

惩罚边界归因性基础设施：罪刑法定原则的一种新制度框架

高鹏 著 · 发表于2026年7月24日 · 三种阅读方式

摘要

——

1 引言

2007年，中国最高人民法院、最高人民检察院联合发布《关于办理受贿刑事案件适用法律若干问题的意见》，将十种“新型”行为纳入受贿罪的规制范围——其中包括“以交易形式收受贿赂”“以收受干股形式收受贿赂”“以委托理财形式收受贿赂”等。这一司法解释并未宣布任何条文被修改或废除，它只是“发现”了已有法条中此前未被识别的含义。从形式上看，罪刑法定的四项子规则——事先公布、条文明确、禁止类推、禁止溯及——无一被违反：条文是在行为发生前就已公布生效的，解释是在条文文字的基础上做出的，解释的效力不溯及发布前的行为。

然而，任何一个诚实的刑法观察者都能注意到，在这一套完美的形式合规之下，发生了一件没有被写进任何立法档案的事：惩罚那些以“交易形式”收受贿赂的行为的权力，并不是由立法机关在制定受贿罪条文时做出的事先判断，它是由司法机关在2007年通过事后解释做出的判断。

这个案例不是孤例。它是过去三十年间全球主要成文法体系中一个更广泛趋势的缩影：在条文一字不改的前提下，通过司法解释权力的渐进扩张，惩罚边界的实质性决定权正从立法端向司法端沉默漂移。这个漂移不需要任何一个标志性事件来宣告它的发生——它只需要解释惯性把“严格解释”从信条降格为“一种值得尊重但不必遵守的立场”。

罪刑法定原则的传统辩护者通常将这种趋势诊断为两类问题之一：要么是立法技术粗疏（条文写得不够精确，所以给了解释者可乘之机），要么是司法越权（解释者明知条文没有这个意思，故意通过解释强加惩罚）。

本文的论证将推进一个不同于这两类诊断的判断：罪刑法定当前面临的最根本威胁，既不是立法技术的失败，也不是司法者的恶意。它面临的是第三种、也是更难用传统法学工具捕捉和修复的退化形式——惩罚边界的决定权被五个各自循规蹈矩的制度行为群体（立法者、检察官、法官、刑法学者、公共舆论）在日常接力中散布式地消解了。没有任何一个环节违反了现行法律，没有任何一个行为者承认自己正在削弱罪刑法定。他们各自在自己的制度位置上、按照各自面临的理性激励做出的正当行为，在系统层面无意间达成了同一件事：让“什么行为应受惩罚”这个问题的最终答案，不再属于立法者在制定条文时做出的一次性判断，而是由这五者在逐案接力中共同做出的、不可审查的、散布式的决定。

本文的目标，就是解剖这套“合作-瓦解”机制的内部结构，并提出一种制度性的回应框架。回应的核心逻辑是：罪刑法定的修复不能靠呼吁任何人“变得更高尚”——不能靠呼吁立法者写得更精确（语言的本质决定了精确有其极限）、不能靠呼吁法官克制解释冲动（博弈结构决定了克制不是占优策略）、也不能靠呼吁学者回归文本忠诚（学术评价的市场逻辑压制了这一呼唤）。修复只能靠改变制度博弈的规则——让每一次惩罚边界位移的代价，从“由未来共同体分摊”的不可归因状态，被拖入“由发起位移的制度节点当下即承担”的可归因状态。

本文将这个思路操作化为“惩罚边界的多层可归因性”框架——一种由四层程序性制度组成的归因性基础设施。本文的论证将沿着以下结构展开：第二部分回顾现有文献在解释性扩张问题上的主要研究路径及其局限；第三部分构建理论框架，用博弈论的代价不对称模型解释解释性扩张的底层驱动；第四部分解剖五个制度齿轮的分工瓦解机制；第五部分提出四层归因性的工程方法论；第六部分回应可能的反驳；第七部分总结整体贡献并规划未来研究方向。

2 文献综述

2.1 形式解释论与实质解释论之争

过去二十年中国刑法学中关于罪刑法定原则最持久的理论争论，集中在形式解释论与实质解释论之间。

形式解释论主张，刑法条文的解释应当严格限制在法条用语可能具有的含义（即“文义射程”）之内。当一个行为的可罚性在文义上存疑时，解释者应当做出有利于被告的选择——即宣告该行为不为罪——然后等待立法机关通过修法来填补惩罚空白。形式解释论的核心理由植根于罪刑法定原则的民主合法性与权力分立功能：如果惩罚的决定不是在立法过程中由民选代表做出的，而是在个案审理中由法官或检察官做出的，那么罪刑法定就失去了它作为“权力约束装置”的最根本意义。

