

伦理自噬：制度维生机制如何制造自身的合法性危机

不是制度的不同部分之间打架，而是制度的同一个动作在自身内部同时完成了保护与伤害——这不是矛盾，这是自噬

高鹏 著 · SDE 学员 · 制度理论与组织社会学 · 2026年7月27日

摘要 现代制度面临一种独特的困境：制度为维持自身运转和伦理权威而采取的管理动作，恰恰在消耗那些制度无法内生、却又赖以生存的伦理资源。本文通过整合法哲学、道德心理学和组织社会学的理论资源，提出“伦理自噬”这一分析框架，将制度退化的根源锚定在“愧疚工具被反转”这一特定机制上。研究表明，制度的两种基本命令形态——“强迫做好事”与“禁止做坏事”——预设了截然不同的主体伦理响应模式：强迫性命令剥夺个体的道德判断权，而禁止性命令则为主体保留了自主愧疚的空间。然而，这一愧疚自检程序在制度运行中会发生方向性反转：当制度通过“合规收编”（将裁决者个人判断纳入标准化考核）来维持执行链的稳定时，它一方面豁免了裁决者的个人风险以保障制度的持续运转，另一方面却在同一动作中杀死了制度赖以生效的“个人担保”这一异质残余。被制度激活了愧疚能力的主体，最终将这项能力用于攻击制度本身的合法性。本文进一步识别出这一悖论在法学、教育学、管理学和政治学中的五种结构同构现象，并提出了一个可操作的机制证伪设计：通过比较同一法律体系内多个法院在“担保制度化密度”上的自然变异，检验“伦理自噬”机制与“规则过载论”的竞争解释力。这一框架超越了传统制度退化理论将衰败归因于外部冲击或内部僵化的单向叙事，揭示了制度坚固性与脆弱性的双向同源性——两者来自同一组维生动作的不同侧面。

关键词：伦理自噬；制度退化；愧疚机制；合规收编；担保悖论；命令形态

1 引言

1.1 问题的提出：制度完善为何加速合法性流失

当代社会面临一个令人困惑的经验事实：越是精密化的制度，似乎越容易陷入合法性危机。司法体系通过细化法官裁判标准来确保“同案同判”的改革，却引发了法律从业者对“机械司法”的普遍不满；教育系统通过教师评价指标来保障教学质量的努力，反而催生了教师的“道德正确疲劳”和实质教育的空心化；企业通过关键绩效指标管理来提升效率的制度设计，常常以产品质量的异化和员工职业荣誉感的丧失为代价。这些现象并非孤立的组织失败个案，它们共享一个深层结构：制度为维持自身生命力所采取的标准动作——完善规则、细化考核、收编异质资源——恰恰在消耗那些制度无法自我生成、却对其伦理权威至关重要的资源。

传统的制度理论在解释这类现象时，通常诉诸两种路径：一是“外部冲击论”，将制度衰败归因于技术变革、经济危机或政治突变等外生变量；二是“内部僵化论”，用路径依赖、规则过载或官僚主义来解释制度的功能失调。这两种路径各自揭示了制度退化的部分真相，但它们都无法解释一个关键的悖论性事实：在上述案例中，制度并非因外部冲击而被动衰退，也不是因墨守成规而丧失灵活性，恰恰相反——它是在主动“完善自身”的过程中加速了合法性流失。制度的维生动作与蚀生结果之间，存在一种尚未被充分理论化的内部因果链。

本文将这一悖论凝缩为一个核心命题：制度用以维持自身伦理效力的核心工具（愧疚自检机制），会被制度自身的维生动作（合规收编）转化为攻击制度合法性的扳机。我们将这一过程命名为“伦理自噬”，意指制度在管理其赖以生存的伦理资源时，恰好杀死了那些无法被制度化的、却构成制度生命力来源的异质成分——正如免疫系统在过度反应时攻击自身健康的细胞。

1.2 问题的边界：强迫、禁止与制度伦理效力的来源

要理解伦理自噬的发生路径，首先需要回到一个看似基础却长期被忽视的问题：制度命令的形态本身，已经预设了不同的伦理主体响应模式。日常经验告诉我们，“必须这样做”与“不许那样做”对人类道德心理的激活方式截然不同。当一种制度以“禁止做坏事”为基本形态时，它在划出不可逾越的边界的同时，为主体留出了自主判断“何以为善”的空间，个体的愧疚自检工具因而处于活跃但自我管理的状态。而当制度转向“强迫做好事”的形态时，它要求主体交出内在动机的评判权，被塞入一个标准化的“善”的模板——愧疚失去了自我校准的对象，要么退化为对制度的依赖（“我做对了规定动作”），要么异化为对制度的反抗（“你凭什么替我定义什么算好”）。

本文将论证，正是禁止性命令所激活的这种“愧疚自检工具”，构成了制度伦理效力的核心来源——因为禁令的生效不仅依赖文本的清晰性，更依赖裁决者和被约束者各自的伦理自律。而恰恰是这一工具，在制度的维生过程中会被反转方向：当制度为保障执行链的连续运作而将裁决者的个人判断纳入标准化考核体系时，它一方面豁免了裁决者的个人风险从而保证了“可持续的执行”，另一方面却在同一个动作里抽空了裁决者在规则文本失效的关键时刻所必需的“我自己也押上去”的伦理资源。被制度以禁止命令激活了愧疚能力的主体，

最终将这项能力用于攻击制度本身的合法性基础。

1.3 本文的理论目标与主要贡献

基于上述问题意识，本文旨在达成三项目标。第一，在理论层面，建立一个能够同时解释制度的“坚固性”与“脆弱性”来自同一组动作的双向同源框架——超越将制度退化归因于单一外因或内因的单向叙事。第二，在机制层面，精确锚定“愧疚工具反转”的发生条件与因果链：命令形态的预设结构作为反噬的触发条件、担保的不可制度化属性作为反噬的靶心、制度演化的对称性断裂逻辑作为反噬的不可逆性的解释。第三，在实证层面，提出一个可操作的机制证伪设计，允许研究者区分“伦理自噬”与“规则过载”这两种竞争解释，从而将这一分析框架从理论建构推进到可接受经验检验的科学命题。

本文的主要贡献呈现在三个方面。理论贡献上，我们识别出一个未被既有文献以同一结构命名的悖论现象：不是“制度失效”，不是“对抗性主体生成”，而是“制度用以维持伦理效力的核心工具，如何被制度自身的维生动作转化为攻击制度的扳机”。这一命题同时推进了法哲学中关于“权威的悖论”的讨论、道德心理学中关于“愧疚的社会功能”的研究、以及组织社会学中关于“制度退化”的分析。经验贡献上，本文提供了跨越五个学科领域的结构同构案例（法学、教育学、管理学、社会学、政治学），初步验证了这一悖论的跨场景可迁移性。方法论贡献上，本文展示了一种“可证伪的理论建构”路径——在提出理论的同时，明确指明何种经验证据可以推翻它，以此回应社会科学中理论过多而裁决不足的困境。

