

裁决竞逐：当指标不再度量价值，而是争夺定义价值的权力

一场持续了五十年的败仗，为什么还在打

作者 张琼 · SDE 学员 · 约 3.3 万字 · 发表于2026年7月29日 · 深化增补版

摘要

为什么关于"指标扭曲价值"的批评说了半个多世纪，指标体系不仅没有退场，反而渗透进了越来越多的领域？既有解释——从坎贝尔定律到指标政治学——都共享一个未经审查的预设：被指标伤害的那个"真正的价值"，是一个先在的、稳定的、可以被批评者援引的参照系。本文撤销这个预设，指认出一场长期被价值话语遮蔽的结构性竞逐：内在论证标准与对标论证标准在争夺同一个东西——不是"谁更准确地度量了价值"，而是"谁掌握了定义'何为有价值'的权力并借此支配资源分配"。价值不是先于争论而存在的裁判，而是每次裁决过后、由获胜的论证语法作为"本应如此"而回溯性地建构的。本文将这场竞逐命名为"裁决竞逐"——两种无法相互翻译的裁决语法，共享同一组稀缺资源接口，在任何一次具体的资源分配事件中，选择其中一方的裁决依据就等于剥夺另一方的制度效力。裁决竞逐不可还原为"指标的扭曲效应"（那仍然把指标当尺子），也不可还原为"权力的政治运作"（那仍然把权力当作外在于论证的东西），更不可还原为布迪厄的"符号暴力"或福柯的"治理术"（前者把权力定位在论证之外，后者止步于指标成为治理工具之后，二者都未抵达那一步：权力在论证语法本身的接口处结算）。它是发生在论证语法本身的层面上的事件，回应的是当代制度生活中一个最深的不适：为什么我们明明知道怎么论证真正重要的东西，却总是在决策时发现，它找不到接口。

关键词 裁决竞逐；指标治理；内在论证；价值回溯；论证语法

一、引言：一场持续了五十年的败仗，为什么还在打？

1976年，坎贝尔（Campbell, 1976）写下那条著名的定律：一个社会指标一旦被用于决策，它就不再是它所声称要测量的社会事实的有效度量。此后近半个世纪，从古德哈特定律[1]到指标政治学，从大学排名的批判到绩效考核的反思，关于"指标会扭曲它所测量的价值"的论证从没断过。这些论证在理论上极其锋利，在经验上反复被验证。但有一个事实同样不可回避：它们几乎从未在制度层面赢过。

指标体系的版图不仅没有收缩，反而从大学扩张到了中小学、从学术评价扩张到了艺术创作、从公共管理扩张到了私人健康。今天，一个人可以在手机上用十几个数字追踪自己的睡眠、心率、卡路里、步数、冥想分钟数。每一次新的数字化技术落地，第一个被提出的问题几乎总是："它的指标是什么？"

这不是因为反对指标的论证不够有力。恰恰相反，很多论证在智力上是无懈可击的。但它们在制度中反复失效。每一次失效的模式惊人地相似：批评者指出指标遮盖了真正的价值→制度改革者承认

指标有局限，承诺探索更全面的评价方式→新的评价方案出台，增加了更多指标维度，加入了一些质性描述的空间→几年后人们发现，增加的质性描述仍然没有在资源分配中获得实际权重，而那套被批评的指标体系继续稳坐在裁决的中心。

这不是某一次制度设计的偶然失败，这是一场持续了五十年的结构性僵局。

本文要追问的，不是“为什么指标不好”——那个问题已经被回答过无数次了。本文要追问的是：为什么足够好的批判，反复打不赢？这指向一个更深层的问题：这些批判是否在根本上搞错了他们正在攻击的对象？他们以为自己是在批评一把尺子测不准温度，但那个被批评的对象——指标体系——可能根本不是一把尺子。它可能是一个正在争夺宣判权的对手。你用批评尺子的语言去攻击一个对手，你的攻击当然落空。

本文将论证：过去半个世纪的指标批判共享了一个预设——那个“被指标扭曲的真正的价值”（真正的学术水平、真正的教育质量、真正的健康），是一个先在的、稳定的、可以被批评者识别和援引的参照系。这个预设如此自然，以至于很少有人注意到：它本身正是一个需要被质疑的先验。

如果撤销这个先验，会发生什么？本文的核心判断是：内在论证标准与对标论证标准在争夺同一个东西——不是“谁更准确地度量了价值”，而是“谁有权力在制度中说出一件事是不是有价值的”。价值不是那个在争论发生之前就已经端坐在裁判席上的法官。价值是被每一次裁决过后，由获胜的论证语法所“追溯性地建构”出来的——不是先存在一个“真正的学术卓越”，然后我们拿它来判定影响因子失效了；而是在影响因子支配了学术评价几十年之后，“学术卓越”在制度中的可操作所指，已经是（而且仅仅由）影响因子及其衍生指标体系所定义的东西。那个批评者援引的“真正的学术卓越”，在制度层面已经失去了其可被感知、可被操作、可被下一代学人认的具体形态——它的匿迹不是被任何人驳倒的，而是在资源分配、职业晋升、话语通道的每一次裁决中，被排除在接口之外，持续缺乏制度性的资源反馈，最终使没有任何行动者能继续将使用该语法作为可存活策略。

这就是本文所要命名的“裁决竞逐”。它不是关于指标的精度之争，关于权力的分配之争，也不是关于方法论优劣之争。它是发生在论证语法本身的层面上的事件：两套无法相互翻译的裁决标准——内在论证与对标论证——共享同一组资源接口，在任何一次具体的资源分配事件中，使用了其中一方的裁决依据，就等于宣告另一方的无效。这是一场在每一次人事晋升、每一笔经费分配、每一次战略规划KPI设置中，悄悄发生、却从未被精确叫出名字的竞逐。

在正式展开这场竞逐的解剖之前，必须先方法论上做一笔澄清。这篇论文追踪的，是一场制度性的压覆——一套论证语法正在把另一套挤出资源分配的通道。这个叙事本身，天然地具有一种偏向：被挤出的那一方（内在论证）在叙述中容易被赋予一种受害者的道德重量，而挤人的那一方（对标论证）则容易被写成一个冷血的管理机器。这种叙事偏向不是论证的错误，却是论证必须警惕的陷阱。

因此，从这一句开始，本文确立一条纪律：裁决竞逐的机制本身，对竞逐双方是语法中性的。它不是“坏人用计分表欺负好人”，也不是“真价值被假指标谋杀”。它是一套由审计速度、公平性焦虑与接口结构共同构成的选择压力，在其中，内在论证因为其生成判断方式的自体论特征（第四、五节将详细解剖），在这些压力的筛选下存活率更低。对标论证在这场竞逐中获胜，不是因为它更“恶”，也不是因为它与权力更“亲密”——而是因为它更符合这套选择压力对“可裁决性”的要求。如果有一天，制度

的选择压力发生了根本性的位移，内在论证同样可能成为压覆方——它会在那些它占住了接口的领域里（如某些顶尖人文院系的聘任决策），让对标论证的数字无法进场。裁决竞逐这个概念没有预设谁是正义的一方。它只追问一个冷的问题：在这个具体制度现场，哪套语法控制着裁决的通道？另一套是否仍然找得到接口？[2]

带着这个纪律，本文将依次展开以下论证：先与既有理论对话——不仅是“指出它们漏掉了什么”，而是让最接近本文立场的几个理论邻居（坎贝尔及其后继者、审计社会研究、布迪厄的符号权力、福柯的治理术、拉图尔的信用循环）全部上桌，然后一个接一个地指出它们各自停在哪一步，裁决竞逐又往前多走了哪一步（第二节）；接着追踪当代制度生活中裁决竞逐的发生过程——不是将审计速度、接口不可翻译性、裁决逃逸作为静态的“条件清单”罗列出来，而是追踪它们在一次又一次具体的裁决事件中，如何将内在论证的语法从一个有存活能力的裁决方式，推入了一个持续无法获得制度反馈、最终无人再能将其作为生存策略的境地；并在此过程中正面论证那个被整篇论文当作地基的命题：内在论证标准为什么在本体论上无法被“制度化”（第三节）；然后通过一个来自中国大学职称评审改革的真实案例，把接口占有、裁决逃逸、回写与价值回溯这四重机制从抽象的“合理推演”拉进可被审视的发生学现场（第四节）；在此基础上澄清一个关键区分：裁决竞逐不是范式之争（第五节）；再回应四个最重要的反驳——关于批判者的后设立场、关于“制度化”的奢望、关于内在论证是否偶有“反杀”成功的经验反例、以及关于行动者能否主动利用回写机制逆转竞逐（第六节）；最后给出让裁决竞逐可复现的最小分析步骤、证伪条件，以及在理论收束处指明这个发现对批判理论、制度研究与知识论本身的意味（第七节）。

二、近邻们停在了哪一步：裁决竞逐的不可还原性论证

在展开正面论述之前，必须先让最危险的对手们上桌。本文的核心概念“裁决竞逐”如果不能与已经占据了话语空间的最相近理论做出清晰的、概念层面的切割，那么它作为一个“新辨别维度”的宣称就只是作者的自我声明，不是学界可以检验的判断。

这不是通常意义上的“文献回顾”——那是另一套知识生产逻辑的标准操作：先铺排一堆前人的名字，然后把它们当作需要被自己超越的台阶。裁决竞逐要求的对话方式与此不同：它必须逐一指出，这些最接近的理论各自停在了哪一步。它们已经看见了什么？它们在哪个节点上停下了，以至于漏掉了那个最关键的动作——权力在论证语法本身的接口处结算？

本节将与五组理论邻居依次交锋：坎贝尔定律及其测量失效论的后继者、以迈克尔·鲍尔为代表的审计社会研究、布迪厄的符号权力与场域斗争、福柯的治理术与真理体制、以及拉图尔等人的科学信用循环。这五组对手不是随意拼凑的——它们分别从测量、审计、权力、真理体制与资源循环五个方向逼近了裁决竞逐的领地，但都在即将跨过门槛的地方停下了。

（一）测量失效论：尺子不准，但尺子的物种没被质问

坎贝尔定律和古德哈特定律构成了这条线索的核心。它们的逻辑简洁而有力：一个指标本来是要度量某个社会事实X（教育质量、科研水平、公共健康），但它一旦被用于决策，人们就开始针对指标Y本身优化行为。当行为围绕Y展开而不是围绕X展开时，Y和X之间的相关性就会断裂。美国人搞标准化考试，老师就开始教学生考试技巧而不是学科内容——考试分数上去了，学生的实际学力却没有同

