摘要

当人工智能开始产出人类无法逐行检验的数学证明时，“数学知识”的定义被迫从“可被独立重建的公开论证链”向“被信任的黑箱输出”滑动。本文追问的不是“AI能不能做数学”，而是这种滑动暴露了一个此前被数学哲学长期忽视的发生学结构：一条真正的知识证明，不仅在符号层面自治，还必须承载并改写一个将其绷紧的、由理念界、现实界与自我界交织而成的三界纠缠土壤。本文将AI的数学产出诊断为只满足单方程的“瘦发生”，并在与乔治·波利亚“合情推理”、汉斯·哈恩“同义反复”数学观、拉卡托斯“概念拉伸与反驳驱动修正”、以及Lakoff与Núñez“概念隐喻与数学认知基础”的正面交手中，提出一条核心判断：数学认知的发生，不可还原为对已存逻辑蕴含关系的解析或归纳性猜测，它是一份被三界追问绷紧的认知土壤，在自由律的驱动下，对于自身路径的不可逆重裁；这种重裁的标志，是提出者在回答三界追问时所做出的、无法被现有概念框架吸收的“改口”，这种改口不仅使概念本身不可逆地膨胀，更使修改者自身无法再以旧方式提问——这正是它与拉卡托斯的概念拉伸之间的分离线。本文进一步分离出“被绷紧的认知土壤”（Taut Cognitive Soil, TCS）作为控制变量，提出其与“推理链条的广度”这一第二轴构成的二维辨别格，并给出可裁决的经验判据。论文上半篇完成对代理变量的坍缩与控制变量的提取。

关键词：数学认知；证明纯粹性；发生学结构；被绷紧的认知土壤；合情推理；启发式搜索；概念拉伸

编者按（合稿说明）：本文为「上半篇」，同日有两个版本——初稿与打磨修改稿。本页正文取打磨修改稿，它比初稿多出两位正面交手的对象（拉卡托斯的概念拉伸、莱考夫与努涅斯的体验性认知基模），承重命题也由排除三个代理变量扩展为排除五个。打磨稿的末尾在传输中断开，第六节的「条件一（历史案例反向检验）」、第七节结语与参考文献表均未出现；第五节末尾的最后一句「……那么本论文的核心变量——被绷紧的认知土壤（TCS）作为数学认知发生学独立变量——」也断在此处。编辑部据同日初稿把这三部分接回，断句按初稿原文补完，第四、第五节的小节标题同样取自初稿；第四节末尾另有一处原稿断裂（「这是一种本体论框架的重新裁定」之后），按原句收束，未增补内容。参考文献表在初稿十四条之外，补入正文新引的拉卡托斯（1976）与莱考夫、努涅斯（2000）两条。除错字与标点订正外，正文未作改写。

一、引言：当证明的黑箱化击穿了认知契约

赖欣巴哈（Hans Reichenbach）曾对科学的合法性做过一个著名的内部—外部二分：发现的语境属于心理学，而辩护的语境属于逻辑。这个区分在二十世纪的数学哲学中被加固成一道常识的墙垣：一个命题的真值，独立于发现它的那个头脑，证明链条从公理出发，如磐石般可被任何理性主体检验。人工智能在数学领域的进展，正在悄无声息地击穿这道墙垣。当四色定理的证明第一次被交给计算机时，人们的焦虑还停留在“我们是否该信任机器的验算”。但今天，问题的性质变了。AI已经可以生成人类数学家短期内无法逐行验证的、长度惊人的证明。正确性不再是那条被公开检查的符号链，而变成了对一个黑箱模型输出结果的“信任”或“风险评估”。这一下子击穿了数学共同体内部一条从未被明说、却一直运作者的认知契约：一个数学主张若要被称为“知识证明”，其正确性必须建立在

一条可被任何同行独立重建的公开论证链之上。AI的产出在这里完成了第一次功能性替换：用“算法预测的正确性”置换了“知识证明的正确性”。面对此景，主流的辩护路径往往滑向功能主义：如果AI证明了黎曼猜想，并且货币化、工程化应用都确认它是正确的，那么它就是知识。但这里头藏着一个被跳过去的追问：那个被撤除的“公开论证链”，仅仅是一个验证工具，还是它本身就是数学认知的载体？本文将论证后一种立场：证明纯粹性不仅是一个逻辑要求，更是数学认知的发生学引擎——自由律（主体生成一条此前不存在的理解路径的行为）运行所必需的土壤。

我们必须先定位几位最深刻的“敌意近邻”，它们都从不同的角度试图将我们命名的“证明纯粹性”予以降解或重新解释，以此来瓦解“数学知识的认知核心正被黑箱化掏空”这个判断。

第一位是乔治·波利亚（George Polya）的合情推理（Plausible Reasoning）。波利亚将数学发现的过程描述为一系列“猜测—检验—修正”的归纳循环。在这个框架里，数学家那种“本该如此”的匿名必然性，不过是一种“成功概率极高的、基于类比与归纳的经验性猜测”。也就是说，真看见的“绷紧感”被还原为一种高效的模式识别与类比记忆。波利亚的代理变量，是“启发式模式匹配的成功概率”。用他的框架来解释AI：只要AI在更大规模的数据上训练出更强的模式匹配能力，其产出的“必然性”就应该与人类数学家的“直觉”在概念上趋于同构，那么所谓“自由律裁出新路径”这个独特发生，就仅仅是一种尚未被我们用算法充分建模的“启发式效率”。

第二位是汉斯·哈恩（Hans Hahn）等逻辑实证主义者所持的“数学即同义反复”观。在这个观点里，所有的数学知识都是逻辑公理与定义的隐蔽蕴含。证明，只是将这些蕴含关系展开的过程。因此，一条“新的理解路径”从未在存在论意义上被“创造”出来，它只是揭示了公理系统中一个我们此前未发现的逻辑变体。哈恩的代理变量，是“证明在逻辑系统中的可推导性”。用他的框架来重新描述AI的证明：AI的高效搜索，不过是在逻辑空间里快速定位了那个本就存在的推导链。当AI给出的证明人类无法理解时，问题不在于它复越了人类的认知功能，而在于人类的解析工具——比如“直觉”——跑得太慢。那么，所谓“发生学的土壤”，在此刻就成了一个冗余的装饰概念。

