯论证如下：如果任意递归可枚举系统 F 从未能证明它的哥德尔语句，那么人类心灵常常能察觉到哥德尔语句为真。也就是说，哥德尔语句对 F 来说是绝对可证明的，但在 F 中不是可证明的。据此，卢卡斯得出的结论是人类心灵远超每个递归可公理化数学理论证明概念。然而，为得到该结论，卢卡斯需要假设，人类心灵能察觉到 F 是协调的，也就是说， F 是协调的，是绝对可证明的。这个假设是不可行的，因为人类心灵难以察觉每个递归可公理化数学理论的协调性。

彭罗斯看到，既然问题与绝对可证明性概念有关，那么他要找出有关绝对可证明性概念的基本假设。斯特恩的做法是：在假设彭罗斯的绝对可证明性假设成立的情况下，探寻彭罗斯的新论证是否是支持第一个析取支的连贯论证。据我们所知，彭罗斯曾经在两个地方详细呈现过他的论证过程。([20, 21]) 斯特恩把彭

⁴科尔纳认为哥德尔析取式论题在绝对可证明性确定真理论下是可形式化的且可证明的。

⁵科尔纳认为彭罗斯的新论证在绝对可证明性确定真理论下无效。

⁶科尔纳把卢卡斯和彭罗斯 ([18]) 的论证称为第一代论证，把彭罗斯 ([19, 20]) 的新论证称为第二代论证。

罗斯的新论证分为六步：第一步，如果我是系统 F ，那么系统 F 是可靠的；第二步，在 F 的基础上，加上语句“我是 F ”， F' 仍然是可靠的；第三步，我察觉到，从我是 F 的假设出发，能推断哥德尔语句 $G(F')$ 为真；第四步，我察觉到 $G(F')$ 不是 F' 的后承；第五步，我刚刚察觉到“如果我碰巧是 F ，那么 $G(F')$ 为真”；第六步，由于我是 F ，我推断我毕竟不是 F 。⁷ 审视整个证明过程，它会用到可靠性概念⁸、真规则⁹和必然规则¹⁰。不幸的是，这两条规则是联合不协调的。([1, 16, 17, 24]) 这实际上建议人们需要找到联合协调的绝对可证明假设。人们有时也把这个联合不协调结果称为内涵悖论。

哥德尔曾建议解决内涵悖论是确立反机械主义论题的关键。斯特恩尝试把内涵悖论解决方案应用到彭罗斯的新论证，以便为彭罗斯的新论证辩护，也为哥德尔的建议辩护。人们对内涵悖论通常有两种解决方案：首先，把概念的原则或者模式限制到特殊语句集，把好的实例从坏的实例中挑选出来，由此避免悖论；其次，为自用 (self-applicable) 概念选择弱的原则或者模式，这些弱原则或者模式是联合协调的。斯特恩认为这两种方案都不太可行。前者的两种子方案都不可行¹¹，后者太弱以致无法复活彭罗斯的新论证。既然常规的解决方案不可行，斯特恩尝试一种新的解决方法。他的想法是把内涵悖论与语义悖论捆绑到一起，使得内涵概念的悖论性 (paradoxicality) 仅仅依赖真概念的悖论性。由此人们可以使用真谓词表述绝对可证明性的原则。真原则就变为对所有语句来说，如果某个语句是绝对可证明的，那么该语句是真的。¹² 如果修正版绝对可证明性原则与协调真理论相结合，那么人们能得到协调的真与绝对可证明性理论。([12, 25–27])

斯特恩正是在此框架下工作，且研究人们能否为第一个析取支提供连贯 (coherent) 论证。科尔纳从必然规则、绝对可证明真规则和真引入规则¹³ 出发，形式化彭罗斯的新论证。¹⁴ 由此科尔纳得出第一个析取支是独立的结论。斯特恩认为