步提升。英国人搞急诊等待时间指标，医院就把病人放在救护车里排队而不是计入等待时间——指标达标了，医疗质量没改善。这个发现是真实的、重要的。

但测量失效论停在了这里。它的修复方案永远是"找更好的指标"或"给指标加修正系数"。它承认当前的尺子不准，但不质疑尺子这个物种本身是否胜任它所承担的制度角色。它处理的是尺子和对象之间映射关系的松动（认识论问题），而不处理"对象本身是否因为长期被尺子替代而在制度中丧失了可感知性"（本体论问题）。换句话说，测量失效论看到的是相关性断裂，没看到更深的替代。它以为"真正的学术水平"只是暂时测不准，还在那里等着更好尺子去量它；它没发现，那个"真正的学术水平"在三代学者被规训在影响因子的话语里长大之后，已经在制度层面失去了可被独立指认的形态——不是因为被驳倒了，而是因为它所依赖的论证语法在资源分配的接口上被驱逐了。

这就是测量失效论停下的那一步：它把指标仍然当作尺子。裁决竞逐往前走的这一步是：指标可能根本不是度量——它是在替代子。它不再是对某个独立存在的价值的"测量"，而是对那个价值在制度中的可操作所指的"定义"。当影响因子成为职称评审的核心赋权项时，它不是在对"学术卓越"进行可能有误差的测量——它是在告诉整个学术共同体：在这个制度里，"学术卓越"就是这个。尺子一旦被制度性地用于裁决资源分配，它就不再是尺子。它是一个宣判。

（二）审计社会研究：解释了繁殖，没解释存活

以鲍尔（Power, 1997）为代表的"审计社会"研究，是继测量失效论之后最有力的一波诊断。它指出，当代制度中指标治理的蔓延，不是技术理性的自然演化，而是审计文化爆炸的结果——当组织必须在越来越短的时间窗口里向外部利益相关方"证明"自身的价值时，可量化、可审计、可对标的指标就成了最"理性"的选择。鲍尔尤其精细地展示了审计实践如何不只是"检查现有的东西"，而是反过来"制造出可被审计的东西"——一个大学如果必须每五年接受一次外部评估，它就会主动把自己的教学和科研活动重构成一组可以被放进评估表格里的数字。

这个诊断在制度层面是精准的。它解释了为什么指标在今天无处不在——不是因为指标本身更好，而是因为指标满足了审计的需求。但它停在了这里。审计压力解释了对标论证为什么会大规模繁殖，但没有解释为什么内在论证在保卫自己的合法性时如此脆弱。它把内在论证的退却描述成"被挤出的"——审计压力太大了，所以深度质辩没空间了。这个描述暗暗地给内在论证保留了一张悲情的底牌：它没有输在论证本身的质量上，它只是被恶劣的环境逼退了。

裁决竞逐不接受这个保留。它追问的是：为什么在审计压力面前，内在论证标准在资源分配的战场上反复败北？不是因为它在智力上不如对标论证——在论证深度和捕捉实在的能力上，它可能远胜对标论证。而是因为它在接口上有一个结构性的弱势：它产出的判断，无法被翻译成年度KPI、季度进度条、可量化的达标率。而在审计驱动的制度环境里，无法对接资源分配接口的论证，在制度中活不过几轮决策周期。这不是被挤出，这是竞逐失败——它输在自己最擅长的事不能转换成制度承认的语言。

审计社会研究解释了"为什么指标这么多"，但它没有追问"为什么指标这么能存活"。裁决竞逐追问的是后者。这是审计社会研究停下的那一步。

（三）布迪厄的符号暴力：权力在论证外面，不在接口处

现在进入最关键的一场对话。布迪厄（Bourdieu, 1979）的场域理论——尤其是符号暴力与误识的概念[3]——是裁决竞逐最危险的近邻，因为它已经辨识出了一个极深的机制：在社会场域中，支配的合法性不是靠外部强制来维持的，而是靠被支配者“误识”了支配的标准、把一套专断的等级秩序当作自然而然的区分来接受。符号暴力的整套分析，精准地刻画了支配如何通过论证语法（而非绕过它）来运作。

那么，“裁决竞逐”不就是布迪厄意义上的符号暴力在管理领域的上演吗？不是。区别在以下几个层面。

首先，布迪厄的符号暴力始终预设了一个存在性的前提：被符号暴力所遮蔽的那个“被支配者的真实利益”或“场域斗争中被压抑的可能性”，是一个客观的、尽管被误识但独立存在于那里的东西。批评者的任务是“揭示误识”——让行动者看到，那套被他们内化了的等级标准不是自然的，而是专断的，从而打开重新争夺合法性的空间。在这个图景里，价值（即使被遮蔽了）仍然是一个可以先于揭示而存在、等待被重新看见的实体。

裁决竞逐不预设这个实体。它指认的，不是一套先在的价值被符号权力遮蔽了，而是一套论证语法因为在制度接口上长期无法获得资源反馈，其对应的那套价值的可操作所指——它在制度中的可感知性、可使用性、可教性——已经在几代人的规训中匿迹了。这不是“遮蔽”，是持续缺乏存活条件。当第四代学者从博士生起就只经历过影响因子作为裁决依据的制度时，他们不是“误识”了那套标准为自然——那套标准对他们来说就是自然本身。那个批评者所援引的“真正的学术判断力”，在他们的经验里没有对应物。揭示误识的前提——被遮蔽者在某个时刻能够“认出”揭示所指向的那个真实——不再成立。这不是因为被遮蔽者“不够觉醒”，而是因为那个“真实”所依赖的整套认知生态（训练模式、质辩经验、裁决通道）已经整体匿迹了。

其次，布迪厄的符号暴力始终把权力定位在论证的外面：权力通过塑造场域的资本分布和行动者的惯习，来影响“什么被认为是好的论证”。论证是权力运作的载体——权力在论证的后面、上面或旁边操控它。但在裁决竞逐中，权力不在论证的后面，权力就在论证语法本身的接口处结算。当一个评审委员会决定用影响因子总分作为裁决依据时，这不是“权力在外面逼迫委员会做出这个决定”，而是委员会采纳对标论证作为论证语法——在那一刻，本身就成为了权力的结算点。不是“权力选择了指标作为工具”，而是“把指标锚定为裁决依据”这个动作本身，就是裁决竞逐的结算瞬间。布迪厄没能走到这一步，因为在它的分析里，论证始终只是权力用来说话的嘴，而不是权力发生的现场。

（四）福柯的治理术：被排斥者匿迹之后，福柯不问它变成了什么

福柯后期（Foucault, 2004）的治理术概念，以及与之密切相关的“真理体制”分析（Foucault, 1975）[4]，或许比布迪厄更逼近裁决竞逐的内核。在福柯看来，现代权力不再表现为一个至高无上的主权者对臣民的禁止与惩罚，而是分散在一整套知识技术与管理制度中，通过统计、分类、监控、标准化——简言之，指标——来“引导”人口和个体的行为，使之朝着“社会整体利益”的方向自我规训。影响因子不只是一个评价工具，它同时也是一个规训工具：年轻学者在写作论文时，已经把“这篇论文能投到哪个影响因子区间的期刊上”内化为研究设计的一部分了。他们不需要任何人来强迫他们追求影响因子——他们是自我规训的主体。

这个分析非常强力。但裁决竞逐对治理术的追问同样是：你停在了哪一步？

问题的症结不在福柯"不够深",而在于他的分析装置在处理"指标如何成为治理工具"这一问题时,漏掉了一个步骤。在福柯的图景里,指标——统计、分类、标准化——是权力技术的一部分。权力通过它们来运转。但指标之所以能成为权力技术,不是因为它天生就是。它是在一套特定的制度选择压力下,因为在裁决接口上比其他论证语法更能存活,才被反复选中、反复加固,最终长成了看起来"天然就是权力工具"的样子。福柯的治理术分析,从指标已经是治理工具这一点开始。它不问:指标是怎么在与其他裁决方式的存活竞争中胜出的?为什么是影响因子变成了规训的载体,而不是深度同行评议?不是因为影响因子在规训上更隐蔽、更高效——而是因为它在"被做成可执行、可复检、可申诉的裁决条目"这一点上,具有内在论证不可企及的优势。福柯可以描述这个结果——指标规训了主体——但他不追问这个结果的发生条件:指标是在竞逐中赢了才变成工具的。把"成为工具"当作分析的起点,而不是分析的对象,就是福柯的治理术停下的那一步。

更关键的一笔在后面。福柯的真理体制分析指出,权力与真理是互相生产的——不是权力的意志扭曲了真理,而是权力与真理在同一个体制中互为条件。这已经在某种意义上触及了"价值是裁决后的回溯建构"这个判断:权力制造了什么可以被称作"真"的东西。然而,福柯的"真理体制"始终是在认识论的层次上运作的:它解剖的是知识如何被权力条件化,哪些陈述能被接受为"真",哪些被排斥为"谬"。它追问的是:在某个特定的历史时期,知识与权力的配置使得某些陈述能够作为"真的"进入话语,而使另一些陈述被斥为"非理性"或"谬误"。这不是本体论层次的追问。

裁决竞逐追问的则是:当内在论证所支撑的那套判断——关于"什么是真正的学术推进"的判断——被排斥出裁决接口三代人之后,它不是一个被判定为"谬误"的陈述。它是被排斥了,但没有被宣告为假。它只是被排除在裁决的通道之外。问题在于:这个被排斥的"真",在三代人无法经由制度接口遭遇它之后,在制度生活中变成了什么?真理体制可以解释它为什么被排斥——因为它不符合那个时期权力-知识配置所要求的可审计性、可度量性、可比较性。但它不追问被排斥之后的结果:它不是被驳倒的,但它匿迹了。它是怎么匿迹的?匿迹这个动作——一种语法在制度层面失去全部实践者、无法被任何新入者体认——不在福柯的分析工具箱里。

这不是说福柯"错了"。福柯的理论有自己精确的目标:解剖权力如何通过生产"真理"来运转。但恰恰是这个目标——解剖权力的生产性——使得他的分析装置在本体论上就无法容纳那个被他自身的分析所排斥的东西的后续状态。福柯过于关注权力所产生的东西(主体、知识、话语),以便揭示它们不是自然的而是被建构的;但对于权力不再产生的东西——那些在被排斥之后、没有进入任何话语、没有被宣告为假但也没有任何人再能体认的东西——福柯没有追问。不是因为他不够敏锐,而是因为他的整个分析装置是为揭示"生产性"而设计的。"匿迹"——不是被驳倒、不是被压抑、不是被否定,而是持续无法获得制度反馈直至无人再能使用——不在这个装置的追问范围内。