实质解释论则主张，刑法条文的解释应当以法益保护为依归，以处罚的必要性（当罚性）和处罚的合理性（需罚性）为实质判准。当一个行为的法益侵害性达到了值得以刑法回应的程度，且该行为与法条文字之间不存在不可逾越的语义障碍时，解释者可以做扩张解释，将该行为纳入该条文的规制范围。实质解释论的核心理由是：如果每一处立法的语义“缝隙”都需要修法来填补，那么刑事立法将不堪重负，而大量具备同等法益侵害性的行为将在“缝隙”中逃脱惩罚——这反而是对罪刑法定所保护的法益价值的背叛。

这一争论延续至今已经产生了大量有价值的学术成果，但它在本轮讨论中的核心局限在于：争论双方都把罪刑法定的实效锚定在“解释方法”这一技术层面上——锚定在“文义解释优先还是目的解释优先”“扩张解释的边界在哪里”这类问题上。然而，正如本文将详尽论证的，解释性扩张不是一个解释方法的问题，它本质上是一个制度博弈均衡的问题。当一个解释者面对一个疑难案件时，他选择扩张解释还是选择宣告无罪，并不是由他信奉形式解释论还是实质解释论决定的——它是由他面临的两种选择的当下成本比较决定的。而当前的制度博弈结构，恰恰让扩张解释的成本远低于宣告无罪的代价，让扩张解释成为任何理性解释者——无论他个人信奉什么解释哲学——的占优策略。

2.2 制度主义路径

不同于聚焦于“解释方法”的争论，刑法学中还有一支关注“制度设计”的研究传统。这一路径认识到，罪刑法定的实效不取决于条文的精确性，而取决于解释权力的制度约束机制。

在国内学界，张明楷指出，罪刑法定原则的实质侧面要求“禁止处罚不当罚的行为”和“禁止不均衡的、残酷的刑罚”，这一判断不仅是对立法者的要求，也是对司法者的约束；司法者在使用兜底条款时，必须受制于“同类解释规则”——即被纳入兜底条款的行为必须在类型上与法条已列举的行为具有同质性。这一“同类解释规则”可以被理解作为一种制度化的自我约束机制，但它的核心问题在于，它的约束力完全依赖于解释者自身的适用意愿——当“同质性”本身可以被弹性解释时，这套自我约束在硬案件中就丧失了判别力。另有学者从比较法视角引入富勒“法律的内在道德”命题和德沃金“整全性”解释理论，试图为中国的刑事解释实践提供规范判准。其洞见在于，将罪刑法定的保障功能从“条文的语言精确性”转移到“法律共同体的解释伦理”上——这正是富勒“法律的内在道德”的核心关怀。

在国际层面，美国学者罗伯特·C·波斯特和尼尔·S·西格尔提出的“司法自制”理论，将司法解释的边界问题锚定在“司法角色认知”上；德国法上的“明确性要求”判准——由联邦宪法法院在系列判决中确立——要求刑事立法必须为公民提供“可预见的行为指引”，违反该要求的条文可被宣告违宪；日本法上，对“公诉权滥用”的判例积累则从另一个制度入口（检察裁量权约束）触及惩罚边界决定权归属的问题。

这批制度主义取向的研究，比形式解释论与实质解释论之争更准确地识别了罪刑法定的要害——要害不在于“如何解释”，而在于“谁有权做出最终的决定”。但它们的共同局限在于，它们倾向于把“制度约束”等同于“某个中央化的审查机构（如宪法法院）在终端对解释产品进行否决”。这类中央化解构的设计有两个问题：第一，它的覆盖面极窄——绝大多数惩罚边界位移发生在日常个案中，从未走进宪法法院的视线就已随着判决生效而固化；第二，它只能碰到五个齿轮中的最后一个（终端的解释产品），碰不到前四个——立法的弹性化出厂、检察的罪名自由选择、法官的个案扩张、学术的事后合理化——这些日常的、散布的、低可见度的惩罚边界位移，不受制于任何审查，也不需要任何人事后对它们负责。

2.3 博弈论与制度分析

在法学与社会科学交叉的前沿，有一支研究将制度行为的均衡化输出用博弈论工具进行模型化分析。在刑法领域，这一路径的应用主要集中在“犯罪与惩罚的威慑均衡”上，而较少被用于分析“制度行为者自身的策略互动”。