⁷证明：1. 根据 T ，第一步成立。也就是，如果 $K \vdash \phi \rightarrow \phi$ ，那么 $\forall x(Fx \leftrightarrow Kx) \rightarrow (F \vdash \phi \rightarrow \phi)$ 。根据演绎定理， $F'(\vdash \phi) \leftrightarrow F(\vdash \forall x(Fx \leftrightarrow Kx) \rightarrow \phi)$ 。2. 根据弱化规则和假言推理，人们得到第二步。也就是， $\forall x(Fx \leftrightarrow Kx) \rightarrow (F' \vdash \phi \rightarrow \phi)$ 。3. 通过把第二步的模式例化到 F' 的哥德尔语句，人们得到第三步。也就是， $\forall x(Fx \leftrightarrow Kx) \rightarrow G(F')$ 。4. 根据哥德尔语句的构造，第三步等价于第四步。第四步有两个版本。也就是， $\forall x(Fx \leftrightarrow Kx) \rightarrow \neg F' \vdash G(F')$ ， $\forall x(Fx \leftrightarrow Kx) \rightarrow \neg F \vdash \forall x(Fx \leftrightarrow Kx) \rightarrow G(F')$ 。5. 根据绝对必然性必然规则，人们从第三步得到第五步。也就是， $K \vdash \forall x(Fx \leftrightarrow Kx) \rightarrow G(F')$ 。6. 根据经典逻辑规则，从第五步得到第六步。第六步有两个子步骤。也就是，首先， $\forall x(Fx \leftrightarrow Kx) \rightarrow F \vdash \forall x(Fx \leftrightarrow Kx) \rightarrow G(F')$ ，其次，这同第四步矛盾，由此得到 $\forall x(Fx \leftrightarrow Kx) \rightarrow \perp$ 。证毕！

⁸系统 F 是可靠的当且仅当所有它的定理都是真的。它的元语言版本是 $F \vdash \phi \rightarrow \phi$ 。

⁹绝对可证明的每个语句都是真的。也就是， $K \vdash \phi \rightarrow \phi$ 。

¹⁰如果人类心灵能产生定理，那么人类心灵能产生绝对可证明的定理。也就是， $\phi \rightarrow K \vdash \phi$ 。

¹¹第一个子方案是把原则或者模式限制到算术语句，但这会导致彭罗斯论证的部分步骤无效。第二个子方案是把语句分为悖论语句和非悖论语句，但这会削弱他论证中的归谬法策略。

¹²表达式从 $K \vdash \phi \rightarrow \phi$ 变为 $\forall x(Kx \rightarrow Tx)$ 。

¹³如果 ϕ ，那么 $T \vdash \phi$ 。也就是， $\phi \rightarrow T \vdash \phi$ 。

¹⁴证明：1. 根据绝对可证明真规则， $\forall x(Fx \leftrightarrow Kx) \rightarrow (Fx \rightarrow Tx)$ 。2. 根据真引入规则， $\forall x(Fx \leftrightarrow$

科尔纳在新框架下没有认真检验彭罗斯的新论证。实际上,在新框架下彭罗斯的新论证会有新的问题出现:首先,从第一步到第二步的推理会出现麻烦。其次,人们需要新的假设才能推出第三步。斯特恩通过两步来修补彭罗斯的新论证:首先,尽管形式系统 F 在假言推理下是封闭的,但人们不知道真谓词是否在假言推理下是封闭的。为保证真谓词在假言推理下是封闭的,人们需要引入新的假设。我们把这条假设称为真蕴涵规则。¹⁵有了这条规则,人们就可以再次从第一步推出第二步。其次,为推出第三步,人们需要假设没有假算术语句是真的。实际上这就是塔斯基双条件规则。¹⁶有了这条规则,人们就可以再次推出第三步。¹⁷由此斯特恩完成对科尔纳论证的修补。

然而,弗里德曼(H. Friedman)和希尔德(M. Sheard)表明,真引入原则、真蕴涵原则和塔斯基双条件原则是联合不协调的。([5])问题出在彭罗斯的原论证似乎依赖素朴真概念。既然不能依赖素朴真概念,人们可以采用类型化真原则。斯特恩使用简单化类型改造绝对可证明真原则、真引入原则和必然原则。这里再次出现修正版原则联合不协调的情况。斯特恩表明新绝对可证明真原则、新真引入原则、真蕴涵原则、新必然原则和塔斯基双条件原则是联合不协调的。([28], 第30页)他得出的结论是彭罗斯的新论证是不连贯的。既然彭罗斯的新论证行不通,人们自然要寻找可替代论证。也就是,人们要找到合理的真与绝对可证明性理论,以产生支持第一个析取支的论证。科尔纳诉诸费弗曼的确定真理论,而斯特恩诉诸彭罗斯的归谬策略(reductio strategy)。斯特恩需要形式系统 F 是经典递归可公理化理论。这产生两个后果:首先,“我是 F ”变为 $\forall x(Pr_F(x) \leftrightarrow Kx)$, Pr_F 是理论 F 的自然可证明性谓词;其次,它会加强机械主义论题,使得人们只考虑输出在经典逻辑下封闭的图灵机。