1.4 论文结构

本文的论述将按以下顺序展开。第2节进行文献综述，在法哲学的权威悖论、道德心理学的愧疚机制、组织社会学的制度退化论三个领域中定位本文的对话位置。第3节构建“伦理自噬”的理论框架，提出三个核心命题，并显式说明跨学科调用的有效性与边界。第4节详述本文的分析方法论——概念分析与多学科结构同构比较的结合。第5节是论证的核心，分为三个子论点和综合讨论，逐一展开“命令形态的预设结构”“担保的合规收编”“担保不对称的演化逻辑”三个子机制，并设置独立章节回应反驳预期。第6节给出本文的结论，包含理论贡献的凝缩、实践含义、研究局限的诚实自陈，以及三个可生长的未来研究方向。最后，第6.6节单设“可证伪性检验”专节，完整保留竞争解释排除的检验设计。

2 文献综述

本文的核心问题——制度维生动作为何导致伦理效力丧失——横跨三个研究传统：法哲学对“权威悖论”的讨论、道德心理学对“愧疚”社会功能的探究、以及组织社会学对“制度退化”机制的分析。本节分别在每个传统中梳理其核心洞见与盲区，最终定位本文的理论切入点。

2.1 法哲学中的权威悖论：强制与自主的张力

法律权威的正当性基础是法哲学的核心关切。自韦伯（Weber）以来，法律社会学便区分了法律的“事实有效性”与“规范正当性”（Weber, 1978）。这一区分揭示了一个根深蒂固的悖论：法律若要有效运行，必须借助外部强制；但若仅依靠强制，法律便不过是暴力的一种形式，丧失了真正的权威。哈特（H. L. A. Hart）的“内在视角”是这一讨论的里程碑：法律的规范性不仅来自制裁的威胁，更来自至少部分社会成员对规则作为行为标准的“接受”（Hart, 1961）。后继学者如拉兹（Raz, 1986）提出“权威的服务性概念”，主张权威的正当性在于它帮助个体更好地遵循那些本就适用于他们的理由，而非替代个体的实践理性。肖尔（Schauer, 2015）则从“规则的强制执行”角度指出，规则之所以有效，恰恰因为它不需要在每个个案中重新开启道德推理。

然而，这一传统在解释本文所关注的现象时存在两个盲区。其一，法哲学主要讨论法律义务的“来源”之正当性（为何应当守法），较少触及法律命令的“形态”差异——禁止与强迫——如何预设了不同的主体响应模式，以及这种预设如何反过来塑造法律权威的可持续性。其二，法哲学对“权威的衰落”的解释多诉诸外部因素（如社会整合的瓦解、价值多元主义的挑战），缺少对权威内部自毁机制的系统分析。本文正是要补充这一内部视角：制度通过禁止命令激活了主体的愧疚自检工具，而这一工具在制度完善自身的过程中会发生方向性反转。

2.2 道德心理学中的愧疚机制：从道德情感研究到“内在制裁”

道德情感作为行为调节的关键机制，在心理学和神经科学中得到了广泛研究。愧疚（guilt）与羞耻（shame）的区分是这一领域的经典命题。刘易斯（Lewis, 1971）率先指出，羞耻指向整体自我（“我做了坏事，所以我是一个坏人”），而愧疚指向具体行为（“我做了一件坏事”），因此愧疚相较于羞耻更可能引发修复行为而非退缩。坦格尼和迪林（Tangney & Dearing, 2002）进一步发展了这一区分，发现愧疚倾向与亲社会行为正相关，而羞耻倾向更常与外化愤怒和回避相关。神经科学研究表明，愧疚涉及前扣带皮层和脑岛的激

活 (Shin et al., 2000), 这提示愧疚具有一种“内部警报”的功能——当个体感知到自己的行为与内化的道德标准偏离时, 这套神经回路被激活并驱动修正行为。

然而, 道德情感研究的重心一直是“愧疚如何促进个体遵守规范”, 而非“愧疚如何可能成为质疑规范本身的扳机”。换句话说, 现有研究将愧疚默认为制度内部的一个“约束工具”, 其功能在于降低违规率。但本文关注的经验现象——被管理制度的伦理主体用制度自身的伦理标准攻击制度——提示愧疚可能具有一种未被充分研究的“外部功能”: 当个体将自己的行为与“本应如此”的标准进行比较时, 这项比较能力也可以被用于比较制度本身与它自己宣称的标准之间的距离。这是一个方向性反转: 愧疚从“我审视自己是否够好”转向“我审视这套让我觉得自己不够好的制度是否够好”。

2.3 组织社会学中的制度退化论: 路径依赖与规则过载

制度退化是组织社会学的核心议题。该领域的主导解释框架经历了从“外部冲击论”到“内部矛盾论”的转变。

路径依赖理论强调制度早期的偶然选择会自我强化, 形成“锁死”效应。皮尔逊 (Pierson, 2004) 系统化了这一机制: 制度设置使退出成本递增, 制度学习效应使替代路径难以想象, 制度互补性使局部改革因与其他制度不兼容而失败。路径依赖论的核心解释变量是“可选择路径范围的收窄”, 其预测是: 制度越是沿着某一路径发展, 就越容易丧失适应外部变化的灵活性。然而, 这一逻辑无法解释本文关注的现象: 在司法体系、教育系统、企业管理等案例中, 制度的衰败并非因为选择太少而被锁死在次优路径上, 恰恰相反——制度在不断“拓宽”其规则覆盖范围、表现出高度的管理创新性。其衰败的动力不是来自外部变化的“不适应”, 而是其维生动作本身在制造伤害。

规则过载论从组织行为学的角度提出制度退化的另一路径: 当规则密度超过个体的处理能力时, 执行者会从“依规行事”滑向“应付规则”。这一视角在公共管理研究中得到广泛讨论 (Bozeman, 2000)。规则过载论的核心变量是“规则密度”, 其预测是: 规则越多, 伦理判断的空间越受挤压, 制度最终因执行僵化而衰败。这一逻辑较之路径依赖论更贴近本文的经验起点, 但它仍将制度退化视为一种量变的结果 (“太多了”) 而非质变的结果 (“同一动作生成了相反的效果”)。本文的论点是: 制度的伦理失败不是规则总量挤压了判断空间, 而是制度在收编那些无法被规则化的异质资源 (裁断者的个人担保) 时, 通过“合规收编”这一具体动作把活的伦理判断变成了死的管理标签。问题不是“太多”, 而是“收编”这一动作本身具有消灭收编物的内在属性。

新制度主义的矛盾论在以上两条路径的基础上, 已开始注意到制度的内部矛盾性。斯科特 (Scott, 2014) 指出制度的规制、规范、文化-认知三大支柱之间存在张力, 瑟伦 (Streeck & Thelen, 2005) 揭示了制度变迁中的渐进转型和内生的变革动力。这些研究为本文的“内部矛盾”视角提供了理论资源, 但它们的分析仍停留在制度要素之间的“不相适应”或“摩擦”层面, 未触及本文所锚定的那个特殊机制: 不是制度的不同部分之间打架, 而是制度的同一个动作——为保护自己而收编异质资源——在自身内部同时完成了“保护”与“伤害”两个效果。这不是矛盾, 这是自噬。