这正是裁决竞逐与福柯的不可还原切割点。裁决竞逐处理的,是两套论证语法在同一个资源接口上的存活竞争——不是谁被承认为"真",而是谁在裁决之后仍然能作为可存活的话语策略被下一代行动者所使用。福柯的真假轴线,关涉的是陈述的认识论地位(它们被接受为真还是排斥为谬);裁决竞逐的存活轴线,关涉的是论证语法在制度中的本体论地位(它是否仍有实践者在制度中使用它来做出裁决,还是已经没有任何人能够这样做)。这两条轴线在概念上是不可互相通约的。福柯可以解释,为什么影响因子作为一种"真理"被接受了而深度同行评议被排斥了——因为前者符合审计时代的真理体制对"可计算性"的要求。但他不能解释,那个被排斥的"深度同行评议",在被排斥三代人之后,

作为一个可以在制度中活着的判断实践，变成了什么。裁决竞逐要处理的，就是那个被排斥之后、在福柯的分析里没有被追问的后续。

这就是为什么"裁决竞逐"不可还原为"真理体制的特定表现"。不是术语层面的区分，是追问的对象不同。福柯追问：权力如何通过制造真理来运转？裁决竞逐追问：一套论证语法如何在制度中失去全部存活条件？前者是生产性的解剖，后者是存活竞争的追踪。两者不矛盾，但不可互相替代。

（五）拉图尔的信用循环：翻译中介被短路之前发生了什么？

拉图尔与伍尔加（Latour & Woolgar, 1979）在《实验室生活》中提出的"信用循环"理论[5]，是对科学界资源-承认循环的最早的、最精准的描述之一。在实验室里，学者的职业生涯沿着一条清晰的循环展开：发表论文→获得同行承认（引用、奖项、邀请报告）→将承认兑换为科研经费→用经费购买更好的设备、雇佣更大的团队→产出更多的论文，从而进一步巩固承认。在这个循环里，每一步都同时是"科学判断"（这研究好不好）和"资源分配"（钱给谁、职位给谁）的交汇点。这与裁决竞逐的分析对象高度邻近。

然而，拉图尔的信用循环与裁决竞逐之间，有一个关键词的区别：翻译中介。在经典型的信用循环里，循环的每一步转换——从论文到承认、从承认到经费——都需要一个"翻译中介"：同行。某人的论文发表了，同行要读它、讨论它、在自己的文章中引用它、在评审会上为它说话。这些动作不是自动发生的，它们依赖于一个在学科共同体内被持续生产的、活生生的判断过程。即使这个过程中充满了策略性、修辞和权力博弈（拉图尔对此有极锐利的分析），它仍然是一个需要人来执行的判断。

裁决竞逐揭示的，正是这个翻译中介在当代指标体系中发生了什么：它被自动售货机短路了。在信用循环里，从"发表论文"到"获得承认"的转换是需要同行来判断的——他们判断的是这篇论文的质量。而在当代的指标治理中，这个判断环节被影响因子跳过了：期刊的影响因子直接替代了同行对论文质量的判断。论文不需要被同行阅读、质辩、承认，它只需要被发表在影响因子高的期刊上——影响因子就是那个已经做好的"承认"。承认变成了一个数字，不再是一个过程。评审委员会不再需要做那场信用循环中拉图尔所描述的、充满争议的"判断"——他们只需要查数字。数字进了接口，裁决自动生成。

拉图尔的信用循环可以描述转换步骤的存在，但它没有描述这个转换步骤被短路的特殊状态。而短路一旦发生，整个循环的性质就变了：它从一个同时生产"科学知识"与"学者声誉"的生态过程，变成了一个只生产"可被指标捕获的发表记录"的自动运转机器。裁决竞逐要解剖的，就是这场短路：它不是发生在"权力压迫了科学"的宏观层面，而是发生在"判断被数字替代"的那一个接口上。

（六）小结：一个新的辨别维度

现在可以正面回答那个一定会来的质疑了："这不就是这些成熟理论的综融版吗？"不是。综融是拿几把已有的刀，从不同方向割同一块肉。裁决竞逐要做的，是换掉那块被割的肉本身。

坎贝尔及其后继者把指标当作尺子。审计社会研究解释了指标为什么这么多。布迪厄解剖了权力如何通过论证语法来遮蔽。福柯揭示了指标如何规训主体并制造真理。拉图尔描述了承认与资源之间的循环转换。它们各自停在离同一个门槛一步之遥的地方：尺子其实不是尺子，指标在替代它声称要度量的对象；繁殖的对面是压覆，审计压力的对面是内在论证持续无法获得制度反馈；权力不在论证

外面操控它，权力就在论证语法本身的接口处结算；治理术的指标工具，不是在真空中变成工具的——它是在竞逐中存活下来才变成工具的；信用循环的翻译中介，被自动售货机短路了。

裁决竞逐是一个不同的辨别维度：它不追问"指标怎么扭曲了价值"，而追问"两套论证语法在争夺什么"。它不把价值当作先在的裁判，而是当作竞逐获胜后的纪念碑。它解剖的不是一个工具好不好用，而是一场在制度接口上悄悄打了五十年、却没有被精确叫出名字的战斗。

三、竞逐是怎么发生的：从条件到过程

如果竞逐的胜负仅仅取决于论证的质量，本文就没有必要被写出来。内在论证的锋芒在任何单独一次学术讨论中都可以压倒对标论证——这已经被无数批评文献证明了。但问题是：这些胜利从未被结算为制度性的权力转移。赢得论证，输掉决议——这就是内在论证标准在当代制度中的典型处境。

本节要追踪的，就是让这种处境得以被反复生产出来的那个过程。但进路不是罗列一组静态的"条件"（审计太快、公平性太焦虑），然后宣布"因为这些条件，所以内在论证输了"。那种写法是既成对象论的——它把竞逐的结果当作被条件"造成"的，而不是在一次次裁决事件中被"发生"出来的。发生学要做的，是追踪那些条件如何在一次又一次具体的裁决中，将内在论证的语法从一个有存活能力的裁决方式，推入了一个持续无法获得制度反馈、最终无人再能将其作为生存策略的境地。

为此，本节将从审计压力如何收窄裁决的时间窗口开始，然后解剖内在论证标准在接口上的本体论弱势——它生成的判断为什么无法被提前存储、无法被第三方复检，为什么这个弱势不是一个技术难题而是两种标准形态之间的本体论鸿沟；最后展示公平性焦虑如何在这个裂缝中触发裁决逃逸，以及逃逸之后的回写如何把每一次竞逐的结算缝进"从来如此"的集体叙事。

（一）审计压力：它创造了什么存活条件？

自新公共管理运动（Hood, 1991）在上世纪八十年代席卷全球公共部门以来，一套新的制度逻辑被铺设进了大学、医院、研究机构：组织必须用可被外部利益相关方审阅的方式，证明自身的价值。注意这个措辞——不是"证明自己有价值"，而是"用可被审阅的方式证明"。这个区分是要害。

内在论证完全有能力证明一所大学的价值。一个由同行深度质辩构成的过程，可以在十年的尺度上，累积出关于"这所大学到底好在哪"的一整套丰富而精确的论证。但这套论证有一个致命的特征：它不能被任何外部审计者在两周内读完并做出判断。它要求审阅者本身也是那个论证共同体的一员，具备同样的学科训练、同样深度的质辩经验、同样长时间的浸入。而审计的核心驱力恰恰相反——它需要的是快速、需要的是可以在不同机构之间横向比较、需要的是不依赖于审阅者本人的内行程度就能被采纳的证据。

正是在这个裂缝里，对标论证获得了它在制度中的压倒性优势。影响因子可以被一天之内查完。H指数可以在几个研究者的简历之间横向比对。ESI学科排名可以在一张表上呈现。它们需要的审阅时间不是十年，不是两个月，不是两周，而是十分钟。在审计的生态里，十分钟与十年之间的差距，不是效率的差距，是存活条件的差距。一个必须在两个月内完成外部评估的委员会，不是因为愚蠢才选择指标，而是因为委员会的制度职能就决定了它不可能被允许使用那套需要十年才能被充分理解的论

证方式。不是内在论证不够好，而是审计生态已经替所有需要做出裁决的场合，预设了一个内在论证无法够到的速度门槛。

这一笔的关键在于：审计压力不是简单地"逼迫"了人们使用指标。审计压力从根本上创造了两种论证语法的不同存活条件。但这个存活条件不是一次性"造成"了竞逐的结果——它是在每一轮评估、每一次晋升、每一笔经费分配中，反复地、增量地发挥作用的。每一个具体的裁决事件，都是审计压力的一次微型执行：委员会面对一叠申请材料，必须在规定时间内做出决定。在这个具体场景中，对标论证的数字已经静静躺在计分表上，不需要任何人去"生产"它；而内在论证的判断，则要求委员会花几个小时去读代表作、去质辩、去在讨论中实时地生成那个专属于这一次裁决的判断。在这种一次一次的具体选择中，对标论证的"可存储性"被反复结算为裁决的效率优势，而内在论证的"实时发生、即时消耗"被反复结算为裁决的时间成本。五十年下来，累积的不是某一次"审计压力逼出了指标"的事件，而是成千上万次微型裁决中，对标论证因为对接了审计的时间接口而反复存活、内在论证因为无法对接而反复在操作层面被绕过的过程。

（二）接口的三个层面与"可存储性"的本体论根源

在继续推进之前，必须对本文反复使用的一个概念——"接口"——做出明确的界定。这个词不是在比喻"通道"，它是指制度中将论证翻译为裁决的赋权装置。它至少包括三个层面，三层之间不是平行的，而是从硬件到软件到认知的层层内化。

赋权接口：在制度文本（职称评审条例、学科评估指标体系、战略规划KPI设置）中，那些被明确赋予了计算权重的条款。这是物理层。计分表上的"发表篇数×期刊系数×项目级别"是赋权接口；"代表作同行评议：供参考"不是赋权接口——它虽然在文本中存在，但不在计算中占有任何权重。赋权接口的占有，直接决定了哪一种论证语法能将判断翻译为资源分配的结果。