由此人们似乎会得到彭罗斯的归谬策略。绝对可证明真原则蕴涵全局反射原则(Global Reflection Principle) $\forall x(Pr_F(x) \leftrightarrow Tx)$ 。根据哥德尔第二不完全性定理,没有递归可公理化理论 F 能证明全局反射原则成立。由于“我是 F ”连同绝对可证明性真理论蕴涵全局反射原则,而后者蕴涵矛盾。这似乎实现了彭罗斯的

$Kx) \rightarrow (F'x \rightarrow Tx)$ 。3. 根据 $G(F')$ 的定义, $\forall x(Fx \leftrightarrow Kx) \rightarrow G(F')$ 。4. 再次根据 $G(F')$ 的定义, $\forall x(Fx \leftrightarrow Kx) \rightarrow \neg F' \neg G(F') \neg$ 。5. 根据 4, 再根据必然规则, $K \neg \forall x(Fx \leftrightarrow Kx) \rightarrow G(F') \neg$ 。6. 根据 4 和 F' 的定义, $\forall x(Fx \leftrightarrow Kx) \rightarrow \neg F' \neg \forall x(Fx \leftrightarrow Kx) \rightarrow G(F') \neg$ 。7. 根据 5, $\forall x(Fx \leftrightarrow Kx) \rightarrow F' \neg \forall x(Fx \leftrightarrow Kx) \rightarrow G(F') \neg$ 。8. 根据 6 和 7, $\forall x(Fx \leftrightarrow Kx) \rightarrow \perp$ 。

¹⁵真蕴涵规则的表达式是 $\forall x(Sent(x \rightarrow y) \rightarrow (T(x \rightarrow y) \rightarrow (Tx \rightarrow Ty)))$ 。

¹⁶塔斯基双条件句是 $T \neg \phi \neg \leftrightarrow \phi$ 。

¹⁷证明: 1. 根据绝对可证明真规则, $\forall x(Fx \leftrightarrow Kx) \rightarrow (Fx \rightarrow Tx)$ 。2. 根据真引入规则, 再根据真蕴涵规则, $\forall x(Fx \leftrightarrow Kx) \rightarrow (F'x \rightarrow Tx)$ 。2a. 根据替代规则, $\forall x(Fx \leftrightarrow Kx) \rightarrow (F' \neg G(F') \neg \rightarrow T \neg G(F') \neg)$ 。2b. 根据塔斯基双条件, $\forall x(Fx \leftrightarrow Kx) \rightarrow (F' \neg G(F') \neg \rightarrow G(F') \neg)$ 。3. 根据 2b, 再根据 $G(F')$ 的定义, $\forall x(Fx \leftrightarrow Kx) \rightarrow G(F')$ 。4. 再次根据 $G(F')$ 的定义, $\forall x(Fx \leftrightarrow Kx) \rightarrow \neg F' \neg G(F') \neg$ 。5. 根据 4, 再根据必然规则, $K \neg \forall x(Fx \leftrightarrow Kx) \rightarrow G(F') \neg$ 。6. 根据 4 和 F' 的定义, $\forall x(Fx \leftrightarrow Kx) \rightarrow \neg F' \neg \forall x(Fx \leftrightarrow Kx) \rightarrow G(F') \neg$ 。7. 根据 5, $\forall x(Fx \leftrightarrow Kx) \rightarrow F' \neg \forall x(Fx \leftrightarrow Kx) \rightarrow G(F') \neg$ 。8. 根据 6 和 7, $\forall x(Fx \leftrightarrow Kx) \rightarrow \perp$ 。