2.4 文献缺口的精确诊断与本文的定位

综合以上三个研究传统, 本文识别出两个精确的文献缺口。

第一, 在法哲学与道德心理学的交叉处, 缺少对“制度形态与愧疚方向”的匹配性分析。法律理论关注命令的正当性而不关注命令形态如何塑造主体的道德情感结构, 道德心理学关注愧疚的个体功能而不关注愧疚如何被不同制度形态激活并可能反转方向。本文将这两个领域咬合在一起, 提出“命令形态预设愧疚方向”的命题。

第二, 在组织社会学的制度退化论中, 缺少对“维生动作即蚀生动作”这一自噬类机制的理论化。现有理论要么将退化归因于外部冲击, 要么归因于内部僵化或规则过载, 都无法解释制度在“主动完善自身”时为何反而加速衰败。本文以“担保的合规收编”与“担保的不对称性演化”为核心机制, 填补这一理论空白——退化不是制度被撞坏了, 是制度自己在修自己的过程中拆掉了承重墙。

基于上述诊断, 本文的理论定位如下: 在法哲学的权威悖论与道德心理学的愧疚机制之间, 建立制度形态的中介分析; 在组织社会学的退化论中, 引入一条新的机制路径——“伦理自噬”, 用以解释制度的坚固性与脆弱性如何来自同一来源。

3 理论框架

本节给出“伦理自噬”理论的完整结构: 首先定义三个核心概念, 然后提出三个环环相扣的命题, 最后显式说明本文跨学科调用的实质咬合基础与有效边界。

3.1 核心概念的操作化定义

命令形态: 制度向主体传递行为规范的语法形式。本文区分两种基本形态——“禁止性命令” (“不许做X”) 仅划出不可逾越的边

界，为主体保留自主判断“何以为善”的空间；“强制性期待型命令”（“必须做Y”）要求主体交出内在动机的评判权，被嵌入标准化的预设中。两种形态的核心差异不在行为要求的内容，而在其对主体愧疚工具有截然不同的激活模式。

担保：禁令生效所依赖的、裁决者本人承担伦理风险的个人判断。担保的核心属性是“不可制度化”——它无法被写进制度文本保存并被标准化复制，因为它本质上是裁决者在特定情境中用自身信誉或风险“押在判断上”的身位性动作。担保是制度无法内生生产的“异质残余”，只能由具体的、可受伤的个体在具体的情境中提供。

「不可制度化」这一属性必须与两个极近的既有概念划清界线，否则本文的核心命题会被认为只是它们的重新包装。其一是默会知识。波兰尼所指出的那类知识——我们知道的比我们能说出来的多——其定义性特征正是不可被完全形式化编码，这与担保的不可制度化在表面上几乎重合。但两者的不可编码性来源不同：默会知识之所以说不出，是因为它的载体是长期练习积累的身体技能，其障碍是**表达上的**；而担保之所以不能被写进制度，不是因为它难以描述——「我为这个判断承担后果」这句话非常好写——而是因为一旦它被写成必须履行的条款，承担后果这件事就从裁决者自己决定的动作变成了被指派的动作，它的伦理重量恰恰来自「没有人要求他这么做」，其障碍是**性质上的**。默会知识被写下来只是损失了精度，担保被写下来则改变了种类。其二是古德哈特定律：当一项指标成为目标，它就不再是一个好指标。这条定律描述的是被测量者的策略性反应——人们会针对指标优化，从而使指标与它原本代理的东西脱钩。本文所论的收编要更深一层：它不需要任何人做出策略性反应就已经完成了。即便所有法官都完全真诚、没有任何一个人人为考核而调整行为，「我以我自身担保这个判断」在被写进考核维度的那一刻，也已经变成了「我完成了一项合规判断」——性质的转换发生在指标设立的瞬间，而不是发生在被测量者开始博弈之后。古德哈特定律解释的是指标为什么会失真，本文解释的是某一类被测对象为什么在被测量的同时就已经不存在了。

愧疚工具的反噬：制度的禁止命令在主体内部激活了愧疚自检程序——“我是否做了不该做的事？”——这项自检能力在制度运行中可以发生方向性反转，自检对象从“自身是否违规”转向“制度自身是否够好”。反噬的发生条件是：制度在试图通过合规收编来维持执行链的稳定性时，恰好暴露了制度自身与被激活的主体的愧疚标准之间的差距。

3.2 三个核心命题

命题1（命令形态的预设结构）：“禁止做坏事”与“强迫做好事”两种命令形态，预设了截然不同的伦理主体位置。禁止性命令通过划出边界的方式，将“何以为善”的判断权交还给主体自身，因而激活主体内部的愧疚自检工具并使其保持自我管理的状态。强制性期待型命令则要求主体交出评判权，被纳入标准化的预设中，从而使愧疚退化为对外部标准的依附或拒绝。因此，制度的伦理效力在禁止性命令形态下更高——因为禁令的生效需要被约束者在愧疚的驱动下主动参与。

命题2（担保的合规收编悖论）：禁令生效不仅依赖规则的清晰性，更依赖裁决者以个人身份押上的“担保”。这一担保是制度无法自我生产的异质残余，只能由可受伤的具体个体提供。当制度出于维持执行链稳定的需要，将这一担保残余“收编”入标准化管理体系（如制定细化裁量权的考核指标）时，收编这一动作本身会杀死收编物——因为在被指标化的瞬间，担保从活的情境化伦理判断变成了死的合规标签。担保的性质被转换，而不是被保存。这就是“维生即蚀生”的原型操作。

命题3（担保不对称性的演化逻辑）：制度在演化过程中，为保障执行者的持续、安全运转，逐步豁免裁决者的个人风险（用职业身份担保替代个人身体担保）。这一“保护”动作创造了“可持续的执行者”，但同时抽空了裁决者在规则文本失效的极限时刻所必需的“我自己也押上去”的判断资源。这种担保不对称性使制度能够在常态下高效运行，但在面临规则文本无法覆盖的临界情境时，执行者发现自己手中只剩下规则文本，任何超越规则文本做出判断所需的伦理资源已被制度自身的保护动作预先拆解。

三个命题之间存在严密的因果链条。命题1给出反噬的“弹药”——愧疚工具被禁止命令激活并交到主体手中。命题2给出反噬的“靶心”——制度收编担保的动作暴露了制度自身与它所宣称的伦理标准之间的裂隙。命题3给出反噬的“不可逆性”——制度自身在每一次“完善”改革中都在加速这一过程的进程，因为完善改革的实质往往是更深、更系统性地收编残余担保。

3.3 跨学科理论调用的实质咬合基础与边界

本理论框架同时调用了法哲学、道德心理学和组织社会学三个学科的理论资源。这里显式说明跨学科调用的有效性基础，以避免“装饰性引用”的嫌疑。

法哲学的调用基础与边界。本文从法哲学调用的核心资源是“权威的悖论”——法律若要有效需借助强制，但强制又消解法律作为权威的规范性。这一逻辑与本文机制的实质咬合点在于结构同构：法律权威的悖论是外在强制与内在接受之间的张力，而本文的“伦理自噬”是制度的标准化管理与不可被标准化的担保之间的张力。两者都是“制度自身的有效性条件与制度自身的管理动作相冲突”这个母题下的子代。有效调用边界在于：法哲学主要处理“应然”层面的正当性问题（权威为什么值得被服从），本文则处理“实然”层面的机制问题（权威为什么在实际运行中衰落）。这两个层面不可直接互推。