第三位，是定理自动证明（ATP）领域的先驱拉里·沃斯（Larry Wos）等人所建立的“启发式搜索”路径。它不跟你谈玄学。它对数学直觉的操作定义是：一套基于启发性规则的、对证明空间进行剪枝和探索的搜索算法。随着计算力的提升，更强的启发式函数可以被发现。沃斯路径的代理变量，是“搜索空间中被有效剪枝的分支数量”。其解释是：数学家的“艰难抉择”、“走哪条路”，本质上是在他大脑的神经网络里，被长期训练调优出的一个能量极低的搜索策略。

第四位，是伊姆雷·拉卡托斯（Imre Lakatos）在《证明与反驳》中提出的概念拉伸（Concept-stretching）与反驳驱动的修正机制。拉卡托斯通过对多面体欧拉公式证明史的精彩重构，揭示了数学概念如何在“证明—反例—修正”的辩证循环中被不可逆地修改：一个看似完备的证明引出局部反例，反例逼使数学家重新定义概念以排除异常，重新定义后的概念又引来新的反例，如此螺旋上升。这个机制极其强大，因为它将我们命名的“不可逆改口”置于一条可被理性重建的历史逻辑之内——改口不是神秘的天启，而是反驳压力下概念自我辩护的策略性调整。拉卡托斯的代理变量，是“反驳压力驱动的概念重定义”。若他的解释成立，则TCS所谓的“不可逆重裁”不过是理性辩证法的一个特例，不需要一个独立的发生学变量。

第五位，是乔治·莱考夫与拉斐尔·努涅斯（George Lakoff & Rafael Núñez）在《数学从何而来》中提出的概念隐喻与数学认知基础理论。他们从认知科学的实证研究出发，论证了数学理解根植于人类

普遍的体验性认知土壤——　例如，算术源于对物体集合的操作隐喻，逻辑源于对空间包含关系的图式，微积分源于对运动与流动的体验性理解。在这个框架里，数学家“被逻辑的必然性绷紧”的感觉，并非来自一个异质的三界追问，而是其长期习得的、映射自身体经验的隐喻网络在概念整合过程中产生的认知张力。这是一种体验性的绷紧感，而非发生学的绷紧。Lakoff与Núñez的代理变量，是“隐喻映射的基模化程度”，即数学概念的抽象程度取决于其赖以形成的体验性隐喻基模能被多深地整合进一个自洽的网络。若这一框架充分成立，那么本文以TCS为起点提出的“三界追问”，便可以被还原为“感受运动基模对抽象域的投射约束”———　土壤之喻在此失去了其存在论上的特殊性。

这些立场强大无比，它们在各自的代理变量范围内完全自治。但它们都无法解释一个“失灵点”：当一个数学物理学家如彼得·舒尔茨（Peter Scholze）将他自认为最重要的直觉——液体张量积的证明——交给Lean系统进行形式化时，他并非在测试一个高效的“启发式搜索”能否重建他的证明。他在做的事完全相反：他认识到自己无法将他“看见”的那个东西完全还原为已知搜索空间里的一个区域，但他又意识到他必须把它交出来，让它成为一个可被独立重建的“公共符号基础设施”。这正是自由律在极度绷紧的三界束缚下的经典运行：他没有把自己无法言传的直觉当作私人财产，而是以极端“自虐”的方式，将其外化。本文追踪的，就是那一个“他看见了”的东西，在被符号化之前，究竟在哪个结构位上。我们的命题将表明，“三界纠缠的绷紧”是这些代理变量都无法指代的、独立的控制变量。

二、代理坍缩：为何“启发式模式匹配”“推导性”“搜索剪枝”“概念拉伸”与“体验性认知基模”都解释了别的

当我们称一个数学家在证明构造中“被逻辑的必然性绷紧”，并且这种绷紧是不可还原到算法上的，我们在说什么？这个问题的答案不在波利亚、哈恩、沃斯、拉卡托斯或莱考夫与努涅斯的解释范围之内，但他们的解释很有力，我们必须走到他们解释力的尽头，才能真正看清那个我们想指认的东西是什么。

我们先看波利亚。“启发式模式匹配的成功概率”这个变量，度量的是一个认知主体在结合以往经验与当前问题结构后，找到一条正确解路径的可能性大小。AI在这个维度上没有理论上的天花板。一个足够大的模型不仅能模仿人类所有的合情推理，还能发现人类从未注意过的跨模块类比。那么，在什么情况下，一个数学家会说“虽然我通过类比找到了这个结构，但它在我心里还没“对”，我必须把它拆掉重来”？这就是分离点。数学史上，埃瓦里斯特·伽罗瓦（Evariste Galois）的工作是一个印记。他那些零散的、在决斗前夜仓促写下的手稿，其“合情推理”的概率按当时的数学框架来看，几乎是零。他没有找到类比，他是在那些早已被接受的操作规则（代数方程的可解性）与一个完全异质的结构（群）之间，强行建立了一种在当时的启发式空间中根本不存在的关联。他“看见”的东西，让他否定了此前所有合情推理导向的死胡同。那份否定性，那份“你们这条路不可能通”的冷硬判断，恰恰不是任何启发式模式匹配的产物，而是主体理性在“此路不通”的张力里，被逼着重新定义了他所处的整个概念空间。也就是说，他的“看见”不是一个正面的推测，而是一次对整个解题空间的本体论重裁。启发式模式匹配，是在给定的空间里找路；而我们要指认的这份认知发生，是当所有已知的路都被“不对”的张力封死时，主体被迫重画地图。AI目前能做到前者的极致，但它的优化函数里没有一个量纲，是用来度量“整个已知解空间的崩坍所引发的主体性改口”的—— 因为它没有一个会被崩坍刺痛的自我界。