归谬策略，且表明第一个析取支成立。斯特恩指出该策略的两个问题：首先，人们能证明，如果 F 证明全局反射原则，那么 F 是不协调的，并不意味着人们能证明， F 证明全局反射原则蕴涵矛盾。其次，在执行归谬策略的过程中，有人不能假设他实际上在 F 中推理，因为这不仅预设“我是 F ”，而且预设我知道我是哪个系统 F 。这意味着即使理想化人类心灵与形式系统 F 相符，系统 F 仍不能把可证明性谓词当作它的可证明性谓词。由此在归谬论证中，人们不能假设洛布可推导性条件（Löb's derivability conditions, [2]）。否则彭罗斯的新论证可能会是连贯论证。¹⁸

然而，斯特恩认为利用洛布可推导性条件得到的彭罗斯论证，也不是令人信服的论证。因为它实际上确立的是，我是 F 且我知道我是 F 是不可能的，而非确立哥德尔析取式论题的第一个析取支。问题的关键是，在不引入关于递归可公理化理论的任意特殊假设的情况下，人们能从全局反射原则推出矛盾。斯特恩尝试在弱假设下，从真与绝对可证明性理论出发，找到执行彭罗斯归谬策略的办法。他找到的真与绝对可证明性理论是协调克里普克-费弗曼理论。它有两个特点：首先，除了否定，它满足所有逻辑联结词的交换律；其次，它有真消除原则¹⁹。令 λ 是标准说谎者语句。²⁰ 斯特恩尝试在协调克里普克-费弗曼理论下，论证反机械主义论题成立。²¹

此论证的关键步骤是第三步，它说的是，在没有任何前提的情况下 $\lambda \vee \neg \lambda$ 是可证明的。这表明对任意经典递归可公理化理论 F , $Pr_F(\ulcorner \lambda \vee \neg \lambda \urcorner)$ ，也就是说，重言式在 F 下是可证明的。由此，斯特恩认为该论证确保反机械主义论题成立。人们质疑该论证是否是好论证。这实际上取决于两个要素：首先，他是否把协调克里普克-费弗曼理论当作适当理论；其次，他是否把该论证当作直观可行的论证。斯特恩认为答案都是否定的：首先，相对于每个递归可公理化理论的标准，协调克里普克-费弗曼理论是可证明不可靠的（provably unsound）。而如果有人要讨论哥德尔析取式论题，他需要可靠理论。其次，由于人们选用的理论是病态理论，这会削弱论证的可信度。斯特恩的结论是，人们总是试图利用哥德尔不完全性定理证明反机械主义论题成立，然而屡屡碰壁。他的建议是，人们可以在完全放弃哥

¹⁸证明：1. $\forall x(Fx \leftrightarrow Kx) \rightarrow (F\ulcorner \phi \urcorner \rightarrow \phi)$ 。2. $\forall x(Fx \leftrightarrow Kx) \rightarrow (F'\ulcorner \phi \urcorner \rightarrow \phi)$ 。3. $\forall x(Fx \leftrightarrow Kx) \rightarrow G(F')$ 。4. $\forall x(Fx \leftrightarrow Kx) \rightarrow \neg F'\ulcorner \forall x(Fx \leftrightarrow Kx) \urcorner \rightarrow G(F')\urcorner$ 。5. 根据 3 和洛布可推导性条件， $K\ulcorner \forall x(Fx \leftrightarrow Kx) \urcorner \rightarrow G(F')\urcorner$ 。6. 根据 4 和 5, $\forall x(Fx \leftrightarrow Kx) \rightarrow \perp$ 。证毕！