道德心理学的调用基础与边界。本文从道德心理学调用的核心资源是“愧疚的内部自检功能”——愧疚驱动个体对照内化的道德标准

修正行为。该资源的实质咬合点在于可迁移性：愧疚的社会功能不限于个体内部，它同样可以被“外部对象”激活和反向应用——制度本身也可以成为愧疚自检的对象。有效调用边界在于：道德心理学研究的是愧疚的普遍心理机制，而本文研究的是制度形态对这一机制的不同激活方式以及激活后的方向反转。前者是后者的必要条件，但并非充分条件——制度形态的中介作用是本文的独有贡献，不能从道德心理学直接推导出来。

组织社会学的调用基础与边界。本文从组织社会学调用的核心资源是“制度退化的内生性”——路径依赖与规则过载都揭示了退化可以不需要外力而来自制度自身的运作。该资源的实质咬合点在于共同的前提：退化可以是内生的。本文在此基础上提出一条新的内生退化路径——不是在数量上“太多”或选择上“太窄”，而是在性质上每一个维生动作用于异质资源时都在同时执行蚀生。有效调用边界在于：路径依赖与规则过载仍然将退化链接到“量变”（路径范围、规则密度），而本文讨论的是“质变”——收编这个动作本身的属性。不能将本文的机制视为规则过载论的一个子类。

4 方法论

本章说明本文的论证方法及其局限性。本文主要采用概念分析与多学科结构同构比较相结合的路径，属于理论建构型研究，而非实证检验型研究。

4.1 研究设计：理论建构型的概念分析

本文的核心研究任务是提出一个尚未被既有文献以同一结构命名的理论机制——“伦理自噬”。这一任务的性质决定了其方法论路径应是理论建构而非经验检验：研究目标不是用数据去“检验”一个已有假设，而是通过概念分析与逻辑推演去“生成”一个新的假设路径，并给出可供后续实证研究使用的检验设计。

具体而言，本文的分析逻辑包括四个步骤。第一步，概念区分：将“命令形态”（禁止/强迫）作为独立的分析维度从法哲学的权威讨论中切出，将“担保”从制度执行的条件变量中切出，将“愧疚反转”从个体道德情感研究中切出。第二步，机制链建构：将上述三个概念用因果逻辑串联为“命令形态激活愧疚自检→制度收编担保以维生→收编杀死担保、愧疚被反转方向”的完整链条。第三步，竞争解释排除：从既有理论中识别出对本现象最具解释力的竞争假设（规则过载论），并对比说明两个机制在预测方向和因果路径上的差异。第四步，可证伪条件设计：提出一个可操作的检验设计，在控制竞争解释的变量后，若本文机制独有变量与因变量之间不存在剂量-反应关系，则本文理论为伪。

4.2 分析策略：多学科结构同构比较

为初步验证“伦理自噬”机制不是特定学科领域的偶然发现，而是具有跨场景的可迁移性，本文采用“多学科结构同构比较”策略。操作方法是：在法学、教育学、管理学、社会学、政治学五个领域中，各自识别出一个与本文理论机制具有相同结构内核的经验现象——即制度的维生动作（完善规则/指标化/收编非正式实践）导致蚀生结果（伦理效力丧失/基本功能的实质异化）——并逐一说明同构性所在与学科特异性所在。这一策略的目的不是用五个案例“证明”理论为真（个案数量不足以建立因果推断），而是展示该机制在不同制度环境中反复出现的结构模式——这构成了理论可迁移的初步证据。

4.3 方法局限的预先承认

本文的方法论存在三个主要局限，需要在论证过程中诚实面对。其一，理论建构阶段的概念分析对机制内部的逻辑自治性可以给出较为严密的论证，但无法替代后续的经验实证。本文在可证伪条件设计中提出的 $N \geq 12$ 法院的比较设计，只是“可被检验”的承诺而非“已被检验”的结论。其二，多学科结构同构比较虽然展示了机制的跨域可迁移性，但由于各学科案例是作者基于文本资料建构的“理想型抽象”而非第一手经验研究，可能存在选择性偏差——只选取了支撑理论的案例而忽略了反例。这一局限的应对策略是：在下文的综合讨论中主动考虑可能的反例场景，并给出“如果存在反例，本文机制应如何调整”的说明。其三，本文的三个核心命题目前主要以逻辑推演和说明性案例支撑，尚未经过系统的比较个案研究或定量检验。这意味着本文产出的不是“已被验证的理论”，而是“可供检验的命题”；其学术价值在于提出了一个更精确的假说，而非提供了一个被验证的解释。

5 分析与讨论

这是本文的核心论证部分。前三节（5.1-5.3）逐一展开三个核心命题的完整论证；5.4节主动设置反驳预期并予以回应；5.5节进行综合讨论，将三个命题整合为统一机制，并与竞争解释做系统性比较。

5.1 命令形态的预设结构：禁止如何激活愧疚的自反潜能

任何制度命令都不只是对特定行为做出“应当”或“不得”的指示，它同时向被命令者传递了一个隐晦的关于“你是什么样的人”的预设。禁止性命令——“你不得杀人”——预设了一个有能力理解这一边界、有能力控制自己不跨越这一边界、并被信任有充足的内在资源去判断“何以为善”的主体。强制性期待型命令——“你要乐于助人”——则预设了一个需要被告诉什么是善、且其内在动机评判不足为主体保留判断权的主体。

这一区分的实质不在于命令内容的严厉程度——某些强制性命令可能比禁止性命令更严苛——而在于命令对主体愧疚工具的不同激活模式。愧疚是一种内部自检程序：当个体感知到自己的行为与自己内化的道德标准之间存在差距时，愧疚被激活并驱动修正行为。禁止性命令恰好为这一自检程序提供了操作空间：它给出的是什么“不”能做，但没有代替主体决定什么“该”做。这要求主体在每一次面临选择时都重新启动自己的道德判断，并在判断可能的失误后面对愧疚驱动的修正。强制性命令则相反：它为主体提供了“善”的标准答案，好的行为被预定，坏的偏离也被预定。主体只需将自身行为与外部标准对照，偏差即是错、符合即是对——这恰恰是羞耻而非愧疚的运作模式。在羞耻模式下，“我做得不够好”并非导向自主的修正努力，而是导向对外部标准的依附或对标准的憎恨。

由此可以得到“愧疚工具的反噬”的第一环：禁止性命令所激活的愧疚自检程序，具有方向性不锁定的属性。自检动作本身是形式中立的——它是“我是否做了不该做的事”与“我是否本应做得更好”之间的自我比较——而这项比较能力可以脱离原初的目标对象（自身行为是否违规）而应用于任何被主体关切的对象上。最关键的迁移是：主体可以开始用这项能力去比较制度宣称的道德标准与制度实际兑现的道德标准之间的距离。“你说不许我偷盗，你把你自己的执行者保护到连为我主持公道时都不需要押上自己一份——那你自己的标准在哪里？”这不是主体在外部攻击制度，这是主体用制度自己交给他的内部工具在制度内部撕出裂口。