再看哈恩的“可推导性”。用逻辑蕴含来重述一切数学真理，在原则上永远是对的，就像说一个死去的人的所有行动曲线都能被一组确定的微分方程重述一样—— 这是“发现的语境”与“辩护的语境”的古老混淆。本文要区分的，是从未知到已知的那一跃的发生学结构。哈恩的框架里没有“不知道”的本体论位置。对于身处某个历史时间的数学共同体来说，在P ≠ NP或任一千禧年难题被解开之前，那个证明在逻辑上是“已存”的；在发生学上，它是“尚未存在”的。我们说AI缺失了“知识证明的发生土壤”，指的不是AI无法调取那条最终被写出的逻辑链，而是说，在它输出那个正确字符串的千万次计算中，它没有一次是在“三界追问”的绷紧下，被迫做出一个不仅是“正确”而且是“不得不如此”的、改写了自身可能性的抉择。这个“抉择”不在最终的证明文本里，它在证成者对于所有被放弃的路径的沉默记忆里。传统的逻辑实证主义无法区分这两种“推导性”：一个是被黑箱搜索到的可推导性，另一个是被三界纠缠的土壤逼出来的、带有路径截断痕迹的“肥发生”的可推导性。而AI正是借用了前者，僭取了后者的名分。

然后是沃斯的“搜索剪枝”。在计算机科学内部，“启发式搜索”对付的是组合爆炸。它的量纲是“被剪掉的搜索节点数”。这是一个极其强大的变量，它证明了许多人类的“直觉”不过是有效的剪枝。那么，在证明过程中，一个数学家说“这条路看起来明明是对的，但它感觉不对”，然后因此

改道，与一个高级剪枝算法因为评估函数赋分过低而放弃了一条分枝，二者的差别在哪？分离点在于：算法对一条分枝的放弃，是因为它的评估函数基于历史数据判定它最终收敛的成功率极低—— 这是一个向前看 （收敛预期）的功利判断。而那位数学家说“它感觉不对”时，她指的是，尽管沿着这条路走，逻辑符号上没有任何一步是错的，但它与她脑子里正在形成的，那个尚未符号化的“新的概念骨架”在深层咬合不上。这个“咬合不上”，是三界纠缠（尤其是理念界关于概念简约性的约束，和自我界关于“是不是我真正要的”的内感判断）给出的一个无法被评估函数计算的张力信号。换句话说，算法是“因为看不到前景”而放弃，数学家是“因为与尚未存在的骨架不兼容”而放弃。前者是对已知空间的优化搜索，后者是对尚未生成的空间边界的一次负反馈。我们由此看到了“搜索剪枝”的代理边界：它无法解释那种为了接生一个新结构，而主动放弃一堆在已知评价体系里“前景光明”的旧路径的认知行为。

拉卡托斯的“概念拉伸”是迄今最深、也最危险的挑战者。他以多面体欧拉公式的证明史为实验室，揭示了一条不可逆的改口轨迹：有人提出一个看似完美的证明，有人找出全局反例或局部反例，反驳压力迫使证明者重新定义“多面体”以排除反例，新定义又催生新的反例与新的修正。这条轨迹的外在形态，与本文所描述的“改口”极为相像。但我们须将二者之间的分离线刻得分明。拉卡托斯的概念拉伸，归根到底是概念本身的内涵与外延在反驳压力下的重新调整——多面体的定义从“一个由多边形面围成的立体”收缩为“一个简单多面体”，再收缩为“一个凸多面体”，每一步都是为了使原证明重新合法化。这个调整是理性的、逻辑的、发生于公共论证空间中的。然而，本文所言的“改口”，不是对概念的单纯调整，而是调整者本人在调整发生于自身之后，再也无法用调整之前的旧方式去提问。这不是概念的膨胀，而是提问者的位移。伽罗瓦在引入群这个概念之前，面临的是一个被代数操作规则封死的空间——他并非在试探着“重新定义可解性”，而是被迫放弃了“可解”这个提问方式本身所预设的、关于方程根与系数之间关系的全部旧的认知框架，转向了一个此前不存在、且在当时几乎所有一流数学家看来“不是数学”的提问方式：不问这个方程的根能否被根式表达，而问这组根之间的排列关系携带了什么不变量。这个位移，不是概念的拉伸，而是提问者的自我改口。而拉卡托斯的模型不要求、也不描述提问者自身的认知结构被修改——在他的图景里，数学家始终站在概念之外，以理性的修辞技巧调整概念的边界以应对反驳。本文要指认的、TCS绷紧之下所发生的那件事，正是站在外面调整边界与被迫发现自己的立足点已经移动之间的那条裂隙。

莱考夫与努涅斯的“体验性认知基模”，换了一条完全不同的路径来瓦解TCS的特殊性。他们不说“改口只是修辞”，他们说“整个数学理性都是人的身体与经验在概念层面的投射”。在这个框架里，格罗滕迪克在概形理论中承受的那种“被未知之物绷紧”的彻骨之颤，无非是一套极其复杂的、源于空间包含与抽象转换的隐喻基模在极高的整合压力下产生的认知张力。这个张力的来源，是人类共通的身体结构与经验图式（如容器图式、源头-路径-目标图式），而不是一个特定问题域的三界追问。我们在这一挑战面前，须同时给出与拉卡托斯方向不同的两条分离线。其一，体验性认知基模解释的是人类如何能够理解数学，而不是一个人为什么在这个特定的历史时刻、在这个特定的问题上，被逼到不得不改写自己提问方式的绝境。前者是类的认知能力（或认知语法），后者是此一认知事件的发生契机。伽罗瓦的身体并不比别人特殊，他的容器图式与路径图式也并无比拉格朗日更发达，他在死前那个夜晚写下的那几页手稿之所以成为数学史的转折点，不是因为他的体验性基模此刻达到了某个阈值，而是因为他在与一个特定问题的长期撕扯中，被逼到了旧有的提问方式（“如何用根式解五次方程”）的绝对尽头。这个尽头，是那个具体问题在这位具体的人身上生成的，不是全人类的隐喻网络在那一刻自动涌现的。其二，更为关键，拉卡托斯的“概念拉伸”与本文的“改口”之间的区

分，同样适用于这里：Lakoff与Núñez的理论解释了概念在头脑中如何被构造与整合，却不解释构造者自身如何被这种构造所改变。他们描述的是概念的发生（conceptual emergence），而非发生者的发生（the emergence of theconceiver as a changed subject）。我们说一个数学家在TCS中被绷紧了，这不是在说他在认知上“很累”，而是在说他将自己压进一个问题后，最终不得不以一种彻底于己不利的方式来作结：交出那个他曾以为是自己的、赖以思考的旧框架，承认它不够用，并亲手推毁它的一部分。这不是概念隐喻能煮熟的石头。它要求一个在承受追问的位置上做出的第一人称抉择，且该抉择一旦做出便不可逆—— 这正是我们以“三界追问”和“自由律”所指认的那块飞地。