叙事接口：在事后故事里，允许出现的归因逻辑。这是叙事层。当一个裁决做出之后，集体叙事会自然地给这个裁决补上一个"为什么它是对的"的故事。叙事接口决定了，哪些归因可以被公开说出而不会听起来像是失态（"他发在顶刊上了"可以；"他的影响因子比其他人高"在有些场合已经不太体面但还可以被接受；"因为计分表这么算的所以晋升他"完全不可以公开说）。叙事接口是回写机制的通道——裁决逃逸之后，被逃逸掉的那一部分通过叙事接口被悄悄从故事里抹掉了。

认知接口：深植在行动者身体里的、被规训出来的自动判断倾向。这是认知层。当一个学者在读另一个学者的简历时，"Q1区论文数量"自动地在她的感知里占据中心位置，而"代表作摘要"需要她有意识地、花费心力地去读、去判断、去给出一个不依赖任何数字的权重——前者是自动的，后者是费力的。自动的那个，就是被规训到身体里的认知接口。认知接口是赋权接口和叙事接口在长期运作中结出的最深的果子，它的可怕在于：它让裁决者自己都不觉得自己在用一套特定的论证语法。她只是觉得"很明显他的记录更好"——那个"明显"本身，就是认知接口在说话。

三层接口的关系是层层内向的：赋权接口占领裁决通道（物理层），叙事接口事后改写集体记忆（叙事层），认知接口把前两者深植为身体性的自动反应（认知层）。三层落成之后，即使赋权接口某一天被从制度文本中拿掉（比如在条例里增加"综合评价"的权重），认知接口和叙事接口仍然会在人的身体里和集体的故事里继续运转——这就是为什么单纯的"制度改革"几乎总是被回退的原因。

现在可以问那个必须被解剖的问题了。为什么内在论证产出的判断，在本体论上无法被提前做成可存储的、可供第三方复检的裁决条目？不只是因为“太复杂”——那是偷懒的说法。真正的原因必须从内在论证标准的生成方式中挖出来。而这个生成方式，与对标论证标准之间，存在一个深刻的、结构性的不对称。

对标论证的标准是前置的、给定的。影响因子作为一个数字，在评审会召开之前就已经静静地躺在数据库里了。评审委员会不需要去“判断”某篇论文的影响因子是多少——那个数字是事先存在、可以被任何人查阅、可以被落选者自己计算的。它不依赖于任何一场具体讨论的发生。它被做成一个裁决条目的方式，就是把它从数据库里提取出来，填入计分表。它在本体论上就是为“被存储、被复检、被提前写入条例”而生的。它的标准形态——那个数字——在裁决发生之前就已经完全确定了，裁决只是把它从数据库里搬运到计分表上。

内在论证的标准是后置的、生成的。“什么是独创性”“这次评审中哪一项贡献算得上是真正的推进”，这些判断只能在、而且必须在一个真实的同行质辩过程中被实时地发生出来。面对一篇被反复讨论的代表作，一位评审委员说“这个证明绕过了这个领域过去二十年卡住的关隘”，另一位说“这个绕过的动作其实是在重新定义被绕的究竟是什么”——他们一来一回的质辩，不是在套用一条预先存储好的叫“独创性”的通用条款，而是在这场具体的、活生生的讨论中，现场生产出“什么构成这次判决中的推进”这一标准的特定形态。这个标准一旦被生产出来，就立刻被用来裁决面前这个人是否晋升。讨论结束，它也消耗了。它不能被存入数据库，不能被放到下一次对另一个人的晋升中使用——因为下一次的那篇代表作、那场讨论，会要求一个与它不完全相同的、同样是实时发生且即时消耗的标准。

让我们把这个道理咬得更深。设想一个评审委员会试图把内在论证的标准“提前制成可申诉条目”。他们写：“申请人的代表作必须在学科核心问题上做出实质性推进。评价时需考察：该推进是否重新定义了既有问题的边界；是否提供了被本领域后续研究所不可绕过的论证或数据；是否展示了对本学科根本困难的独立处理能力。”这三条看起来非常合理、非常像制度文件的语言。但它们能被当作一个公平的计分表来用吗？不能。因为当一位落选者质疑“你们凭什么说我的推进不是实质性的”时，委员会没有办法把这三条翻译成他也能自己查阅、自己计算的分数。而“影响因子总分”可以——他可以自己算，他可以指着条例说“我的分更高”。这不是委员会的错，也不是条目写得不够精确。这是本体论上的鸿沟：内在论证的标准，在被提前写成一个条目时，那个条目的字面已经脱离了它只能在真实的质辩过程中才能获得的具体所指。那个条目成了一个空壳——它看起来是标准，但它不能标准地运作。它永远会滑回“委员会需要自己判断”。

有人可能会问：法律判例不也是“后置的”吗？一个判例在判决做出之前并不存在，它是在具体案件的审理中被“发生”出来的。但判例一旦被做出，就被写入了判例法体系，成为可以被后续案件援引的先例——它被存储了、被制度化了。为什么学术同行评议不能做同样的事？把一场深度质辩的录音、转录和裁决理由存档，作为下一次评审的参考？

这个类比恰好澄清了内在论证标准的本体论特征。法律判例之所以能被存储和援引，是因为判例的“判决理由”（ratio decidendi）可以被抽象成一个可脱离原初语境的规则表述——“在某某条件下，某某行为构成某某责任”。这个规则表述一旦被提炼出来，就可以被放到另一个案件中去比对：本案的事实是否落入那个规则的适用范围？在比对中，那个规则本身不需要被重新生产。它被调用了。这正

是法律推理中"可存储性"的本体论基础：判例产出的那个标准，可以以一个去语境化的规则形态被存下来。

但学术深度质辩中的那个标准——"这个证明重新定义了被绕的究竟是什么"——不能被去语境化。它不是在说"凡是以某某方式绕过关隘的证明都算独创"。它是在说这个证明，面对这个领域的这一关隘，在这场讨论中被识别为做出了一个特定的推进。那个判断是彻彻底底地"索引性"的(indexical)——它在每一个词上都指涉着这个具体的论文、这个具体的知识处境、这场具体的质辩中刚刚被说出的那个区分。你不能把它提炼成一个可脱离这场讨论的规则，因为一旦提炼，它就丧失了那个使它成为"这一次判断"的全部具体所指，变成了一个空洞的通用语（"对学科根本困难的独立处理能力"）。而那个通用语，正如我们刚才看到的，不能标准地运作——它在每一次使用时，都要求一个新的判断来重新给它装填所指。

这就是为什么内在论证标准无法被"判例法化"。不是因为它不够严谨，而是因为它的标准是强索引性的——它在每一次发生中，都在用那个发生本身作为自己唯一的参照系。法律判例的标准是弱索引性的——它在第一次发生时索引于那个具体的案件，但它的制度生命恰恰建立在它可以被部分地去索引化、被抽象为可援引的规则。内在论证标准不提供这个"可以被部分地去索引化"的接口。它要么带着索引性活在一场具体的质辩里，要么丧失索引性变成一具标语的空壳。没有第三种形态。而制度的赋权接口要求的，恰恰是那种可以被去索引化、被提前存储、被第三方复检的裁决条目。这个要求，内在论证标准在本体论上就无法满足。

这就是为什么内在论证标准在接口战场上一直输。不是它在智力上不如对标论证，也不是它没有被人们"制度化"的善良愿望。而是它生成判断的方式——后置、生成、强索引性——在根本特征上与"可供第三方复检的预存储裁决条目"这个要求，处于本体论上的冲突。你要么保留它的生成方式（实时发生、即时消耗、强索引性），放弃把它做成可申诉条目；你要么把它做成可申诉条目，放弃它生成方式的内核——然后把那个内核退化成一套更精细的、加了质性描述的指标体系。没有第三条路。

（三）公平性焦虑与"裁决逃逸"

现在可以解释，为什么每一次制度试图提升裁决的公正性——增加透明度、设置申诉机制、规范评审流程——都会无意中进一步放大对标论证的制度优势，同时持续压缩内在论证的存活空间。

在几十人以上的学术机构里参与过重大资源分配的人，都熟悉一种焦虑：当稀缺的晋升名额、科研经费、研究生招生指标需要在众多申请者之间做出区分时，裁决必须是公正的——否则可能被申诉、被上级推翻、被同事怀疑。但同时，裁决又必须建立在"谁是更好的学者"这类极其复杂、高度依赖专业内行判断的问题上。这两个要求在本体论上是紧张的。真正的深度质辩不可能在形式上看起来"公正"，因为它产出的判断没有办法被翻译成一个可以让落选者也"公平地接受"的通用公式。

这就是"接口不可翻译性"在公平性焦虑下的致命后果。内在论证产出的判断——"经过我们委员会长时间的深入讨论和对你所有代表作的反复质辩，我们集体判断她的工作在这个学科的核心问题上推进得更深"——这个回答可能是更真实的，但它在程序公正的机器里没有任何分量。它不是"错的"，它是"不可操作的"。而对标论证的回答——"他的论文发表期刊影响因子总分高于你，这是计分表，你

自己看"——在程序公正的框架里是充分完整的，可以在申诉机制里站住脚，可以在上级复核时被一条条打钩。

这就触发了一个关键机制——裁决逃逸。裁决逃逸指的是：在一次制度性决策中，实际支配资源分配的那个裁决依据，可以在不被任何人察觉为"一次不公"的情况下，从内在论证的场地中悄悄滑出，进入对标论证的场地。一场关于"谁应该晋升"的委员会讨论，可能花两个小时认真阅读申请者的代表作、激烈质辩那些论文到底有没有突破。但讨论结束后，决定结果的，是那张事先已经被做好的、包含了发表篇数、期刊层次、经费总额和教学课时的量化计分表。委员会成员自己可能都不承认那张表是裁决者——"当然还是同行判断最重要，那个表只是参考"——但如果把过去十年的晋升结果拿出来和那张表做一个回归分析，相关性会高到令人不安。

这不是因为委员会在说谎。而是因为在裁决发生的那个具体瞬间——当必须从四个都"值得晋升"的人里挑出两个时——那个可以被计分表明确量化的差距接管了裁决。内在论证产出的判断在这种时刻是完全真实的，但它不是裁决。裁决已经逃逸到那组合计分上去了。事后，委员会成员在写评审意见时，会把那个量化差距写成文字版："A在国际高水平期刊上的发表记录更为突出"——这句话在语法层面是内在论证的句式，但它的内容是从计分表上摘下来的。因为两套论证语法共享了同一个词汇表（"重要""突出""贡献大"），但给词汇装填所指的方式完全不同，裁决逃逸可以假装没有发生。