5.2 担保的合规收编：制度维持自身的动作如何歼灭异质残余

禁令生效并非单纯依赖文本的清晰性与制裁的确定性。一项禁令之所以能够让被约束者在内在道德情感层面感到“我不该这样做”而不仅仅是“我不能这样做因为会受罚”，依赖的是一个隐蔽但关键的条件：裁决者本人承担了“我以我自身的信誉或风险押在这个判断上”的身位性动作。传统社会中，一位长者的“你不该做这件事”之所以具有道德分量，不仅因为他说出了规则，更因为在说出的同时，他将自己的威望、与听者的关系、甚至自己在社区中的安全押在了这一判断上。如果他的判断被证明偏私或不公，受伤的不仅是规则——他自己也要承担后果。这就是“担保”的本源形态。

担保的根本属性是“不可制度化”。它之所以能承载伦理重量，正因为它无法被事先写成标准文本，无法被纳入考核表格，无法在脱离具体情境和具体裁决者的身体之后仍然存活。试图将担保“制度化”——即将其转化为可管理的、可考核的、可复制的内容——是收编而非保存。当法院出台内部规范将“法官裁量权的审慎行使”细化为五个评分维度、每个维度三个等级时，发生的不只是管理精细度的提升。发生的是担保的质变：此前法官在个案中做出的“我认为这个判断是正确的，我为此负责”的个人押注行为，此后变成了“我按内部考核要求完成了一项合规判断”的组织任务。担保的“身体性”——那种“我以自身作为担保物”的风险暴露——在前一种情况下在场，在后一种情况下被抽离。担保被保存了形式，却失去了实质。

这就是“维生即蚀生”最浓缩的操作原型。制度的维生动作——细化裁量标准以保障“同案同判”、减轻个别法官被追责的风险、使执行链可复制和可持续——其效果是真实的。制度确实因此更稳定、更可管理。但蚀生在同一动作里完成：由于被收编的担保残余恰是制度稳定性来源中不可被自身生产的那一部分，收编本身就在消灭收编物。每项旨在“保存”担保的合规指标，都在把担保推得更远。而由于这一蚀生过程以“完善制度”的形式出现，制度的每一次自我强化都加速了蚀生——制度很难在这一悖论中刹车，因为刹车本身也意味着停止“完善”，而“停止完善”在制度内部的评价逻辑中就是退步。

5.3 担保不对称性的演化逻辑：执行者如何在被保护中丧失判断资源

禁令的担保结构在制度化演进中经历了一次关键的断裂。在口头传统或早期法律共同体中，担保是大致对称的：裁决者用自己的身体或社会信誉担保裁决的公正性，同时被约束者用自己的身体或社会信誉担保自己遵守裁决。裁决者与被约束者都在同一个担保机制中暴露于风险。而现代法律体系的一个核心制度成就是将裁决者从个人担保中解放出来——用职业身份担保取代个人身体担保：法官判案不需要拿自己的身家性命做押注，他的判决即使后来被证明错误，职业身份本身的存续（在多数情况下）也不受威胁。

制度这么做的理由十分正当：保护裁决者免于因裁决而受个人报复、确保裁决可以不带个人恐惧地做出、使法官能够独立地基于法律而非基于人际关系判决。这确实极大地提高了法律系统的稳定性和可复制性——维生动作有效。但同时，这一“保护”也在裁决者的伦理判断力中制造了一个空腔：当规则文本清晰、适用明确时，职业身份担保足以支持判断。但当规则文本失效或产生荒谬结论的临界时刻——那种“按法条应该这样判，但任何一个没有被制度抽空伦理直觉的人都看得出来这是荒唐”的情境——裁决者发现自己没有可押的自身。他的身体受制度保护不暴露于风险，但恰恰因为不暴露于风险，他失去了用来抵抗制度文本荒谬性的唯一资源：他不能说“我以我自身担保这个条文的这种适用方式是错的”，因为“我以我自身担保”这个句式在他的职业伦理中已被撤销。

担保不对称性构成了伦理自噬链条的最后一个环节：它解释了为什么这一悖论是系统性的而非偶然的，为什么制度不会从自己的失败

中学习而纠正。因为每一次“纠正”——加强法官培训、细化裁量标准、追加考核指标——都恰恰在进一步掏空执行者手中可用于超越规则文本的残余担保。完善不是在解决问题，而是在解决方案的名义下推进问题。制度在长期演化中，为自己制造了一群高效、可持续、但手无寸铁的裁决者——而这个“高效可持续”正是蚀生的最终代价。

5.4 反驳预期与回应

本节主动设置三项对本文核心论点的有力反驳，并逐一回应。

反驳一：这不过是“规则过载”的另一种表述。制度出的问题不是收编，而是规则太多了，任何把裁量权纳入考核的做法只是在增加规则量——真正起作用的是规则密度而非收编动作本身。

回应：本文承认规则过载是制度退化的一个重要路径，但伦理自噬机制指向的是一个独立于规则密度的因果环节。核心区隔在于因果路径的差异。规则过载论的因果链是“规则太多→处理不了→执行走样”，其驱动力是认知负担超限所致的被动失效。伦理自噬的因果链是“制度收编异质担保→收编动作本身性质转换了被收编物的属性→伦理资源死亡”，其驱动力是维生动作本身包含的歼灭效应，且失效在规则执行层面上可能表现为“一切合规”——执行走样不是因无能而是因过分精准。如果伦理自噬机制为真，则我们应当在经验上观察到：在控制了规则密度之后，担保制度化密度（收编强度）仍独立地与逆向愧疚爆破率正相关。这一预测将在结论部分转化为可操作的检验设计。如果未来的实证检验发现这一独立相关关系不成立，本文坦然接受被证伪并将制度退化归入规则过载论的管辖范围。

反驳二：本文夸大了裁决者担保的重要性。如果裁量权越大、越依赖个人担保的制度事实上越不可靠——这正是现代司法制度用规则拘束裁量权的原因。本理论似乎是在倡导回到人治。

回应：这是一个重要的误解防范。本文的分析目标不是政策倡导——不是在说“应该回到个人担保而不应该使用规则拘束”。本文的分析目标是机制解释——在说“当制度用规则收编个人担保时，它同时发生了什么呢？”本文承认规则拘束裁量权带来的稳定性和可复制性收益是真实的，并且其价值在许多情境中高于被牺牲的个人担保。本文指出的只是：这个牺牲是真实发生的，不能假装担保被“保存”在规则里而没丢；如果某一天在某些临界情境中制度因缺少担保而失效，这种失效不是由于外部偶然冲击，而是制度自身的维生动作所必然积累的结构性的空腔。理解这一机制，并不等于倡导撤销规则回到人治——正如理解免疫系统的自噬不等于倡导停掉免疫系统。理解是为了更准确地评估在何种情境下担保的收编代价已经超出其收益，以及是否可能存在不通过“收编”而通过“保护其不被收编”的方式来保有残余担保。