¹⁹真消除原则是 $T\ulcorner \phi \urcorner \rightarrow \phi$ 。

²⁰说谎者语句是 $\neg T\ulcorner \lambda \urcorner \leftrightarrow \lambda$ 。

²¹证明：1. 根据绝对可证明真原则， $F \vdash \forall x(K(x) \leftrightarrow Pr_F(x)) \rightarrow \forall y(Pr_\sigma(y) \rightarrow T(y))$ 。2. 根据替代原则， $F \vdash \forall x(K(x) \leftrightarrow Pr_F(x)) \rightarrow (Pr_F(\ulcorner \lambda \vee \neg \lambda \urcorner) \rightarrow T(\ulcorner \lambda \vee \neg \lambda \urcorner))$ 。3. 根据重言式， $F \vdash Pr_\emptyset(\ulcorner \lambda \vee \neg \lambda \urcorner)$ 。4. 根据 2 和 3，再根据假言推理， $F \vdash \forall x(K(x) \leftrightarrow Pr_F(x)) \rightarrow T(\ulcorner \lambda \vee \neg \lambda \urcorner)$ 。5. 根据析取分配律， $F \vdash \forall x(K(x) \leftrightarrow Pr_F(x)) \rightarrow T(\ulcorner \lambda \urcorner) \vee T(\ulcorner \neg \lambda \urcorner)$ 。6. 根据说谎者语句和真消除原则， $F \vdash \forall x(K(x) \leftrightarrow Pr_F(x)) \rightarrow \neg \lambda$ 。7. 根据说谎者语句， $F \vdash \forall x(K(x) \leftrightarrow Pr_F(x)) \rightarrow T\ulcorner \lambda \urcorner$ 。8. 根据真消除原则， $F \vdash \forall x(K(x) \leftrightarrow Pr_F(x)) \rightarrow \lambda$ 。9. 根据 6 和 8, $F \vdash \neg \forall x(K(x) \leftrightarrow Pr_F(x))$ 。

德尔不完全性定理的情况下,去探索直接证明反机械主义论题的可能性。([28],第35–36页)

4 我们表明斯特恩第二次论证无效

我们看到斯特恩先后为反机械主义论题给出两个论证。([28, 29])我们首先来看两个论证的共同点和不同点。它们采用的理论都是真与绝对可证明性理论,是真理论与绝对可证明理论的结合体。第一论证采用的是克里普克–费弗曼真理论,第二论证采用的是协调克里普克–费弗曼真理论。协调克里普克–费弗曼真理论承认谓词的协调性,在克里普克–费弗曼理论中等价于真消除原则。也就是说,协调克里普克–费弗曼真理论自身包含真消除原则。两个论证都使用真谓词析取分配律。它们都使用绝对可证明真原则。它们也都使用说谎者语句。

两个论证的最大差异体现在对机械主义论题的表述上。第一论证把机械主义论题表述为 $\forall x(Pr_{\sigma}(x) \leftrightarrow Kx)$, σ 是某个理论的递归可枚举公理集的表示。第二论证把机械主义论题表述为 $\forall x(Pr_F(x) \leftrightarrow Kx)$, F 是经典递归可公理化理论。前者是通用机械主义论题,后者是强机械主义论题,说的是在假设经典逻辑的前提下,机械主义论题成立。科尔纳认为第一论证的问题就出在对理论的规定上。本来只有确定语句,后来却出现不确定语句。为避免这种情况出现,斯特恩对原条件进行限定,加入经典逻辑这一条款,以避免不确定语句的再次出现。斯特恩的努力表面上成功了。但是问题还是出现在对不确定语句的理解上。不确定语句指的是类似 $\lambda \vee \neg \lambda$ 的语句。在经典逻辑下,它是重言式,也就是一个恒真语句。但是在二值逻辑的情况下,它还可以表达或真或假的情况,而这表达的是不可判定的情况。当然斯特恩也意识到这个问题。他把这种情况称为内部逻辑与外部逻辑的分离。²²既然面临重重困难,我们认为斯特恩的第二论证即使是形式上可靠的论证,也终究不是一个好论证。由此我们再次回到科尔纳的裁定:人们需要继续寻找能证明反机械主义论题的可行免类型真与绝对可证明理论。

5 结论

反机械主义论题含有三个核心概念:理想化人类心灵、理想化有限机器和真。由于图灵贡献,理想化有限机器概念具备科学根据。真概念也得到了充足的发展。人们不仅有塔斯基的类型化真理论,而且有各种各样的免类型真理论。([8])在理想化有限机器概念和真概念具备科学根据的前提下,理想化人类心灵却不具备科学根据。这里有两个原因:首先,意识科学尚未真正发展起来([13, 18, 19]);

²²内部逻辑是真谓词辖域内部的逻辑,外部逻辑是经典逻辑。

其次，人们不仅对绝对可证明性概念认识不足，而且尚未有成熟的内涵悖论解决方案出现。〔28〕哥德尔建议人们从解决内涵悖论入手，进而证明反机械主义论题。由于没有令人满意的内涵解悖方案出现，人们实际上诉诸语义悖论解决方案，也就是诉诸真理论。