三、承重命题：作为被绷紧的土壤与不可逆的改口

至此，我们已分别将“启发式模式匹配的成功概率”、“逻辑系统中的可推导性”、“搜索空间中被有效剪枝的分支数量”、“反驳压力驱动的概念重定义”以及“隐喻映射的基模化程度”这五个代理变量，从数学认知发生的最终解释位上剥离。我们并不是在说它们错了，而是说，当它们试图僭越为那个终极解释项时，我们看到了一个失灵区：一个数学家在某一次不可逆的高认知张力抉择之后，对他（她）自己认知路径的不可还原的“重裁”。这把我们导向一个需要被正式命名的控制变量。

我们将这个变量称为被绷紧的认知土壤（Taut Cognitive Soil，TCS）。它不是指数学知识的逻辑基底（如ZF/ZFC公理系统），而是指在一次具体的、生成新知识的行为中，被激活的三界纠缠的紧绷状态。它有如下三个维度的定义性特征：第一，在理念界，它表现为既有概念网络对该猜想形成最大约束的饱和状态。它不是对规则的“了解”，而是对“所有已知的简化方式在此处都不再有效”的直觉性把握。第二，在现实界，它表现为该猜想能够在实证或形式推演层面，给出一个此前无人能做出的、排他性的具体预测（或一个判据性的构造），这一点是它迫使主体负起“可证伪性”责任的扳机。第三，在自我界，它表现为主体必须做出至少一次其本人无法用旧有概念框架重述的、涉及认知路径取舍的“决断”（哪怕只是一次命名），并且在此决断之后，该主体自身对于此问题的提问方式已被不可逆地改变（改口）。这三个特征同时高强度在场，才能称作“TCS 被激活”-。TCS这个控制变量，不是测“这个人聪不聪明”，也不是测“这个结论对不对”，它测的是一个更冷硬的发生学事实：这次认知事件是否已经不可逆地改写了执行者自身的认知路径。

现在，我们将承重命题正式立下：

一种数学认知的发生X，其成为“知识证明”的本质身份，并不来源于它是Y_1（一种高成功概率的启发式模式匹配的结果），也不来源于它是Y_2《公理系统里一个被高效搜索到的逻辑蕴含），也不来源于它是Y_3（一个被优化的搜索策略在巨大搜索空间中剪枝后的输出），也不来源于它是Y_4（反驳压力驱动下对概念的内涵外延的策略性重定义），更不来源于它是Y_5（体验性认知基模在概念整合中的高阶投射），而是Z—— 一个被TCS绷紧的认知土壤，在自由律的运行下，所实现的一次不可逆的、对自身认知路径的重裁，其可被外部观察的操作性签名，是提出者在回应针对理念界、现实界与自我界的三个追问时，所做出的无法被现有概念吸收的“改口”。这种改口的判别标准，是提出者因此而无法再以改口发生之前的提问方式去理解其本人曾经相信过的那个问题—— 这是拉卡托斯的概念拉伸框架从未要求的层次，也是波利亚、哈恩、沃斯以及莱考夫-努涅斯的解释范围均无法覆盖的残余。

这个命题拒绝将“知识证明”还原为任何形式的“被发现物”。它断言“知识证明”是一种“发生事件”，其发生的标志不是输出字符串的正确性，而是那个输出者的认知路径在该事件中发生了不可逆的结构性变更。在这个意义上，AI在当前的范式下，并不是在产出“有缺陷的数学知识”，它是在产出另一种东西—— 一种可以极度逼近正确、却从未被三界纠缠的土壤绷紧过的“瘦发生”。AI产出的“证明 ”，是一个已经坍缩为字符串的终点，但它从未背负那些被放弃的、前景光明却因不兼容而被主人亲手杀死的路径；也正是这些“被杀死的可能”，才是“肥发生”之为知识的不可见的内核。

四、第二轴强制：绷紧的土壤与推理的广度

为了将Z从又一个漂亮的新名字升华为一条可被验证和应用的辨别维度，我们必须为它加上一根结构独立的第二轴。我们找到的候选轴是：推理链条的广度，即一次数学认知活动最终产出的论证可以被展开的独立逻辑步骤的数量与分支的复杂度。

本文预测，TCS的激活与推理链条的广度，在当代“AI辅助证明”的环境下，将呈现一种深刻的相互制约与反向选择。为了检验这一点，我们在此先提出第二轴，并为后续论证建立一个初始的辨别格。我们将这两个维度的强度分为高和低，形成一个2X 2的初始象限。通过这个辨别格，那些“敌意近邻”的预测将与本命题发生可裁决的冲突。

第一象限（高 TCS × 高推理广度）：本文预测，这是“肥发生”知识证明的理想态。其典型实例是安德鲁·怀尔斯（Andrew Wiles）对费马大定理的证明，或格罗滕迪克（Alexander Grothendieck）的概形理论重构。在这种状态下，主体承受着巨大的、来自三界的逻辑与认知张力，并成功地将这种紧张感外部化为一个宏大而精细的、可被独立重建的公开论证体系。其高推理广度，并非来自冗长的计算，而是源于TCS的绷紧迫使主体必须创建全新的概念语言来承载其“看见”的重量。波利亚框架会预测，这种广度是大量启发式推理的串联与并联，其成功概率是累积的。本文预测则与此不同：其成功的核心不是启发式成功率的累积，而是由TCS驱动的那一次不可逆的本体论重裁，在根本上改变了后续所有推理的发生路径。拉卡托斯框架会将其描述为一系列反驳驱动的概念拉伸的累积效应，但无法解释为何怀尔斯在其七年孤独研究中所经历的那种自我改口——他曾公开承认自己在发现最初证明中的漏洞后，一度认为自己“永远无法修补它”，这不是概念边界的外在调整，是主体对自身认知可能性的内在重估。而莱考夫与努涅斯可以解释怀尔斯大脑中的隐喻网络此时经历了何种整合压力，却无法解释“修补不了”这个判断何以成为怀尔斯自我认知的一部分、并反过来迫使他寻求泰勒的合作以更换思路——不是隐喻网络在更换，是他在更换他赖以思考的隐喻网络本身。