反驳三：本文过度依赖愧疚这一特定道德情感，而不同文化中对同一命令形态的情感响应可能完全不同。西方的愧疚文化与东方的羞耻文化存在根本差异，理论的文化边界不清。

回应：本文承认愧疚与羞耻的跨文化差异是一个重要变量，且确实限制了本文机制的适用边界。本文的回应分两步。第一步，重新明确本文所锚定的“愧疚”不是特定文化的日常情感倾向，而是由禁止性命令的预设结构所激活的“自检程序”——“我是否做了不该做的事”。这一自检程序在形式上是跨文化可操作的：只要一个制度采用禁止性命令形态并向被约束者传递“判断权在你”的预设，就会刺激这种自检启动。不同文化可以给出不同的命名（愧疚/内疚/良心不安），其情感色调和应对策略也可以不同，但“将自我与标准做比较”这个抽象操作是一致的。第二步，承认本文的论证材料和说明性案例集中于中国与欧美的法律与教育制度，确实不足以覆盖更广泛的文化变异。这意味着本文的命题3——关于担保不对称性的演化逻辑——可能在不同法律传统（如伊斯兰法、习惯法主导的体系）中有不同的表现形态甚至不成立。这一局限需要在未来研究中通过跨文化比较个案研究来处理。

5.5 综合讨论：伦理自噬的完整链条与理论意义

将5.1到5.3节的三个命题串联起来，伦理自噬的完整链条如下。第一阶段，制度采用禁止性命令形态，在划定行为边界的同时将“何以为善”的判断权交还给主体，激活主体内部的愧疚自检工具。这一阶段是制度伦理效力的来源——因为禁令的生效需要被约束者在愧疚驱动下主动参与自我约束。第二阶段，制度为维持执行链的稳定与可复制，通过对裁决者个人判断进行标准化收编来降低执行风险和裁量偏差。这一维生动作在提升执行效率的同时，将裁决者的个人担保从活的情境化伦理判断转化为死的合规标签——性质转换而非保存。第三阶段，被第一阶段激活了愧疚自检工具的主体，在第二阶段中看到了制度自身的行为与制度所宣称的伦理标准之间的裂隙，将愧疚自检能力的方向从“审视自身是否违规”反转为“审视制度是否够好”——逆向愧疚爆破发生。第四阶段，制度在每一次通过“完善规则”来应对新的合法性挑战时，实质上是进一步深化了对残余担保的收编。这导致担保不对称性加剧，执行者手中的判断资源进一步被掏空，制度越来越依赖规则文本，也越来越暴露于规则文本失效临界时刻的崩溃风险——不可逆的蚀生循环由此形成。

这一机制的深层理论意义在于它揭示了制度坚固性与脆弱性的双向同源性。这一制度的坚固性来自何处？来自禁止命令对主体伦理自律的激活，来自规则化执行对执行链稳定的保障，来自对执行者的保护使其能够不带恐惧地履职。这些都在常规运行中真正有效。制度的脆弱性来自何处？来自同一组动作的另一侧面：禁止命令激活的愧疚可以反转方向，规则化收编杀死担保，对执行者的保护抽空其判断资源。坚固与脆弱的来源完全相同，区别只是观察的时刻：常规状态看一面，极限状态看另一面，但机制在运行之初就已经双面俱在。

这种双向同源逻辑使我们可以重新评估制度退化论的既有诊断。当我们面对一个正在走向衰败的制度时，通常的做法是诊断其“弱点何在”——哪里被外部冲击打穿了，哪里被内部僵化卡死了——然后提出修复方案。伦理自噬的视角给出了一个更不安的诊断：在很多时候，所谓“弱点”并非故障，而是制度正常运行的另一面。衰败不需要来自外部冲击或内部僵化，它是制度在正常维生过程中积累的结构性质副产物。这意味着某些针对衰败的“修复”——增强规则密度、细化考核指标、强化合规管理——不是在解决问题，而是在解决方案的名义下推进问题。这不是在说任何改革都不可取，而是在说改革的方向必须审慎于识别：哪些动作是在收编残余担保从而加速噬生，哪些动作是在保护残余担保不被收编从而延缓噬生。这是两种根本不同的改革逻辑。

最后，将本文机制与最主要的竞争解释——“规则过载论”——做一简要对比，以锁定本文的真值条件。规则过载论预测制度衰败与规则总量正相关，衰败是渐进的、量变的、由认知超载驱动的；伦理自噬论预测制度衰败与收编异质担保的强度正相关（在控制总量后），衰败可突变（当最后一层残余担保被收编时）、是质变的、由伦理资源被歼灭驱动的。两个机制并非完全互斥——一个制度可以同时面临规则过载和伦理自噬——但它们指向不同的干预方向。如果规则过载论为真，干预应在“精简规则总量”上着力；如果伦理自噬论为真，干预应在“改变收编方式、保护异质残余免遭指标化”上着力。下文结论部分将给出区分这两者的具体检验设计。

6 结论

6.1 核心发现的简明陈述

本文的核心发现可以用一句话凝缩为一个可被引用的命题：制度为维持自身伦理效力而采取的标准管理动作（合规收编），恰好会激活并反转制度自身依赖的伦理工具（愧疚自检），从而使每一次制度自我完善都加速了合法性资源的流失。我们称之为“伦理自噬”。

这一命题包含三个子机制，分别揭示了反噬的触发条件、靶心位置与不可逆性逻辑。第一，禁止性命令通过将“何以为善”的判断权交还给主体，激活了主体的愧疚自检工具，而这一工具的方向可以被反转为对制度自身的审视。第二，制度为稳定执行链而对裁决者个人担保进行的“合规收编”，实质性杀死了担保——因为担保的核心属性（不可制度化）使其无法在被指标化后仍然存活。第三，制度在演化中为保护执行者而豁免其个人担保，创造了高效可持续的执行系统，但同时抽空了执行者在规则文本失效的临界时刻所必需的“我自身作为担保”的伦理资源。这一套伦理自噬机制在法学、教育学、管理学、社会学和政治学五个领域中均有结构同构现象，构成跨场景可迁移的初步证据。

6.2 理论贡献

本文对既有研究的具体修正和推进呈现在三个层面。

对法哲学权威悖论讨论的推进。既有研究将法律权威的脆弱性主要定位于“强制与接受的张力”这一外部悖论。本文补充了一个内部悖论：制度用以维持效力的工具（愧疚自检）本身也可能成为拆解效力的扳机。这意味著权威的脆弱性不仅源于“外部人是否接受”，更源于制度自身如何处理其无法内生、却又赖以维生的异质伦理资源。此外，本文通过切入“命令形态”这一被既有法哲学忽视的分析维度，将“禁止与强迫的制度伦理后果”从中抽离为一个独立的评价轴——不依赖权威的正当性来源，而取决于命令形态的预设结构对主体响应模式的形构效果。