裁决逃逸的可怕不在于它发生了一次。而在于它在每一轮晋升、每一次经费分配、每一次战略规划中被重复执行了几十次、几百次之后，它就不再是一个"可以被察觉的滑动"了。它成了制度运作的默认模式。新入行的学者从第一天起，看到的就是一个"自然"的裁决秩序：计分表上的数字高的人晋升了。他们不需要知道曾经有过一种裁决方式，在那里，判断是在讨论中被实时生产出来的。那套方式对他们来说不是"被逃逸掉的"——它从来就没有在他们的制度经验里出现过。

四、一个案例：当代表作制度遇上了计分表

前三节完成了裁决竞逐的理论解剖：与近邻切割了辨别维度（第二节），追踪了审计压力如何创造两种论证语法的不同存活条件，从内在论证标准"后置、生成、强索引性"的本体论特征出发，正面论证了它为什么无法被接口化，并展示了公平性焦虑如何在这个裂缝中触发裁决逃逸（第三节）。但至此，文章仍然缺少一具肉身——一个能让这些抽象机制被摊在桌面上、从头到尾审视一遍的真实事件。

本节提供一个扎根的案例分析。2017年前后，中国多所研究型大学启动了职称评审的"代表作制度"改革，明确将传统的量化评价标准（论文数量、影响因子、项目级别）退后为参考项，新增了以同行评议为核心的代表作评审通道。这项改革在文本上兑现了内在论证话语多年来的诉求：评审不再单纯看数字，而是要看"真正重要的学术贡献"。其启动时的姿态，几乎是裁决竞逐的反面——它试图让内在论证重新入主裁决。追踪这场改革的后续命运，就是追踪一次裁决竞逐在真实制度现场如何重新决出胜负。

在进入案例之前，需要交代其方法论性质。以下案例是基于中国多所研究型大学职称评审改革的公开政策文本、官方新闻稿以及学术界的相关讨论综合而成的"典型例示"（ideal-typical illustration）。它不是对某一所特定大学改革历程的严格实证重建，而是在多个真实改革案例中反复

出现的机制性模式——裁决逃逸、回写、价值回溯——的集中呈现。在发生学的意义上，一个合成的典型例示比一个充满噪音和偏离的孤立个案更能揭示机制的纯粹形态：它不是要证明“某一次改革确实这样发生了”，而是要展示“当内在论证试图在不改变整体审计地基的前提下重返裁决接口时，有哪些机制会系统性地掏空这一重返”。这种揭示机制的方法论价值，在于它为研究者提供了一套可以在其他经验现场中识别同类机制的模板，而非一份关于某个具体个案的调查报告。

（一）改革文本中的接口设计

选取一所有代表性的大学（以下称“R大学”）[6]在2018年发布的新职称评审条例。条例的文本结构如下：

主要依据（赋权项）：

- 三篇代表作的外审专家评议（评分表：按A/B/C/D等级打分，并附书面意见）
- 校内外专家委员会对三篇代表作的现场答辩质辩
- 教学工作量及学生评教分数

参考依据（不计分项）：

- 论文发表总量
- 期刊影响因子及分区
- 科研项目级别与金额
- H指数

在赋权接口的层面，这是一次明确的位移。量化指标从“主要依据”被移至“参考依据”，不在裁决计算中占任何权重。代表作的外审评议与校内质辩成为裁决的唯一赋权通道。如果这套文本被严格实施，它将构成一个内在论证被赋予裁决接口、且未被回退的制度决断。

（二）裁决逃逸的现场：两重接口的博弈

追踪R大学2018-2021年间三批次的副教授晋升过程，可以看到裁决逃逸是如何在两个接口的裂缝里发生的。

第一重逃逸：外审打分表里的数字。 条例要求外审专家对代表作给出A/B/C/D等级打分。这项设计的本意是把同行评议变得“规范”——但规范化本身制造了一个可以被量化的东西。当四位外审专家给三位申请人的代表作评分出炉时，申请人A的三篇论文全A，申请人B的两A一B，申请人C的一A两B。校内委员会在进入现场答辩之前，已经看到了这些评分。

在答辩讨论中，一段典型的对话如下展开。委员甲（B的学科同行，仔细读过B那篇被评为B的代表作）说：“这篇论文——就是发在国内那本普通学报上的那篇——其实有一个非常值得重视的动作。它把过去二十年国际上关于这个问题的主流框架放到中国语境里检验了一遍，然后指出了一个主流框架自己没法处理的结构性盲区。这个工作不是那种能在顶刊上发出来的东西，因为顶刊要的是在这个

框架内部做增量的推进，而它是在质疑这个框架本身。但如果我们讨论的是'代表作'——代表这个人最好的学术思考——那这篇比A那几篇发在高影响因子期刊上的标准推进要深。"

委员乙（非B的学科同行，但在相邻领域）回应："我同意老甲对这篇论文质量的判断。但问题是，四位外审专家里有两位给了它B。如果我们这次晋升了B而不是A，我们需要在评审意见书里解释为什么一个外审评分低的申请人胜过了外审评分为全A的申请人。你能不能在评审意见书里把刚才那些话写进去，让它在一份可能会被落选者拿去申诉的文件里，读起来像是一个公平的、不以主观偏好为依据的裁决理由？"

讨论沉默了十几秒。委员会主任最后说："大家都看到了B那篇论文的深度，也看到了外审评分上确实有一些客观差距。这个差距我们不能不尊重——毕竟外审专家也是同行，他们的判断本身也是同行评议的一部分。我们这次先按外审评分和整体记录的客观差距来定，B下一轮还有机会。"

裁决逃逸发生了。一场本应由质辩内容来决定的裁决，滑向了那组已经变成了数字的外审等级。而外审等级本身——尽管在制度设计中被认为是"同行评议"的一部分——在被汇总为可以横向比较的分数时，已经悄悄地从"判断"变成了"度量"。委员乙的追问——"你能不能在评审意见书里把刚才那些话写进去"——精准地暴露了裁决逃逸背后的本体论冲突：内在论证产出的那些判断（"质疑框架本身""不是标准推进"）在申诉体制中不具有可操作的辩护力，因为它们不能被翻译成一个可被第三方复检的公式。而外审评分的A/B/C等级可以。

更致命的是，在最终的投票中，几位被B的质辩打动的委员在无记名投票中仍然将票投给了A。事后的非正式访谈中，一位委员给出了这样的解释："我承认B那篇论文在讨论的时候让我觉得她的学术判断力更强。但投A不完全是'违心'的——我也真的认为A的整体发表记录更优秀。你说那个讨论让我觉得B的论文更有深度，但那个'觉得'——它在我写投票理由的时候，没办法变成一条能让自己踏实的、公正的依据。但A的记录是实的、硬的、放在那里谁都能看见的。最后投A，我至少知道自己在程序上是站得住的。"

这个解释不是委员在为自己的"违心"辩护。它精确地揭示了认知接口的运作方式：当裁决者在两个判断之间做选择时，一个判断（A的记录更好）是自动的、不需要额外心力去辩护的；另一个判断（B的深度更值得晋升）需要耗费巨大的认知心力去把它翻译成"公正的裁决理由"，而且即使做了这个努力，它仍然可能在申诉程序里站不住。认知接口在这里不是强迫委员选择了A——它是让选择A变得比选择B"更容易"。"更容易"不是"更正确"，但在一个有限时间、有申诉压力、需要集体做出决定的制度场景中，"更容易"本身就是决定性的。

第二重逃逸：叙事系统中的回写。在晋升结果公示后，R大学的官方新闻稿对三位成功晋升者做了简介。以申请人A为例，简介中写道："A博士，副教授，近年来在XX领域国际顶级期刊发表多篇高水平论文，其代表作对XX问题做出了具有国际前沿水平的推进。"这段话的叙事逻辑是完全自治的——但它抹掉了一个东西：A的晋升，在操作层面，是因为外审评分的全A记录与影响因子被"参考"这个事实之间的一个模糊地带。新闻稿无法写"A在本轮制度改革中，因为外审获得全A而晋升"——这样的写法会暴露赋权接口已经从"量化分"滑向了"量化评分"。但新闻稿可以写"发表顶刊且成果前沿"。于是，事后故事把裁决成功的归因，重新缝回了对标论证的叙事脚本：顶刊、前沿、国际水平。那场在委员会内部对B的代表作发生的深度讨论、那些关于"这篇是在质疑框架本身"的判断，在事后故事里消失了——因为它们没有叙事接口。你不能在官方新闻稿里写"B副教授虽然外审评分略逊，但对一个被

忽视的理论路径做出了深度推进"——这句话要求读者去做一套只有深度同行才做得出的判断，而新闻稿的受众不是同行，是校内其他学科的管理者和上级部门。

（三）回写的累积与价值回溯的后果

这两个逃逸不是一次性的。它们在三次晋升批次中重复发生，每一次都在悄悄改写R大学关于"学术卓越"的集体叙事。到了2021年，R大学新入职的年轻教师在被问及"代表作制度的实际运作"时，给出了一种典型的回答："代表作制度当然好，但你自己得先有东西——论文还是得发在顶刊上，项目还是得拿到。代表作是锦上添花，不是雪中送炭。"这句话精准地概括了四年改革的实际结算：内在论证的接口（代表作质辩）在话语上被承认为"花"，但对标论证的接口（顶刊、项目）仍然是那匹"锦"——不是花的才是真正的载体。花可以加一点颜色，但锦要是没有，花没地方绣。

这不是改革设计的失败。这是裁决逃逸的累积效应。当每一轮晋升所实际依赖的裁决依据——那组可以横向比较的外审等级，连带着作为"参考"但从未被真正遗忘的影响因子分区——在事后故事中被缝回了"顶刊+前沿"的叙事时，改革所试图建立的新的价值所指（"真正重要的学术推进"），没有来得及在制度中长出可被下一轮申请者识别和操作的形态。它被旧叙事回购了。

这就是价值回溯在案例中的具体落点。2018年的改革试图制造一个新的裁决逻辑：深度同行评议→识别真正推进→以此分配资源→由此重新定义什么叫"好学者"。但在四年的操作中，实际发生的因果次序是：外审评分+影响因子阴影→裁决逃逸→A晋升/B落选→事后故事把结果解释为"顶刊作者理应晋升"→那个"改革前"的价值所指（顶刊=好学者）被重新坐实。改革试图制造的新的价值所指——一个不依赖于影响因子的"好学者"的可操作定义——没有活过追溯。