第二象限（低 TCS × 高推理广度）：本文预测，这是AI“瘦发生”的典型领地，也是当代“证明黑箱化”焦虑的发生地。AI产出的、人类无法逐行理解的冗长证明，或一项需要超级计算机完成的、无法被直觉简化的巨大分类定理的检验，皆属于此。它们拥有极高的逻辑步骤复杂度，但其生成过程并未伴随任何一个主体在TCS维度上的深层激活。这里，沃斯的“搜索剪枝”框架能极好地解释其高效率。但在本文看来，这是“知识证明”之名被“算法预测的正确性”所僭越的位置。我们的可裁决预测在此展开：如果一个数学共同体长期主要依赖这类产物（高广度-低TCS）来推动前沿，它将不可避免地发生知识路径的“板结化”——人们只知道某些命题是正确的，但失去了调动那份被绷紧的土壤去重新剪裁和质疑这些路径的发生学能力。

第三象限（高 TCS × 低推理广度）：本文预测，这是打破旧范式的“核心引理”或“观念证明”（proof by idea）的发生地。其典型是前述伽罗瓦的遗稿，或黄皓（Hao Huang）对敏感度猜想的不到两页的证明。推理链条的出奇简短，不是因为问题简单，恰恰是因为主体在TCS激活时，被迫完成了一次深层的概念重裁：他（她）否定了所有繁复的旧路径，找到了那条“本就该如此”的、极其稀疏的结构结晶。在这个象限里，哈恩的“同义反复”框架会认为这不过是一个短逻辑链的被发现一它的简短性就是其逻辑简单性的证明。本文的预测截然相反：这种简短性是一个高压的、高TCS的

发生过程的终点，这个过程以“不可逆改口”为代价，换取了最终符号链的极度浓缩。它“短”，是因为它在形成之初就在TCS的压迫下，把所有不能被新骨架容纳的枝节都排斥掉了——这是一种发生学的简短，不是逻辑蕴含的简短。在此，拉卡托斯的概念拉伸模型仍可以描述伽罗瓦对“可解性”概念的修改过程，但它抓不到一个要害：伽罗瓦在引入群之后，已无法再以拉格朗日的方式去提问——对他而言，一个五次方程求根式解的问题，已经不再是一个有意义的问题了。这不是概念的内涵被调整，而是旧问题在发生学上被注销了。莱考夫与努涅斯可以指出伽罗瓦在“置换”这一概念上动用了某种不同于前人的隐喻基模，但他们无法解释为何这种隐喻的转换，竟同时让转换者本人永久地丧失了对旧隐喻的感知力——这种丧失，恰恰是本文用以区分“体验性基模转换”与“追问者自身的不可逆位移”的判据。

第四象限（低 TCS × 低推理广度）：这是常规数学训练的领地，比如本科生习题、标准化的定理应用。此区间由“技巧”而非“判断”主导，不需要TCS的深度激活，也不产生长链的探索。其认知价值在于巩固已知的概念连接，而非改写它们。

在这个辨别格里，最关键的判别格是第二与第三象限的冲突。本文的核心赌注是：在第三象限（高TCS-低广度）发生的那种认知重裁，无法通过不断优化第二象限（低TCS-高广度）的算力来自然“涌现”。因为驱使伽罗瓦或黄皓做出那种“砍掉”动作的，并不是在巨大组合空间里的高效率搜索，而是被TCS绷紧后，对“什么才是一个恰当的数学对象”这一标准本身的一次不可逆的变更。这是一种本体论框架的重新裁定。

五、可裁决判据：让最近邻做出相反的预测

这种起源的遗忘不是时间的流逝造成的，而是逻辑实证主义以来将“发现的语境”与“辩护的语境”严格割裂所付出的代价。在这道割裂的墙垣两边，数学知识的最终形态（被码成逻辑符号的证明链条）被赋予唯一的认识论身份，而那个将其绷紧并推入存在的三界土壤却被归入心理学—— 一个不产生任何认识论分量的废纸堆。本文整个上半篇的论证，就是为了打穿这道墙垣：那片被宣告为“不相干”的土壤，恰恰是数学知识成为“知识”而不是“正确符号串”的发生学发动机。它的停止，才是黑箱化焦虑的真正来源。而本文能够识别出这片土壤，是因为我们与两个具有强大解释力的近邻—— 拉卡托斯的概念拉伸理论、莱考夫与努涅斯的具身认知进路—— 进行了正面交手，并通过他们各自的失灵区，才找到了那条不可还原的分离线。

拉卡托斯在《证明与反驳》中以“多面体欧拉示性数”的历史重构为战场，揭示出一条精密的辩证机制：当一记反例击穿一个被珍视的猜想时，数学家并非抛弃它，而是通过“概念拉伸”将其定义不可逆地修改，将反例排除、或吸收为合法的新成员，从而使猜想在更复杂的形态下存活。这一模型捕捉到了那个我们关心的核心现象：一次数学认知的推进，往往以对核心概念的“不可逆修改”为标志。拉卡托斯确实看见了，证明不只是验证，它是概念被反例逼迫着自我重定义的过程。这正是他作为敌意近邻的致命吸引力：他的“反驳驱动修正”与本文的“不可逆改口”，在现象表层几乎无法区分。我们必须给出分离线。他的模型有一个未被他自己追问的假定：他所描述的那场辩证，是被一个持续在场的“概念-命题-反例”的逻辑三角所驱动的。每一次概念拉伸，都是对这个三角内部张力的一种逻辑结算——反例迫使概念扩大或缩小以消解矛盾。这个结算动作，始终发生在同一个逻辑层面、同一个可定义的概念空间之内。它不追问：如果整个“三角”的底座—— 那个被默认共享的、关于“什么才算一个数学对象”的默会框架——本身在一次拉伸中被摧毁了，该怎么办？拉卡托斯的故事里，那个由命题、证明与反例构成的游戏，其规则是可以自我修正的，但游戏的棋盘——即“什么是多面体”这类最底层本体论—— 始终是修复的对象，而不是被撤离的对象。而我们要指认的那种“改口”，恰恰发生在棋盘本身被撤离、参与者被迫不再用旧定义去“吸收”反例、而是宣告“这类对象从此不再以旧方式被提问”的时刻。伽罗瓦对“可解性”的改口，就是棋盘的更换—— 他并没有去“拉伸”“根式解”这个概念以吸收五次方程的某种新解法；他宣告“可根式解”本身不再是从根部追问的恰当框架。拉卡托斯的“概念拉伸”只能描述棋子在棋盘上的移动与棋盘的局部修整，而不能描述棋盘被更换这一事件本身—— 因为它缺少一个让棋盘本身成为被追问对象的维度，即本文所称的自我界的不可逆位移。