对道德心理学愧疚功能研究的延伸。既有研究将愧疚的社会功能限定在“促进个体遵守规范”的内部维度。本文识别出愧疚的“外部功能”——将道德审视的能力从“审视自身是否够好”迁移至“审视那套让我觉得自己永远不够好的制度自身是否够好”。这不是一个与既有研究相矛盾的发现，而是一个扩展方向：愧疚的社会功能不仅包括保守性的约束违规，也包括变革性的反向检测。

对组织社会学制度退化论的补建。既有退化理论诉诸路径依赖（选择空间收窄）、规则过载（规则总量超限）和外部冲击（外生变量介入）。本文提出了一条新的内生退化路径——不是量变而是质变：制度正常的维生动作本身，在作用于某些特定类型的资源（不可制度化的残余担保）时，会同时产生维生和蚀生两种效果。这条路径补充了既有退化论无法解释的一种经验形态——制度在“主动完善自身”过程中衰败。

6.3 实践与政策含义

将本文的理论命题转化为可操作的实践指导，可以得到三条具体建议。

第一，在司法制度建设中，区分“保护裁量权”与“收紧裁量权”两种改革逻辑。当改革以“细化裁量标准考核”的方式收紧裁量权时，它实质在做的是担保的合规收编——将法官的个人判断从活的伦理担保转化为死的合规标签。这种改革在短期内确实可以降低裁量偏差、提升“同案同判”的统计指标，但长期代价是法官群体伦理判断力的空心化。一种替代方向是“保护裁量权”：不试图将裁量标准完全化约为可考核指标，而是为法官在特定边界内做出的不可被事先指标化的判断提供制度性保护——让“我以我自身担保这个判断”仍然

是一个可操作的伦理动作，而不立刻被合规管理消解。

第二，在教师评价改革中，设置“不可指标化的伦理保护区”。教师育人热忱的本质不是一种可被分解为指标序列的技能（“关心学生”无法被善意量化成“每月谈心次数”），而是一种由教师个人向学生发出的、包含身位担保的回应。将这项热忱指标化——例如将其细化为“育人日志填写完整度”和“学生满意度评分”——所实现的管理收益（可考核、可比较）是以消灭收编物为代价换来的。一种折中策略是：在评价体系中识别并明确划定一块“不允许指标化”的保留地，将教师的部分重要判断行为（如对特定学生困境的自主回应）从考核体系中豁免出来。这不是对教师免于问责，而是承认某些核心伦理资源无法被收编而不死。

第三，在组织管理中，对“合规伦理竞赛”保持警觉并设立反向信号。当企业将员工职业荣誉感指标化并在组织内部建立横向比较机制时，它同时在做两件事：表面上是“激励荣誉”，实际上是把荣誉从自发的职业伦理理解变为外部强制的合规负担。由于被指标化的荣誉不再能承载“我自己以此为荣”的自主意义，员工在履行这些指标时会产生“道德正确疲劳”——一种“我知道我所做的只是在文化表演上得分但我已经懒得区分它与真荣誉”的状态。组织管理可以在设立量化指标的同时，引入反向信号机制：定期追踪“员工对指标本身是否产生嘲讽情绪”或“指标博弈行为是否普遍化”，视之为管理体系进入自噬阶段的早期预警。

6.4 研究局限的诚实自陈

本文存在四个主要局限。

第一，经验实证的阶段局限性。本文的论证处于理论建构阶段，三个核心命题主要以概念分析和说明性案例支撑，尚未通过系统的经验研究加以检验。可证伪条件的设计给出了理论自己指定的“翻车条件”，但该设计本身尚未被执行——这是承诺而非完成。

第二，跨学科同构案例可能存在选择性偏差。5.5节给出的五个学科的结构同构案例，是作者基于文本资料建构的理想型抽象，其选取可能有“寻找支持性案例而忽略反例”的偏差。一个本文未能处理的潜在反例是：某些高度制度化的系统（如现代金融监管体系）同样经历了深度的“个人判断合规收编”，但似乎并未出现明显的逆向愧疚爆破。这一现象的合理解释可能是金融监管的对象（机构而非个体）本身不具备愧疚自检工具，但这也意味着本文机制的适用范围在“被约束主体的类型”维度上存在尚未划清的边界。

第三，文化边界的未充分处理。本文对愧疚机制的讨论，主要基于西方道德心理学和中国法律与教育制度的说明性案例，未系统地处理不同文化中道德情感（愧疚与羞耻的不同权重）和命令形态合法性基础的差异。在愧疚文化较弱的社会中，禁止性命令是否仍然能有效激活愧疚自检工具，并使其具备逆向爆破的潜能，是需要文化比较研究来处理的问题。

第四，干预含义的工程性有限。本文虽然给出了三条实践建议，但尚未细化到可被一线管理者直接操作的程度。例如，“设置不可指标化的伦理保护区”在实践中如何操作化、如何防止“被保护的区域”在下一轮改革中再次被收编，这些是需要进一步工程化研究的问题。

6.5 未来研究方向

基于本文的理论框架和研究局限，以下三个方向可作为具体的后续研究纲领。

方向一：司法体系中的担保制度化密度比较研究。按照第6.6节（见下）的检验设计，选取同一法律体系内多个层级的法院（控住审级、案件类型结构、法官年均负荷等变量），测量各自的“担保制度化密度”（近五年旨在细化裁量权考核的内部规范性文件项数），并将其与“逆向愧疚爆破发生率”（以程序不当为由上诉且二审支持的案件占比）进行回归分析，控制规则总量密度。这项研究可以直接检验伦理自噬机制的独立性，并提供第一手的制度干预证据。

方向二：教育系统内“不可指标化保护区”的实验设计。选取一批同质化程度较高的学校，随机分配于两组：实验组在教师评价体系中明确划出部分重要判断行为（如对特殊困境学生的自主回应）为“不计入考核指标的范围”，控制组保持全部行为的指标化考核。追踪两组教师在后续两年内的“道德正确疲劳”自我报告水平、“实质教学投入”（学生深度访谈评定）和逆向行为（投诉制度、抵触考核等）的发生率。这项研究可以检验“停止收编是否能部分恢复被侵蚀资源”这一对伦理自噬论至关重要的推论。

方向三：不同命令形态制度演化路径的历史比较研究。选取法律传统中存在命令形态差异的体系（如偏重禁止性命令的普通法体系与偏重强制性期待型命令的部分大陆法体系），长时段追溯两者在“担保制度化”的演化速度和“执行者伦理判断力空心化”的症状出现的时点，比较是否存在系统性差异。这一研究可以检验命题1关于“命令形态预设了不同伦理主体响应模式”的推论，并观察命题3关于对称性断裂不可逆程度的文化变异。

6.6 可证伪性检验：排除规则过载论的一个研究设计

作为本文最硬的学术分量，本章节完整给出一个可操作的机制证伪设计，允许后续研究者实证地区分“伦理自噬”与“规则过载”这

两种竞争解释。

（一）本理论面对的最强竞争解释：规则过载论。 规则过载论的承重变量是“规则密度”——制度文本的总量与复杂性。它的底线预测是：在控制了其他相关变量后，制度衰败（伦理空洞化、执行走样、合法性流失）与规则密度正相关。规则过载论对本文所讨论的“逆向愧疚爆破”的解释是：这是一种认知过载的产物——被约束主体因为规则太多、太复杂而增加了对制度不满的概率，并非因为收编担保这一特定动作激发了独立的因果关系。因此，规则过载论预测“逆向愧疚爆破”与“规则总量”相关，而与“担保制度化密度”在控制了规则总量后无独立相关。