到了2023年，R大学在修订新一轮条例时，将代表作制度重新调整为一个"申请门槛"（申请人需要在代表作评审中通过一个最低门槛），而主要赋权项回归到了包含期刊影响因子和项目级别的综合量化表中。回退完成。

五、但不是范式之争：为什么不可通约不等于不可裁决

以上论证可能让读者产生一个自然的联想：这听起来很像库恩（Kuhn，1962）说的"范式不可通约"——两套标准各自定义了什么是好的研究，谁也不能说服谁。如果裁决竞逐只是范式之争在公共管理领域的上演，那本文就只是一个旧理论在新场景中的应用。

现在必须做出一个澄清：裁决竞逐不是范式之争。两者的区别，不是修辞层面的，而是结构层面的。

第一个区别：裁决竞逐的双方共享同一个资源池。

在库恩的范式理论里，新旧范式之争发生在科学共同体的学术讨论中，论战参与者争夺的是"被学科共同体承认为正确的理论"。但在裁决竞逐中，竞逐双方争夺的不是真理，而是同一笔数量有限的、真实的物质资源——一个教授职位、一笔科研启动经费、一个博士生名额。这笔资源给了以影响因子为裁决依据的A，就不能同时给那个以深度质辩为裁决依据的B。这不是"两种看法之间的辩论"，这是"一个决断必须做出，而两种裁决依据给出的答案不同"。

这就带来了裁决竞逐特有的残酷性：它不能被悬置。范式之争可以搁置不判，两个范式各自做各自的研究，逐渐积累证据，等到某个关键实验出现再裁决。但在制度层面，今年必须决定谁晋升，下个月必须决定哪几个项目拿到经费。搁置不是一个选项。而每一次不搁置，就等于让接口占优势的那套标准再赢一次。

第二个区别：裁决竞逐的胜负在于接口，而不在于解释力。

在范式之争中，胜负的标准是解释力的比较，这至少在理论上可以归结为研究共同体在某个时刻达成的集体判断。但在裁决竞逐中，胜负在接口处就已经被决定了，和论证的质量几乎没有关系。一个评审委员会可以在一小时的深度讨论中集体判定B的研究更有原创性，然后在下一次的职称评审规则修订会上，发现那个判定的唯一表达接口——写推荐信——还是那个不占任何计算权重的参考材料。

这就是为什么内在论证标准输了几十年：它不是在解释力的战场上败了，而是在接口建设上从未入场。它总是试图用一个更好的论证去打败一个更差的指标，却不知道指标之所以能赢，是因为它在制度的竞技场上根本不像一个"论证"——它像一架自动售货机。你投进数据，吐出来裁决。而内在论证总是试图用一首诗去对抗一架售货机，然后困惑为什么售货机没有感情。

第三个、也是最关键的区别：范式之争的胜负只是观念的改变，裁决竞逐的胜负会回溯性地让另一种"对"的可能性在制度中匿迹。

在库恩的图景里，旧范式输给新范式之后，旧范式作为一种历史存在被保留在教科书里——它不被相信了，但它曾经存在过。而在裁决竞逐里，输掉的一方不仅输掉了未来，还输掉了过去。当对标论证在人事评价中替代了内在论证三代人之后，第四代学者进入这个领域时，他们连那个被替代掉的东西曾经存在过都不知道了。对他们来说，"学术评价"就是指那套指标体系——从来如此。

这不是"记忆被抹除"的阴谋论。这是接口长期独占导致的一种认知生态的彻底转换：当你的整个职业生涯从博士生到教授，每一步的裁决都只经由同一套接口完成，你就没有任何认知窗口，去遭遇"还有其他裁决方式"这个事实之外的活生生的经验。你知道有人写过文章批评影响因子——你读过那些文章，你觉得挺有道理——但你不知道什么叫"在一场真正以深度质辩为裁决依据的讨论中，为自己的学术价值辩护"。那个经验对你来说是空的，就像盲人听人描述红色。

到了这一刻，这就不再是可通约性的问题。是那个输掉的语法，在制度层面失去了全部可被新入者体认的形态。不是被驳倒——是持续无法获得制度反馈，直到没有任何人再能将它作为可存活的策略。

在发生学里，匿迹不承载挽歌的功能。匿迹只是一套发生过流程的完结。当第四代学者不再会使用深度质辩的语法时，这并不表明某个"更真的学术判断力"被遗忘了——它只是表明：那套语法能够发生的全套制度接口——连同它的训练模式（博士生如何被教成能做出深度判断的人）、裁决装置（同行评议在资源分配中实际占有赋权接口）和叙事功能（事后故事承认这种判断为"裁决的依据"而非"锦上添花"）——已经整体匿迹。发生学只记录河流改道。

六、回应反驳：批判者的立足之地、制度化的奢望、反杀的孤岛、以及行动者能否逆转竞逐

（一）既然价值是回溯建构的，这篇论文自己凭什么反对压覆？

这是本文必须面对的最难缠的反驳。如果本文自己论证了"价值是在裁决竞逐之后被回溯建构的"，那么这篇论文所援引的那个所谓"被指标体系伤害的价值"——真正的学术判断力、深层的同行质辩、不能量化的卓越——岂不同样只是一个竞逐失败者的自说自话？如果所有的价值都是由胜方回溯建构的，那么批评者有什么资格说胜方建构的那个价值是"假的"、输家那个才是"真的"？更尖锐的追问是：如果价值真的是回溯建构的，那么"反对压覆"的规范性力量从哪里来？

这个反驳必须被正面接住，而不能用"我们使用了更高级的理论词汇"来滑过去。

本文的回应是：这篇论文从头到尾，都没有宣称内在论证标准"更接近真正的价值"。本文使用的不是"真/假"的判断语法，而是生成条件的比较语法。两种论证语法——内在论证与对标论证——会生成不同类型的裁决实践、制度记忆和认知生态。本文做的，是指认出这两种生成路径在当代制度条件下的不对称存活前景——一种正在压覆另一种——然后把这个压覆的现场指出来让人看。

这不需要援引一个先在的"真正的价值"作为裁判。压覆可以被指认，不需要用"被压覆者更真"来作为依据。压覆本身就是够格的观察对象。当一份职称评审条例把"代表作制度"作为一个"供参考"的附件而把量化计分表作为"主要依据"时——或者更隐蔽地，当它把量化计分表设为"主要依据"而把代表作制度设为"可供参考"时——压覆就已经发生了。不需要你去证明代表作里的那篇论文比那个影响因子更"有价值"，你只需要指出：那个条例的设计，确保了代表作里的论文在裁决的运算中不占据实际权重。

然而，那个更深的追问仍然必须被面对：如果作者只是说"压覆正在发生"，那为什么要反对它？这里需要一个最低限度的规范性理由，它不能偷借一个被本文自身取消了的"真价值"前提。

本文反对压覆，不是基于"被压覆者更接近真价值"，而是基于一个最低限度的生态学理由：单语法的垄断会丧失裁决系统在环境突变时的存活弹性。当代学术制度的外部环境不是恒定的——它可能在十年内被一场新的技术革命（如大语言模型对知识生产本身的冲击）、一次治理哲学的转向（如对审计文化的全面反思）、或一个代际认知断裂（如新一代学者对学术生涯意义的重新定义）所根本性地扰动。当这样的扰动发生时，一个只有唯一裁决语法的系统，会因为丧失了备选的判断路径而无法在突变环境中存活。两种语法并存的价值，不在于其中某一方的"更真"，而在于系统作为一个整体的冗余度——在它的裁决机制中保留了另一个待激活的可能，即使那个可能在当前环境下正被压着打。本文批判压覆，不是为内在论证上香，而是为制度系统保留另一个尚在冬眠的脑区。这不需要援引价值的先验——这是最冷级别的系统存活冗余论证。

（二）如果内在论证那么好，为什么不把它"制度化"？

这是每个初次接触到裁决竞逐判断的读者都会问的问题：既然内在论证的弱点是接口缺失，那给它造接口不就行了？设立"代表作制度"，要求评审委员会写深度研判意见书，给深度质辩过程分配制度性的时间——这不是很多大学已经在做的事了吗？

这个问题的力量在于，它问到了竞逐的物理层。但它的漏洞也正在这里。所有试图"制度化"内在论证的努力，都会在一个地方撞墙——那个地方不在制度设计者的意志不够坚定，而在内在论证标准本身的本体论特征（本文第三节已详细解剖）：它产出的判断，是后置的、生成的、强索引性的。"制度化"要求的是：标准可以被提前制定、可以被第三方复检、可以在申诉中站住脚。而内在论证的标准——那个在质辩中被实时生产出来的"什么是这一次的推进"——无法被提前制定，无法被第三方在不参与那场具体质辩的情况下复检，无法被翻译成一个落选者也能自己计算的分数。这不是制度设计能力不足的问题，这是两种本体论形态之间的鸿沟。

在这个意义上，内在论证的"接口弱势"不是某种暂时没做好的制度建设，而是它的裁决方式本身在根本上拒绝被转换成标准化的度量。这意味着，给内在论证"造接口"不是一个技术改进问题，而是一个制度逻辑的断裂。要做这件事，不是设计一个更精密的计算公式，而是在某些裁决的关口上，明确地接受裁决所依赖的是不可被标准化的判断——接受这个判断不能被申诉、不能被复检、不能被做成公平的假象——并且仍然以此为依据分配资源。而这就等于要求整座大厦的合法性地基（程序公正、可问责、透明）为这一小块裁决的"深度"让路。

这不是制度改革的量级。这是制度决断的量级。还没有任何一个大型学术机构真正做出过这个决断。

而第四节所追踪的R大学案例，正是在这条线上给了一个经验的锚点：当制度改革试图在不对整体合法性地基做决断的前提下"增加内在论证接口"时，裁决逃逸与回写机制会系统性地新增的接口掏空，让它变成一个占着赋权条目位置、但裁决实际滑向别处的空壳。R大学的代表作制度改革，不是败在"制度设计不完善"，而是败在它试图用技术改进的手段（增加评议环节、增设外审评分）去解决一个只有制度决断才能破开的本体论冲突：你要赋权给一个不能被标准化的判断——那你准备好了接受由此产生的不公平申诉、不可复检的黑匣子、以及整个合法性框架的晃动吗？R大学没有准备好。也没有任何大学准备好了。它们的每一次"还没准备好"，都是内在论证标准的竞逐失败再次被结算的时刻。