莱考夫与努涅斯则在概念隐喻与数学认知基础的研究中，将数学的起源锚定在人类的体验性基模之上：我们的身体经验、运动图式和意象基模，通过隐喻映射，构成了数学概念的可理解性基础。他们正确地指出，像“置换”这样看似抽象的代数操作，其可被思考的前提之一，是主体脑中关于“物体在空间中换位”的某种前概念的体验性土壤。这一框架的代理变量，是“支撑概念理解的体验性基模及其隐喻变换”。用它来解释伽罗瓦，会说伽罗瓦成功地将“置换”从一种对固定物体的外部操作，想象为一种构成物体自身身份的内在关系—— 这是一种隐喻基模的根本转换。本文的分离线，同样落在“转换者自身的不可逆位移”上。莱考夫与努涅斯的模型可以完美地解释一个认知主体如何从一种理解模式“迁移”到另一种理解模式，并把两种模式之间的差别分析得极其精细。但它无法解释，为

什么在经历这种迁移之后，伽罗瓦不能简单地“切换回”拉格朗日的理解模式，以便和同时代人进行更顺畅的沟通。对他而言，那个旧的提问方式已不再是“一种可选的理解模式”，它变成了一个被取消的、不再有意义的问题。这不同于“忘记了一种母语”，而是“撤销了那个语言所设定的提问资格”。一种体验性基模的转换，解释不了“资格的撤销”；它解释的是模式的更迭，而非提问者自身的不可逆位移。而这份位移——那个“我已经不再是可以那样问问题的那个人了”的不可逆改变——正是本文用TCS的“自我界”维度所命名的那个东西。它是被三界同时追问的土壤，逼迫着一个主体做出的、对自身认知资格的自我剥夺。这种自我剥夺，在“概念拉伸”里是棋盘上的精妙走位，在“隐喻变换”里是图式的优雅迁移，而在TCS的框架里，它是一次发生学的本体论注销。

正是靠着这两个分离点，我们才得以锚定TCS这个控制变量——它是那份迫使棋盘本身被更换、迫使提问者自身被注销的绷紧状态。它不是聪明，不是技术，不是模式匹配或逻辑搜索的任何一种优化形态。它是一个主体被逼到认知资源的尽头、被迫否定了那套此前让他能够识别“什么是资源”的默会框架本身时，所进入的那个自我裁决的瞬间。而那个瞬间的产物，就是一次不可逆的改口。

现在，我们将TCS这一控制变量，推进到可裁决判据的刀刃上。本文的核心赌注，不是要去和波利亚、哈恩、沃斯在“谁的解释更优美”上纠缠，而是要押在一个具体的判别格上，让他们做出相反的预测。这个格，位于第二象限（低TCS-高广度）与第三象限（高TCS-低广度）的冲突地带。

我们构想的检验模型如下：一个由大规模计算机穷举得出的、人类无法逐行直观理解的庞大证明（譬如对某个有限群分类中一系列繁冗子情形的穷尽式检验，落于第二象限），其正确性在共同体中已被形式化验证接纳。现在，我们向一群顶尖数学家悬赏，要求他们对此穷举证明给出一个“观念的、本质的、能被人类直觉在几页内把握的”重构。他们可以无限接触那个正确黑箱证明的所有逻辑足迹与中间结果。问题是：这个压缩过程，是只需要高超的合情推理技巧与启发式搜索策略，还是必然至少触发一次不可逆的改口？

波利亚与沃斯的综合预测（波利亚-沃斯联合预测）是：这是一项可被策略性逼近的任务。庞大的穷举证明提供了海量的类比基地与归纳素材。一个足够熟练的数学家，运用逆推、一般化、寻找结构相似性等合情推理技术，辅以对证明空间剪枝模式的精细分析（即沃斯的“找出那条最低能量路径”），应当能逐步“猜出”那个更短的核心出发点，而不必然需要对问题框架发起一次非连续性的重裁。在那个更短的观念证明被写出之后，人们甚至可以回头画出它在原始穷举空间中的“影子路径”，来展示它不过是被高效搜索到的更优解。拉卡托斯的框架会进一步补充，这个压缩过程很可能正是一次概念拉伸的生动案例：数学家发现了穷举证明中某个隐蔽的不变量或统一结构（反例消失），并将它吸收进一个被拉伸后的核心定义中，从而使得繁冗的分支情况全部坍缩。整个过程是对既有概念做出不可逆修改，但这个修改本身仍发生在数学共同体的推论辩证空间之内，不需要摧毁那个空间的棋盘。

本文则做出方向相反的赌博。我们预测，存在一个子类，其原穷举证明所依赖的对象定义，会在“观念证明”的重构中被宣布为不恰当的、有根本性缺陷的，而不仅仅是被绕开或简化。也就是说，那个最终写出观念证明的数学家，她必须公开做出一次改口，其形式可能是：“我们把问题搞错了，这些对象不应该按原来的分类去处理，问题的边界不在于那些分支的具体情形，而在于那个分类标准本身就是冗余的；或是：“那性质 X并不是对象的一个可以被穷举的属性，它根本是属于另一个范畴的结构在错误相位的投影”。这次改口，要求她与那个庞大的、已由计算机证实为正确的穷举证明所安身立命的“本体论地基”一刀两断。如果她不做这一刀，仅仅对穷举证明进行引理压缩和逻辑重组，那

么她的产物将只是一个经过高度优化、但仍然在旧定义空间内部运行的“瘦观念证明”—— 它能被很好地理解，但它没有移除那层冗余。而本文赌的是，要移除那层冗余，把那片巨大的复杂度彻底蒸发掉，只有一个办法：证明原初的问题陈述—— 即那台庞大计算机被指令去搜索的那个东西的定义本身—— 是那个复杂性的来源。一旦她用一次改口重新裁定了对象，那巨大的搜索树就不是被“剪枝”了；它是被连根拔起，扔到了另一个本体论范畴的垃圾桶里，新的证明在那片重新开辟的、干净得多的地基上建立起来。这个动作，是拉卡托斯的概念拉伸无法覆盖的，因为它不是对棋盘的修补或扩张，而是将整个棋局宣告为一场误入歧途的游戏，并另开一局。莱考夫与努涅斯的框架能解释她如何找到“新一局”的体验性基模，但不能解释她为何必须亲手“注销”自己曾精通的那局旧游戏所设定的提问资格。这份注销，是她为更换棋盘付出的不可逆的自我位移。