（二）研究设计。 样本选取：同一法律体系下的N≥12个地级市中级法院，要求样本在“担保制度化密度”上有自然变异（即某些法院近五年出台了大量细化裁量权考核的内部文件，另一些法院则相对保持裁量空间的开放性），同时控制审级、法律文本总量、社会经济水平等变量。自变量操作化：“担保制度化密度”= 该院近五年出台的、旨在细化法院裁量权并将个人判断纳入标准化考核的内部规范性文件项数（按相同编码标准由两名独立评判人计数，信度需达 $\kappa \geq 0.85$ ）。因变量操作化：“逆向愧疚爆破发生率”= 该院在同期内，以“程序不当”或“裁量标准不统一”为由被提起上诉、且二审因同样理由改判或发回重审的案件占该院同期总案件的比例（控制案件本身的复杂度，如排除简易程序案件）。控制变量：法官年均案件负荷、法院信息化建设程度、辖区律师密度、案件类型结构（刑/民/行比例）、规则总量密度（每万字的制度文本条数）。研究假设：伦理自噬论预测，在控制了规则总量密度和其他控制变量后，“担保制度化密度”与“逆向愧疚爆破发生率”之间存在正相关；规则过载论预测，两者之间在控制了规则总量后不存在独立相关关系，因变量的变异应全部被规则总量解释。

（三）何种结果会推翻本理论。 如果上述研究发现“担保制度化密度”与“逆向愧疚爆破发生率”之间不存在显著的正相关（在 $\alpha=0.05$ 水平），则伦理自噬论对司法领域这一机制的独立性主张为伪。这一结果意味着制度的伦理空洞化应全部归因于规则总量负担，而非制度对担保的收编动作——干预方向应从“审慎收编”转向“精简规则总量”。如果结果是正相关，则伦理自噬论获得初步的经验支撑——制度的衰败在一定程度上独立于规则总量，而取决于制度如何处理那些无法被它自身生产的异质伦理资源。两者的差异不仅在理论层面有实质意义，在实践层面也指向截然不同的改革路径：前者要求“做减法”，后者要求“改变收编方式”。

这一设计有一处必须被预先承认的弱点：因变量「逆向愧疚爆破发生率」以二审改判或发回的比例来操作化，而这一比例同时受审级关系、上诉率、律师策略、辖区上诉成本等多重因素影响，与「主体用制度自己的伦理标准攻击制度」这一构念之间存在相当的距离。换言之，即便相关性被观测到，它也可能是别的东西造成的。为收窄这一距离，检验设计应同时纳入两项更贴近构念的辅助指标：其一，裁判文书说理部分中出现「虽合规范但……」「本院注意到……然现行标准」这类执行者自己标记出制度与自身伦理判断之间落差的表述的频次（可由独立编码员按统一手册计数）；其二，法官在匿名执业调查中对「我曾按规定作出过一个我自己并不认同的裁判」一题的肯定率。三项指标若同向变动，构念效度才算被初步支撑；若只有二审改判率单独变动而两项辅助指标不动，则该相关很可能来自审级因素而非本文机制，此时本文应当承认该检验未能支持自身。

参考文献

[1] Hart, H. L. A. (1961). *The Concept of Law*. Oxford University Press.

[3] Lewis, H. B. (1971). *Shame and Guilt in Neurosis*. International Universities Press.

[4] Pierson, P. (2004). *Politics in Time: History, Institutions, and Social Analysis*. Princeton University Press.

[5] Raz, J. (1986). *The Morality of Freedom*. Oxford University Press.

[6] Schauer, F. (2015). *The Force of Law*. Harvard University Press.

[7] Scott, W. R. (2014). *Institutions and Organizations: Ideas, Interests, and Identities* (4th ed.). Sage.

[8] Shin, L. M., Dougherty, D. D., Orr, S. P., et al. (2000). Activation of anterior paralimbic structures during guilt-related script-driven imagery. *Biological Psychiatry*, 48(1), 43–50.

[9] Streeck, W., & Thelen, K. (Eds.). (2005). *Beyond Continuity: Institutional Change in Advanced Political Economies*. Oxford University Press.

[10] Tangney, J. P., & Dearing, R. L. (2002). *Shame and Guilt*. Guilford Press.

[11] Weber, M. (1978). *Economy and Society: An Outline of Interpretive Sociology*. University of California Press. (Original work published 1922)

[12] Bozeman, B. (2000). *Bureaucracy and Red Tape*. Prentice Hall.

本次打磨说明

改动已就地落在正文，此处逐条说明落点与理由。

本篇改了两处。其一关乎不可还原性：本文的核心属性「担保不可制度化」与两个极近的既有概念高度重合，原稿一处未提，

命题会被合理地当作它们的重新包装。第三节因此补入正面切割。与波兰尼的默会知识：两者的不可编码性来源不同——默会知识之所以说不出，障碍是表达上的（载体是长期练习的身体技能）；担保之所以不能被写进制度，不是因为难以描述（「我为这个判断承担后果」非常好写），而是因为一旦写成必须履行的条款，承担后果就从裁决者自己决定的动作变成了被指派的动作，障碍是性质上的。默会知识被写下来只是损失精度，担保被写下来则改变种类。与古德哈特定律：该定律描述的是被测量者的策略性反应，而本文所论的收编不需要任何人做出策略性反应就已经完成——即便所有法官完全真诚，「我以我自身担保这个判断」在被写进考核维度的那一刻也已变成「我完成了一项合规判断」，性质转换发生在指标设立的瞬间，而非被测量者开始博弈之后。其二关乎检验设计的构念效度：原稿把因变量操作化为二审改判或发回的比例，而这一比例同时受审级关系、上诉率、律师策略等多重因素影响，与「主体用制度自己的伦理标准攻击制度」这一构念距离甚远。已补入两项更贴近构念的辅助指标——裁判文书说理中「虽合规但……」这类执行者自我标记落差的表述频次，以及匿名执业调查中「我曾按规定作出过一个我自己并不认同的裁判」的肯定率——并明言三项若不同向变动，本文应当承认该检验未能支持自身。另删去正文从未引用的十四条文献与利益冲突声明。

本次打磨所核验的文献

Polanyi, M. (1966). The Tacit Dimension. London: Routledge & Kegan Paul. （本次新增正面切割）

Goodhart, C. A. E. (1975). Problems of monetary management: The U.K. experience. （古德哈特定律的出处，本次新增切割）

Hart, H. L. A. (1961). The Concept of Law. Oxford University Press.

Raz, J. (1986). The Morality of Freedom. Oxford University Press.

Pierson, P. (2004). Politics in Time. Princeton University Press.

Bozeman, B. (2000). Bureaucracy and Red Tape. Upper Saddle River: Prentice Hall.