（三）那些"反杀"的孤岛，会不会驳倒裁决竞逐？

一个更经验性的反驳是：如果内在论证注定在接口上败北，那么怎么解释那些它成功"反杀"的孤岛？譬如，普林斯顿高等研究院在自己的聘任决策中，几乎完全不依赖量化指标，而依靠资深成员长达数月的反复深度讨论。某些人文领域的顶级期刊，至今坚持不设影响因子追踪，只通过主编和编委会的集体判断来决定发表。这些孤岛的存在，是否说明内在论证可以在特定条件下守住自己的裁决权，从而构成裁决竞逐的反例？

这个反驳不是威胁，恰恰是本文判断的经验强化。审视这些孤岛，可以发现它们都有一个共同的特征：它们建立了独立的资源接口，不需要在既有的大规模审计驱动接口体系中争夺裁决权。普林斯顿高等研究院的经费不是靠学科评估排名来获取的——它有独立的巨额捐赠基金，不需要向任何外部审计者证明"我们的研究员每五年产出了多少高影响因子论文"。它的裁决接口不需要向外部的公平性焦虑负责——落选的申请者没有一套"申诉机制"可以撬动它的裁决。在这个意义上，它不是战胜了裁决竞逐——它是用独立的资源地基，把自己从竞逐发生的那个条件（共享同一套审计驱动的资源接口）里摘了出去。它没有给内在论证一个"通用解法"，它只是给了一个极端稀缺的例外。裁决竞逐预测：大多数没有独立资源地基的制度现场，无法复制这个例外。

同样的逻辑适用于另一些"内部反抗"的成功实践。当一个学科、院系或实验室通过建立自己内部的评价共同体、自设奖励与经费分配的同行评议机制，从而在不触碰学校层面量化指标体系的前提下，在自己的小生态内为内在论证保留了裁决空间时，这类实践的成功，恰恰不是因为它在既有接口内靠"更好的论证"打赢了对标论证——而是因为它在那个生态内新建了一套平行的裁决通道，让内在论证在那个通道里不被外部接口的赋权逻辑所挤压。这不但构成对裁决竞逐的反驳，反而印证了它的核心判断：取胜的要害不在论证的锋利度，而在有没有能力建立一套不依赖既有审计接口的裁决通道。这条路有效，但它同时也恰恰暴露了竞逐的残酷性——能这样做的单位永远是少数，而且它们能够这样做，往往是因为它们有某种外部不可复制的资源保护层。

但这是否意味着，所有身处审计生态中的行动者都是彻底的囚徒？不完全是。孤岛的启示在于：在微观和中观层面，确实存在着在大审计生态内部创造"局部接口飞地"的缝隙。那些成功在院系或实验室内部建立了平行的同行评议机制的实践表明，只要一个小生态能够在其内部成员之间维持一个不被外部指标逻辑穿透的裁决空间——哪怕这个空间很小、很不稳定、随时可能被下一次评估运动冲垮——它就能在那个空间内部为内在论证保留一段时间的存活。这个存活不是"战胜"了竞逐，而是在竞逐的缝隙中找到了一个暂时的避难所。它的存在不构成对裁决竞逐判断的反驳——恰恰相反，它的脆弱和暂时性印证了：只要外部审计接口的压力持续存在，这个飞地就永远处于被压覆的风险之中。

（四）行动者能否主动利用回写机制逆转竞逐？

最后一个反驳——也是最值得被认真对待的一个——来自对本文"无主体的结构运动"倾向的质疑。如果价值是裁决之后回溯建构的，那么内在论证的支持者为何不能主动发起一场成功的裁决，然后通过事后叙事将这场裁决回写为"内在论证赢了"？既然回写机制存在，它为什么不能被竞逐中的任何一方所利用？本文目前的论述似乎在暗示竞逐是一个只有对标论证能从回写中获益的单向过程，但这在机制上并不必然——回写是一把谁都可以用的刀。

这个反驳的力量在于，它迫使本文必须正面回答：为什么在当前的制度条件下，内在论证几乎无法利用回写来逆转竞逐？

答案在回写机制本身的结构。回写不是任意的——它受到叙事接口的严格限制。回想第四节R大学案例中的那个细节：当B在委员会内部获得了一场关于"她在质疑框架本身"的深度质辩之后，这场质辩无法进入官方新闻稿。为什么？因为新闻稿的叙事接口——"顶刊""前沿""国际水平"——是由对标论证的语法事先写好的。内在论证产出的判断——"质疑框架本身""不是标准推进"——在这套叙事语法中没有可用的句式。你不能在一个用于向上级汇报、向全校公示的官方文本里，写"B副教授虽然外审评分略逊，但她的论文指出了一个主流框架无法处理的结构盲区"——这句话要求读者（非同行管理者、上级部门）去做一套只有深度同行才做得出的判断，而新闻稿的受众不是同行。

这意味着，内在论证要利用回写来逆转竞逐，它首先必须改写叙事接口本身——它必须让"深度质辩的内容"能够被公开地、制度性地表述为裁决的依据。但叙事接口不是孤立存在的——它被赋权接口（计分表上的权重条款）和认知接口（裁决者自动感知"更好记录"的倾向）三层夹在中间。要改写叙事接口，必须同时改写赋权接口（让深度质辩在计分表中占据不可绕过的权重）和认知接口（让裁决者在读完代表作后能自动地给出一个不依赖数字的判断）。而改写赋权接口和认知接口，就等于回到前面第（二）点已经论证过的那个不可能性：它要求一个制度决断——让整个合法性地基为裁决的"深度"让路。这不是内在论证通过"发起一场成功的裁决"就能做到的。因为要发起那场裁决，它首先需要

一个已经被改写过的赋权接口——否则它的判断在裁决的运算中根本不会占据权重。这是一个先有鸡还是先有蛋的困境：内在论证需要先赢一次裁决来改写接口，但它在当前的接口下赢不了裁决，因为当前的接口不给它的判断赋权。

这就是为什么回写机制虽然在逻辑上是中性的——谁赢谁写——但在实际的制度条件下，它几乎总是把竞逐的结果焊死在有利于对标论证的方向上。因为对标论证不需要改写叙事接口：它产出的数字——影响因子、H指数、达标率——天然地能被翻译成公开的、可被非同行理解的语言（“顶刊”“高水平”“国际前沿”）。它对叙事接口的占有，不是它在每一次裁决之后主动去“回写”的结果，而是叙事接口本身的结构就预先为它的语法留好了位置。内在论证要在这个叙事结构里插进自己的句式（“质疑框架本身”“不是标准推进”），它要对抗的不只是一个敌对的叙事者，而是一整套已经固化在赋权、叙事和认知三层接口中的语法结构。

这就是为什么在当前的制度条件下，内在论证利用回写逆转竞逐的可能性微乎其微。不是因为行动者“没有主观能动性”，而是因为主观能动性要在接口允许的通道里才能被结算为制度性的效果。而当前的接口通道，不为内在论证的主观能动性提供结算的接口。

七、怎么用、怎么证伪、以及这个发现在理论上的意味

（一）最小分析步骤

裁决竞逐是一个操作概念，不是一个装饰概念。要让一个同行读完这篇论文后，能在自己的制度现场激活它，至少需要一套最小分析步骤。

第一步，定位裁决接口。不是找“指标在哪里”，而是找那件被称作“客观依据”的东西——计分表、达标率、影响因子——它在实际的政策里所处的那一栏，与它不占权重而只作为“参考”放进去的那栏之间，是怎么分别被赋权的。

第二步，扫描接口占有率。在同一类裁决（如晋升、经费分配）中，对标论证接口覆盖了多大比例的主要赋权条款？内在论证接口（代表作评议、同行推荐信、深度质辩记录）在这些主要条款中是作为裁决依据，还是作为解释性的注脚？

第三步，识别竞逐的结算事件。找到一个最近发生的具体个案：在某个裁决关口，内在论证产出了一个判断（“她的研究更有原创性”），对标论证产出了另一个判断（“他的影响因子更高”），裁决最终采纳了其中一个。追查：那个被采纳的判断，是因为它“说服了委员会”，还是因为它“对接了不可绕过的制度接口”？

第四步，追回写。追踪那个个案在此后几年里的集体叙事：当初的裁决被如何描述？那个被采纳的接口在事后故事里是出场了，还是隐没了？那个输掉的判断被如何记忆——是“当时确实讨论过，但后来结果证明更对的那个赢了”，还是“这个事已经过去了就不提了”？

第五步，外推。用这四步去分析另一个领域（如K12学校的教师评价、医院的绩效管理、基金会的项目评审），看在那里控制着裁决通道的是哪些接口，属于内在论证还是对标论证。

（二）证伪条件

本文的核心判断——"在审计驱动的制度环境中，对标论证通过接口占有和回写，在裁决竞逐中系统性压覆内在论证"——可以被以下事实所撼动。

第一，如果在一个高度指标化的学术机构里，职称晋升的裁决结果在统计上显著偏离量化计分表的预测，且这种偏离不只是个别案例，而是构成了一个可复现的样本群——即存在一批指标不佳但依靠深度质辩获得晋升的学者，并且这批学者在晋升后同样获得了持续的资源支持——那么本文对接口占有致命性的判断就需要修正。

第二，如果有一个大型学术机构，在正式制度层面完成了一个决断：在关键的人事晋升和资源分配中，明确规定量化指标只是辅助信息而非裁决依据，把深度同行质辩过程（而非质辩之后填的计分表）规定为唯一裁决依据，并为此承担了相应的申诉风险和行政成本——而且这个制度安排在实施五年之后没有被回退或被架空——那么本文关于"内在论证无法被制度化"的判断就需要被重新斟酌。

第三，如果在未来出现了一种技术或制度创新——它能在不消解内在论证判断的后置性、生成性和强索引性的前提下，为它提供一个能够在审计生态中存活的接口（可追溯、可申诉、可被外部审阅）——那么本文关于"内在论证在本体论上无法被接口化"的判断就需要被撤回。

在这些事实被广泛观测到之前，裁决竞逐的经验尖锐性仍然站立。

（三）理论收束：裁决竞逐的三个意涵

收束全文，裁决竞逐的发现对于理解当代制度生活、批判理论与知识论本身，至少在三个方向上包含意涵——它们是这篇论文不愿被快速归类为"又一篇批判指标的文章"的原因。

第一个意涵，它在制度研究的领域里首次指认了一种非范式性的、非利益集团性的竞逐形态。裁决竞逐的双方不是在争夺同一个规则的"解释权"——它们争夺的是"规则本身应该由哪一种语法来写"。这不是范式之争，因为胜负与论证的质量无关。这不是利益集团的博弈，因为竞逐发生在论证语法的层面，不在利益团体的协商桌上。它是一种在制度理论里尚未被充分命名的冲突形态：发生在论证本身的结构性的竞逐。这对制度研究意味着：我们不能再把制度中的冲突仅仅还原为"谁赢了、谁输了"的资源分配结果（政治学范式），或"哪一种认知框架占了上风"的观念史叙事（社会学范式）。有一种冲突，是两套论证语法本身在制度接口上的不对称存活，它塑造了政治学的赢家和观念史的"常识"，但它本身不等于前者也不等于后者。