要检验这个“改口”是否发生了，不能依赖哲学家对数学家内部认知状态的揣测，必须落在一个可观测的代理变量上。我们提出的可观测代理，是这次重构的前后，关于同一被证明命题的公开讨论中，关键定义句的语义主语是否发生了不可还原的替换。具体而言，我们从该数学家或共同体的公开文本中，提取重构前与重构后对问题域核心对象的定义陈述。如果我们发现，重构后的定义句，其主语的语义范畴已不属于重构前所采用的分类体系——例如，将问题的主语从一个“具有性质A的集合族”替换为“一个拓扑不变量在某种算子下的像”，并且这个新主语不能被旧有分类体系的任何合法复合所生成——那么这就是一次改口的文本痕迹。如果整个重构过程，仅有逻辑压缩和引理合并，所有定义句的主语始终可以追溯到旧框架内的元素，那么TCS变量在此案例上被证伪，波利亚-沃斯联合预测将胜出。

这个可裁决判据，可以立刻在一个现实中的、具有类似结构的案例上被“低压”检验。黄皓（Hao Huang）对敏感度猜想的证明，就是一个已被共同体接纳为“观念证明”的文本，并且它的“短”本身被视为其深刻性的症状。在黄皓之前，对此猜想的攻击持续了三十年，产生了大量关于布尔函数不同复杂度度量的、繁复细致的局部结果。这些结果构成了一个巨大的“已知正确结论的推理网络”，其广度极高。黄皓的证明不到两页，其核心构造是一个基于超立方体子图的谱论证，极其稀疏。这是一个极其接近“高TCS-低广度”的案例。我们现在可以验证：在黄皓证明中，那次“改口”是否发生了？我们审读其论文。黄皓的证明并没有去对已有的、关于敏感度与其它复杂度度量之间局部不等式进行“拉伸”或“重组”。他做了一件让所有那些繁复的旧路径集体失声的事：他引入了一个全新的矩阵构造，并将原问题的陈述直接翻译为一个关于该矩阵最大特征值的性质。原问题中的核心对象定义——“布尔函数的敏感度”一没有被拉卡托斯式地拉伸，而是被一个全新的数学结构（他所构造的那个对称矩阵的谱）在一个更高的层面被重新“映照”了出来。旧框架里所有的局部不等式、所有的渐进估计，在这个映照下没有一个是错的，但它们都变成了冗余的中间层。黄皓证明之后，一个数学家对“敏感度猜想被解决了”的理解，不再需要遍历那些三十年间的局部工作；他只需要一句：黄皓证明了那个矩阵的最大特征值就是根号n。原有的那个庞大推理网络，其功能（证明敏感度猜想的某个局部）没有被否定，但其作为一个“问题域”的存在权重被根本性地降维了。这次重构，完全符合我们所说的“更换棋盘”：问题的定义（敏感度与块敏感度的关系）被重新指向了一个之前不在此问题域内的算子（一个巧妙构造的对称矩阵的谱），而这个新主语“矩阵的特征值”无法由旧有分类“布尔函数的组合复杂度度量生成。这是一次干净利落的改口，一次在公开文献中可以清晰指认的自我夺权。它满足了本文提出的可裁决判据，赋予了TCS在这一案例上的正面读数。

而正是在这件事上，AI展现了其结构性无能。要复现黄皓的壮举，一个AI系统需要做的不只是在已知的布尔函数复杂度度量的搜索空间里进行更高效的匹配。它必须自主地决定：停止在现有的度量框架内寻找新的不等式，而去完全不同的一个数学领域（谱图论）中“借来”一个全新的结构，并将其构造性地嫁接到原问题上。这个“嫁接”动作的触发，不是因为在旧空间中搜索的期望收益降到某个阈值以下——　　波利亚-沃斯式搜索的优化函数可以轻易模拟这个“放弃旧空间”的决定，只要给“失望”赋一个足够高的惩罚权重。但那个决定，只告诉系统“别再搜这儿了”，它不告诉系统“该去哪儿借那个不存在于你当前本体论词汇表里的新结构”。而黄皓的改口，恰恰是“去哪儿借”这一发生学的创造性动作。那“一借”，源于他被那个三十年未解的、被理念界（所有已知联系均已穷尽）和现实界（所有局部结果均已试遍）逼到极点的张力，并在自我界被迫否定了他借以理解该问题的原有范畴本身——　他不再问“哪个度量能压住敏感度”，而是问“如果用一个全新的谱，它能暴露出原来框架里的什么隐藏关系”。这个“不再问”，就是改口。目前已知的任何ATP系统，都不具备在搜索过程中，自主产生一个关于其搜索对象之本质的“为什么-不再问”的否定，并据此切换到另一个不兼容的本体论词典中的能力。因为它的词典，是它搜索的宇宙的边界；而改口，就是扩展这个边界、甚至摧毁旧宇宙的动作本身。这个动作，不是计算，是发生。

本文提出的另一个、现在就能被检验的赌注，与制度变迁有关，并直接关联到“自证焦虑”与“存在感单一锚定”的结构。如果我们将TCS在制度层面的对应物——　即一套迫使研究者持续回到未解问题之核心压力的共同体筛选机制——视为一种需要特定认知生态才能维系的稀缺资源，那么我们可以做一个反向检验：我们可以去寻找那些已经主动用“高广度-低TCS ”的成果产出模式，替代了“观念证明”作为主要学术资本积累手段的数学子领域。在这些领域里，由于发表高广度（长、复杂、依赖大规模计算与协作）的结果更容易、更安全、且更不易被挑剔“缺乏深刻性”——因为深刻性是一个需要被TCS绷紧后才能判断的、难以量化的维度——　　我们预测，该子领域的研究者将系统性地降低对“提出新概念框架”和“裁出核心引理”的激励。这里，可观测的代理变量是：在某个顶级期刊的某一子领域分类下，过去十五年间发表的所有论文中，“引进了新的核心概念或对象定义”的论文，与“对已知框架进行了大规模计算验证或穷举推进”的论文，其比率的变化轨迹。本文的预测是：在任何一个由“瘦发生”主导的子领域，前者的比率将呈显著下降趋势，而同期该领域发表的论文总数量可能不降反升。这个趋势的一个早期信号，是该领域中获奖（尤其是颁给“观念性突破”的奖项）成果的平均年龄的显著增长——　因为“瘦发生”的加速产出，掩盖了“肥发生”的概念资源正在被消耗而未能得到及时补充的事实。板结化，是土壤退化的最终状态。