不管是科尔纳还是斯特恩，他们都试图借助于真理论证明或者否证反机械主义论题。科尔纳借助的是确定真理论，斯特恩第一论证借助的是克里普克-费弗曼真理论，斯特恩第二论证借助的是协调克里普克-费弗曼真理论。科尔纳表明，不管使用哪种免类型真理论，论证的效果都是一样的。真正的差别体现在对机械主义论题的表达上。科尔纳从免类型真理论出发，推断反机械主义论题是独立的，而斯特恩从免类型真理论出发，推断反机械主义论题成立。这实际上提醒我们，诉诸真理论试图证明或者否证反机械主义论题的方案并不可靠。人们还是要回到对理想化人类心灵概念也就是绝对可证明性概念的深度分析上面。

彭罗斯已经想到这一点并作出尝试，也就是为数学心灵提供计算模型甚至物理模型。〔22〕彭罗斯的想法源自图灵（A. Turing）的预言机，他把自己构想的机器称为谨慎预言机（cautious-oracle device），以期容纳绝对可证明概念。彭罗斯认为哥德尔的考虑并非周全，他认为哥德尔的问题在于并未把物理要素考虑进去，仅仅考虑数学因素和心灵因素。彭罗斯认为这种物理要素要具有非可计算的性质。他认为现有的物理学并不具有这种性质。人们需要发展一种新的物理学。我们从彭罗斯的视角来考虑心灵能否被机械化的问题。在薛定谔（E. Schrödinger）的思想实验中，一只猫的状态可能演变成这只猫既生又死的状态。这实际上说的是两个可替代实在事物可以同时出现，比如量子粒子可以同时占据两个不同位置。我们在考虑机械主义论题时也出现了这种情况。这是不完全性现象在不同科学领域的体现。

既然不完全现象普遍存在，我们不会考虑斯特恩的建议，因为他建议人们绕开不完全性现象去证明或者否证反机械主义论题。这里我们再次审视彭罗斯的建议，他把各种各样的柏拉图主义称为程度柏拉图主义（degrees of Platonism）。如果有人要解决连续统假设（the continuum hypothesis）这样的难题，他需要持强柏拉图主义立场。如果有人要解决反机械主义论题，他只需要持弱柏拉图主义立场。我们的看法是即使要解决反机械主义论题，人们也要持强柏拉图主义立场。哥德尔曾经告诉我们，尽管科恩（P. Cohen）已经证明连续统假设是不可判定的，但这并没有解决数学真的问题。〔6〕如果要解决数学真的问题，人们需要持一种强柏拉图主义立场。这种立场在希尔伯特（D. Hilbert）的世界是数学自治（mathematical autonomy, [4]），在科尔纳的世界是数学牵引（mathematical traction, [23]）。

参考文献

- [1] D. Chalmers, 1995, “Mind, machines, and mathematics. A review of shadows of the mind by Roger Penrose”, *Psyche*, **2(9)**: 11–20.
- [2] M. Detlefsen, 1986, *Hilbert’s Program: An Essay on Mathematical Instrumentalism*, Dordrecht: Reidel.
- [3] S. Feferman, 1991, “Reflecting on incompleteness”, *The Journal of Symbolic Logic*, **56(1)**: 1–49.
- [4] C. Franks, 2009, *The Autonomy of Mathematical Knowledge: Hilbert’s Program Revisited*, New York: Cambridge University Press.
- [5] H. Fridman and M. Sheard, 1987, “An axiomatic approach to self-referential truth”, *Annals of Pure and Applied Logic*, **33**: 1–21.
- [6] K. Gödel, 1990, “What is Cantor’s continuum problem?”, in S. Feferman *et al.* (eds.), *Kurt Gödel Collected Works, Volume II: Publications*, pp. 176–189 + 254–271, Oxford: Oxford University Press.
- [7] K. Gödel, 1995, “Some basic theorems on the foundations of mathematics and their implications”, in S. Feferman *et al.* (eds.), *Kurt Gödel Collected Works, Volume III: Unpublished Essays and Lectures*, pp. 304–323, Oxford: Oxford University Press.
- [8] V. Halbach, 2011, *Axiomatic Theories of Truth*, Cambridge: Cambridge University Press.
- [9] L. Horsten and P. Welch, 2016, *Gödel’s Disjunction: The Scope and Limits of Mathematical Knowledge*, Oxford: Oxford University Press.
- [10] P. Koellner, 2016, “Gödel’s disjunction”, in H. L. and P. Welch (eds.), *Gödel’s Disjunction: The Scope and Limits of Mathematical Knowledge*, pp. 148–188, Oxford: Oxford University Press.
- [11] P. Koellner, 2018, “On a purported proof that the mind is not a machine”, *Thought: A Journal of Philosophy*, **7(2)**: 91–96.
- [12] P. Koellner, 2018, “On the question of whether the mind can be mechanized, I: from Gödel to Penrose”, *The Journal of Philosophy*, **115(7)**: 337–360.
- [13] P. Koellner, 2018, “On the question of whether the mind can be mechanized, II: Penrose’s new argument”, *The Journal of Philosophy*, **115(9)**: 453–484.
- [14] S. Kripke, 1975, “Outline of a theory of truth”, *The Journal of Philosophy*, **72(19)**: 690–716.
- [15] T. Maudlin, 2004, *Truth and Paradox: Solving the Riddles*, Oxford: Oxford University Press.
- [16] R. Montague, 1963, “Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability”, *Proceedings of a Colloquium on Modal and Many-valued Logics, Helsinki*, 23–26 August, 1962, pp. 153–167, Acta philosophica Fennica, No. 16, Helsinki.
- [17] J. Myhill, 1960, “Some remarks on the notion of proof”, *The Journal of Philosophy*, **57(14)**: 461–471.
- [18] R. Penrose, 1989, *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*, Oxford: Oxford University Press.