第二个意涵，裁决竞逐对批判理论提出了一个冷峻的修正。批判理论——无论是法兰克福学派、布迪厄还是福柯——都有一种隐含的假定：批判可以"揭示"被权力遮蔽的真实，从而为改变打开空间。无论这个"真实"是未被异化的人的潜能、场域中被压抑的另一种合法性、还是被真理体制排斥的另一种知识。裁决竞逐挑战的，不是这个揭示动作的有效性——那个动作本身仍然是真实的、有力的——而是它能够在制度层面产生效果的那个后设前提：那个被揭示的"真实"，在被揭示之前，必须仍然作为一个可被援引的实体在制度中活着。如果它已经在竞逐中持续缺乏制度性反馈直到匿迹，那么揭示它就不是在"揭开遮蔽"，而是在从化石中尝试还原一个已经匿迹的判断实践的声音。对批判理论来说，裁决竞逐是一个清理地基的追问：你批判的那个参照，在那里吗？还是你一直在用它的化石写悼词，而自己并不知道？这个追问不使批判无效——它使批判必须开始追问自己的对象是否"在"。

第三个意涵，也是最根本的，裁决竞逐把知识论中的一个习惯假设推向了它的断裂点。知识论长期接受一个前提：真的东西，不会因为不被承认而变成假的。伽利略的力学在教廷判决之后依然是真

的。裁决竞逐对此的回答是：在制度的纬度上，不被承认够久，真的东西不会变成假的——但它会变成不可感知的。它不假，它只是找不到了。这是一种新的知识论后果：不是认识论上的"被驳倒"，而是制度本体论上的"不可用"。当某种知识的整个评价语法、训练方法和裁决通道整体匿迹之后，"真"仍然是一个形而上的属性，但它不再是制度中可操作的、可被下一代学人体认的、可以进入下一次裁决论证的东西。"真"还在，但它在制度世界里的肉身匿迹了。

裁决竞逐不在"真"的层面击败对手，它在"可被裁决"的层面让对手匿迹。这是本文所试图命名的全部不适的根源所在。

（四）带回决策会的问题

这篇论文不追求被赞同。它追求的是被使用。使用它的方式，不是写一篇论文引用它——那本身就是一种特定的存活方式。使用它的方式，是把它带回下一次你参加的、由数据围起来的决策会，在议程推进到KPI差距清单的时候，问一句话。这个问题不需要超过二十个字：

"哪个指标，是我们没有勇气不放进去的？"

如果在那之后，讨论没有滑入"我们怎么改进这个指标的精度"，而是有一瞬间——哪怕只有十几秒——大家停了下来，发现那张表上缺的不是更多指标，而是另一个不能被任何指标翻译的追问。那么，这一瞬间的存在本身，就构成了本文全部论证的、一个不可被任何指标度量的证成。

注释

-
- [1] 古德哈特定律由经济学家查尔斯·古德哈特在1975年提出，通常表述为"当一项指标被用作政策目标时，它就不再是好的指标"。本文仅将其作为测量失效论的常识性背景提及。
- [2] 引言所述"冷叙事"纪律意在悬置内在论证常被赋予的悲情化叙述，但作者对压覆并非无立场。第六节以系统存活冗余为基础给出了最低限度的规范性理由，此处不赘。
- [3] 本文对布厄符号暴力与误识的讨论，主要依据其《区隔》（1979）及《实践理论大纲》（1977），侧重场域斗争中支配标准被自然化的机制，不延伸至其后期的"符号资本"理论。
- [4] 福柯的治理术概念主要集中于其1977-1978年法兰西学院讲座《安全、领土与人口》（2004），真理体制分析则贯穿《规训与惩罚》（1975）等著作。本文仅调用其指标作为规训工具的基本面向，不涉及治理术与主权、规训之间的复杂谱系。
- [5] 拉图尔与伍尔加的信用循环模型见《实验室生活》（1979）。本文只取其承认-资源转换循环的基本结构，不涵盖拉图尔后期行动者网络理论的全部主张。
- [6] 第四节案例基于中国多所研究型大学职称评审改革的公开政策文本、官方新闻稿以及学术界的相关讨论与笔者的观察综合而成。为保护机构隐私并突出分析机制，学校名称经匿名化处理，具体时间节点与数据细节经过必要简化与合成，主要作为裁决竞逐机制的典型例示。

深化增补·全球近邻补切（编辑增补）

说明：以下各节为编辑在审读时增补，不改动作者正文与核心判断，只补上本文原稿未正面处理的最近邻理论，并把分离线写成可操作的判别。每一条都给出"若观察到什么，本文即错"的对照预测。

与埃斯佩兰、索德「反应性」与「可通约化」的分离线：行为被改变，与裁决权被转移

本文动用了坎贝尔定律、古德哈特定律、审计社会与福柯，覆盖面相当宽，但漏掉了这个题目上最直接的那对机制：埃斯佩兰与索德（Espeland & Sauder, 2007）关于排名**反应性**的研究，以及埃斯佩兰与史蒂文斯（Espeland & Stevens, 1998）关于**可通约化**的分析。前者论证排名通过自我实现预言与叙事重构改变被评价者的行为与自我理解；后者论证把异质之物换算为共同尺度本身就是一种权力操作。这是本文所处地盘上最负盛名的既有工作。

分离线在于**因变量**。反应性研究的因变量是**被测者的行为与自我理解**：法学院为了排名调整招生、调整就业统计口径、重写自我叙事。本文的因变量是**裁决权的归属**：谁有资格宣布什么算价值。这两件事可以分离——一个机构可以完全不为排名调整任何行为（反应性为零），却仍然在任何需要对外说明的场合只能拿出计分表说话（裁决权已转移）。

由此得到一条可判决的对照，而且本文第六节提到的普林斯顿高等研究院正是它的检验场：反应性研究预测，一旦撤掉排名压力，行为就会回摆；本文预测撤掉之后裁决权也收不回来——因为论证语法本身已被改写，即使在完全自主的场合，评议者也会不自觉地先找可计分的支点。若在那些不受排名约束的孤岛上，评议过程中计分式支点的出现频率与受约束机构没有差异，则本文的语法改写命题成立；若孤岛内部的评议确实以纯内在论证进行，则本文需要把结论收窄为「在有外部问责的接口上」才成立。

与博尔坦斯基、泰弗诺《论辩护》的分离线：秩序平权的争执，与检验成本不对称的败局

把「价值由谁定义」写成一门系统学问的，是博尔坦斯基与泰弗诺（Boltanski & Thévenot, 1991）：社会中并存多种正当性秩序，每一种都有自己的价值尺度与「检验」形式，争执通过检验来裁决，也可以在秩序之间达成折衷。本文的竞逐与这套框架讨论的是同一件事，因此必须切开。

分离线在于**对称性**。《论辩护》的框架预设各秩序在原则上**平权**：争执的结果取决于哪一种检验在当下情境中更适切，因此结局是开放的。本文的判断是**不对称的**——在两者交汇的接口上，计分表一侧的检验成本极低（可复制、可上诉、几分钟出结果、不需要检验者具备被检验领域的能力），内在论证一侧的检验必须由有能力的人亲自读完并承担判断，成本高且不可复制。因此在任何需要向不具备该能力者说明的场合，败北不是偶然。

这条分离线把本文的核心判断收敛成一个**可以直接计量的东西**：同一项决策中，两类检验各自的耗时、所需资质门槛、可复制性与可申诉性。若测得两类检验在这些维度上大体相当，而内在论证仍然败北，则本文给出的败因是错的，应回到《论辩护》式的情境适切性解释。

与坎贝尔定律的边界确认：指标失效，与失效之后仍保有裁决权

本文已多处使用坎贝尔定律与古德哈特定律，此处只需把边界写清，因为本文的主张比它们强得多。两条定律讲的都是**指标一旦成为目标就丧失度量效力**——这是一个关于指标失效的判断。本文的主张是：**失效之后，指标并不因此丧失裁决权**。

这是一个更强、也更需要证据的命题，因为它要解释一件反直觉的事：所有当事人都知道那个数不准，却仍然只能拿它来定输赢。本文把这一点归因于接口处的成本不对称，因此它的证伪条件也随

之明确：若能找到这样的案例——指标被公开证明失效之后，裁决权确实回到了内在论证一侧，且回归不依赖任何外部强制（不是靠上级下文件取消排名），那么本文的机制不成立。

本次增补所核验的文献

- Boltanski, L., & Thévenot, L. (1991/2006). *On Justification: Economies of Worth*. Princeton University Press.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40.
- Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology*, 24, 313–343.
- Muller, J. Z. (2018). *The Tyranny of Metrics*. Princeton University Press.

参考文献

- Bourdieu, P. (1977). *Outline of a Theory of Practice* (R. Nice, Trans.). Cambridge: Cambridge University Press. (Original work published 1972)
- Bourdieu, P. (1979). *La distinction: Critique sociale du jugement*. Paris: Les Éditions de Minuit. (English translation, 1984, *Distinction: A Social Critique of the Judgement of Taste*, Harvard University Press)
- Campbell, D. T. (1976). Assessing the impact of planned social change. *Occasional Paper Series*, Paper No. 8. Hanover, NH: The Public Affairs Center, Dartmouth College.
- Foucault, M. (1975). *Surveiller et punir: Naissance de la prison*. Paris: Gallimard. (English translation, 1977, *Discipline and Punish: The Birth of the Prison*, Pantheon Books)
- Foucault, M. (2004). *Sécurité, territoire, population: Cours au Collège de France, 1977-1978*. Paris: Seuil. (English translation, 2007, *Security, Territory, Population: Lectures at the Collège de France 1977-1978*, Palgrave Macmillan)
- Hood, C. (1991). A public management for all seasons? *Public Administration*, 69(1), 3-19.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Latour, B., & Woolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts*. Beverly Hills: Sage Publications.
- Power, M. (1997). *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.