本文的最后一条证伪路径，是一份写给人工智能的、带有明确时间期限的赌注。我们将站到波利亚、哈恩与沃斯的阵营里，赌他们输。到二零三五年截止，如果一个完全自主运行、未在关键节点被人类输入新本体论的AI系统，能在一个公认的、期待“观念证明”的开放问题上（例如，为朗兰兹纲领中某个长期以来仅有计算证据支撑的函子性猜想，提供一个揭示“为什么”的简短证明，而不仅仅是数值验证），产出被数学共同体一致评为具有黄皓证明那种“观念绝杀”特质的成果；如果在整个过程中，人类合作者事后无法在公开访谈中指出任何一处“AI让我们把问题的定义本身给换了”的改口节点，那么本论文的核心变量——　被绷紧的认知土壤（TCS）作为数学认知发生学独立变量——将失去其必要性。因为这个结果将证明，那种发生学的“改口”，本质上可以被一个不承受三界追问、只操作形式符号的系统，在足够庞大的搜索与优化下复制出来。此刻，波利亚的精神将合并沃斯的技术，宣告本文的失败。

六、证伪条件

条件一（基于历史案例的反向检验）。找出数学史上一个公认的、彻底改变了领域认知框架的“观念证明”。追溯该证明的提出者的公开文本（论文、信件、访谈），如果在整个发现过程的记录中，我们没有找到任何一次不可逆的“改口”陈述——即提出者从未在关键的定义、分类标准或问题构成上，表达过“我之前那样想是错的”“这个问题以前被问错了”或“这个对象不该被那样定义”——而整个发现过程可以被完整描述为一连串“我又看到了一个更漂亮的构造”的优化组合，那么本文关于TCS与前文分离点的论述，将在此案例上被证伪。这个检验是低成本且可滚动的：它可以由任何一位数学史学者，选任一重大突破进行正面核验。

这四条路径（案例的正面读数、制度的反向检验、写死时间的AI赌注、历史文本的反向检验）彼此独立，各属不同的检验方式：文本证据、行为结果、制度效应。它们共同构成了本框架的可反驳性边界。

七、结语

回到那个迫使本文动笔的原初问题：AI能做数学证明，那么数学家还有存在的必要吗?

这篇论文的回答是：这个问题本身问错了。它暗含了一个未被审查的前提——数学家的存在，仅仅是一种“证明生产”的功能。一旦这个功能被机器替代，这个职业就失去了其存续的理由。我们试图用整篇论文的发生学结构来论证，这个前提是对数学认知的窄化。数学作为一种人类活动，其不可被算法化外包的内核，不在于它有能力产出“正确的命题”，而在于它有能力在被逻辑必然性绷到极致的土壤上，栽出一条此前不存在的、并且会迫使其栽种者自身首先被它改写的理解路径。这种改写，不是对已有逻辑空间的更优搜索，而是对那个空间的本体论构成——什么该算作一个恰当的对象、一个合法的操作、一个能被问的问题——的不连续的重新裁断。

波利亚的合情推理描述了在既有框架之内的、极其高明的策略性跳跃。哈恩的同义反复拆掉了关于数学真理最终基础的形而上迷雾。沃斯及其后继者的自动推理，将这片土地上的微观耕作推进到了人类肉身无法企及的精细度。它们都是对的，但它们共同漏掉了那个裁断框架的、发生在它们起跑之前的动作。

那个动作，需要一个主体。这个主体不一定非得是一个人，但它必须是一个能承受被三界张力同时追问、并据之以做出抉择的位置。这个位置，就是此刻还坐在那张旧书桌前、被一个未解的问题攫住、强迫自己重新想象“它本该是什么”的那个数学家。AI可以在数秒内证明她眼前那个猜想的十万个推论，但它尚未被赋予那个问出“哦，也许我该把问题换成它”的、自我否决的瞬间。

那一瞬的否决，就是土壤被绷紧的证据。而那份被绷紧过的土壤，才是数学认知在历史中留下沟壑的真正犁铧。守住这个位置，不是为数学家守住饭碗，而是为人类的集体认知，守住一种在黑箱化洪流中随时可能被遗忘的精度——一种不是“计算得正确”，而是“追问得恰当”的精度。

八、参考文献

1. Bourbaki, N. (1950). The Architecture of Mathematics. American Mathematical Monthly, 57(4), 221–232.
2. Hahn, H. (1980). Empiricism, Logic, and Mathematics: Philosophical Papers. D. Reidel Publishing Company.
3. Huang, H. (2019). Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture. Annals of Mathematics, 190(3), 949–955.
4. Lakatos, I. (1976). Proofs and Refutations: The Logic of Mathematical Discovery. Cambridge University Press.
5. Lakoff, G., & Núñez, R. E. (2000). Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being. Basic Books.
6. Pólya, G. (1954). Mathematics and Plausible Reasoning. Princeton University Press.
7. Scholze, P. (2022). Liquid Tensor Experiment: A challenge for formalized mathematics. Experimental Mathematics, 31(S1), 1–10.
8. Wos, L. (1996). The Automation of Reasoning: An Experimenter's Notebook with OTTER Tutorial. Academic Press.
9. 站内资料：《SDE本体论入门：世界是如何发生的》，SDE Universes 专著摘要导读。
10. 站内资料：《数学：发现还是发生？》，SDE Universes，2026年7月。
11. 站内资料：《SDE内部的三角关系：从子三角到理想边界情形》，SDE Universes，深度长文。
12. 站内资料：《二级延申·承重柱解构·十篇》，SDE Universes。
13. 站内资料：《第五章 伞模型：SDE计算数学的统一框架》，出自《SDE数学导论》专著选读。
14. 站内资料：《存在与流变的共同根源——对一次区分动作的发生学追溯》，胡敏，SDE Universes，2026年8月。