- [19] R. Penrose, 1994, *Shadows of the Mind: An Approach to the Missing Science of Consciousness*, Oxford: Oxford University Press.
- [20] R. Penrose, 1996, “Beyond the doubting of a shadow”, *Psyche*, **2(23)**: 89–129.
- [21] R. Penrose, 2011, “Gödel, the mind, and the laws of physics”, in M. Baaz *et al.* (eds.), *Kurt Gödel and the Foundations of Mathematics: Horizons of Truth*, pp. 339–358, Cambridge: Cambridge University Press.
- [22] R. Penrose, 2016, “On attempting to model the mathematical mind”, in S. Cooper (ed.), *The Once and Future Turing*, pp. 361–378, New York: Cambridge University Press.
- [23] C. Rittberg, 2015, “How Woodin changed his mind: New thoughts on the continuum hypothesis”, *Archive for History of Exact Sciences*, **69(2)**: 125–151.
- [24] S. Shapiro, 2003, “Mechanism, truth, and Penrose’s new argument”, *Journal of Philosophical Logic*, **32(1)**: 19–42.
- [25] J. Stern, 2014, “Modality and axiomatic theories of truth I: Friedman-Sheard”, *The Review of Symbolic Logic*, **7(2)**: 273–298.
- [26] J. Stern, 2014, “Modality and axiomatic theories of truth II: Kripke-Feferman”, *The Review of Symbolic Logic*, **7(2)**: 299–318.
- [27] J. Stern, 2016, *Toward Predicate Approaches to Modality*, Cham: Springer.
- [28] J. Stern, 2018, “Penrose’s new argument and paradox”, in M. Piazza (ed.), *Truth, Existence, and Explanation*, pp. 17–38, Switzerland: Springer.
- [29] J. Stern, 2018, “Proving that the mind is not a machine?”, *Thought: A Journal of Philosophy*, **7(2)**: 81–90.

(责任编辑：袁之)

The Modal Argument on Mind, Machine and Truth

Mou Bo

Abstract

Lukas and Penrose both think that from the Gödel's incompleteness theorems people can deduce the thesis of "the mind cannot be mechanized". This is a strong assertion. The weak assertion was proposed by Gödel. He thinks that from his theorems people can only conclude either "the mind cannot be mechanized" or "there exists an undecidable mathematical proposition". This disjunctive assertion is equivalent to "if the idealized human mind is the idealized finite machine, then people cannot decide whether this mathematical proposition is true or false". Gödel suggests people should justify the anti-mechanism thesis beginning from the intentional paradox. Koellner's success in mechanistic thesis motivates Stern try to prove the anti-mechanistic thesis from alternative theories of truth. Koellner then shows Stern's argument does not hold. After that Stern tries to improve on his previous work. We show Stern's revised argument also does not hold. Stern's two failed arguments stem from the fact that people have no method to recognize the notion of idealized human mind, i.e. the notion of absolute provability.