张维迎从信息经济学和激励理论出发，分析了法律作为一种激励机制的内在逻辑，提出了法律制度的有效运作依赖于其能够在多大程度上改变行为者的成本-收益计算；法律与社会规范的研究传统则揭示了“法律秩序的维持不仅依赖于国家强制，也依赖于社会成员的内部化规范与相互监督”。这些研究为分析“解释者为什么选择扩张解释而不是宣告无罪”提供了底层工具，但它们尚未将这一工具系统性地运用于罪刑法定问题的诊断。

本文的理论框架部分，将借鉴博弈论的基本逻辑——特别是“代价不对称”这一核心概念——来模型化解释性扩张的底层驱动机制。本文提出的核心模型是：在每一个疑难案件中，解释者面对的不是“依法解释与违法解释”的二元选择（这个选择太容易——大部分解释者会选择前者），而是两种合法行为之间的代价权衡。宣告无罪是合法的，扩张解释也是合法的（因为解释者对“文义射程”的判定本身拥有裁量权）。宣告无罪的代价——舆论反弹、民意压力、治理绩效受损——完全由当下的决策者承担。扩张解释的代价——对“法律文本约束是真实的”这一集体信仰的微小松解——由整个法律共同体在未来分摊。这种代价的结构性不对称，让扩张解释成为理性决策者的占优策略：当你可以把代价分摊给别人时，你没有任何理由独自承担。

这一模型解释了为什么“呼吁法官克制”“呼吁检察官审慎”是无效的——这些呼吁试图改变决策者的偏好，但没有改变决策者面对的代价结构。只要代价结构不对称，扩张就是占优策略——无论法官或检察官的个人偏好如何。

2.4 文献缺口与本文定位

综上所述，现有文献在解释性扩张问题上留下了以下三块尚未被充分填补的空间。

第一，解释方法之争忽视了博弈结构——形式解释论与实质解释论都把问题锚定在“什么解释方法是正确的”之上，但解释性扩张不取决于解释者信奉哪种解释方法，而取决于制度博弈中的哪种选择成本更低。这是一个博弈均衡问题，不是解释哲学问题。

第二，制度约束路径集中关注终端否决而忽视前端归因——现有制度主义研究倾向于将修复锚定在“建立中央化审查机构对司法解释进行事后否决”上，这只能碰到第五个齿轮的终端产品，而碰不到前四个齿轮的日常操作。更关键的是，“事后否决”面对的是已经过终审判决固化、已形成执法惯性、已获得学界论证合法化的事态——此时否决的社会成本高到几乎不可承受，否决的概率趋近于零。

第三，缺乏一套操作化的跨齿轮程序性解决方案——现有文献提供了诊断（指出问题在哪）和规范目标（指出应该往哪走），但极少提供精确到操作步骤的工程解决方案。罪刑法定的修复如不能在制度层面被翻译为“谁、在哪个节点上、启动哪个程序、会产生什么样的约束成本”这类可被司法行政文件写定的指令，它就永远停留在学术论文的结论段里，不会进入制度。

本文的定位是：在上述三个缺口上补入实质性的增量——将罪刑法定从“解释哲学争论”和“宏观制度呼吁”这两个既有层次，推入“制度博弈的诊断”和“操作化程序工程的设计”这两个新层次。本文的核心增量不在于提出了一个新的“解释方法”判准，也不在于提出了一个新的“制度约束”理想。本文的核心增量在于：将惩罚边界的漂移机制还原为一个可被博弈论精确模型化的“代价不对称”问题，然后针对五个齿轮中每一个的启动节点，提出具体到操作程序的归因性制度设计，并给出每一层设计的可观测的成/败判据和先验失败模式与修复路径。最终将这套设计编织进一个命名研究纲领——AIPB——并给出未来五到十年内可被跨国实证检验的分阶段预测。

3 理论框架

3.1 核心概念定义

惩罚边界：本文用“惩罚边界”来指涉刑法禁止性规范的外延——即某一条文所能覆盖的、可被合法施加刑事制裁的行为的最大范围。罪刑法定原则的核心功能，就是确保这个边界是由立法机关在制定条文时的事先判断划定的，而不是由司法机关在个案审理中的事后解释重新划定的。

然而，任何法律文本的词义都有一个从核心到边缘的连续结构，精确划界在语言层面是不可能的——所有的争议案件都准确无误地落在词义的边缘地带，而不是核心地带。因此，“惩罚边界的程序归属”——即惩罚边界在某一个具体案件中是被谁、以什么身份、在什么时间点上做出的决定——就比“惩罚边界的语义位置”更能捕捉罪刑法定的实效。语义位置（这个行为在文义之内还是之外？）是解释者永远可以争论的；程序归属（这个惩罚的决定，是由立法者在事先做出的一次性判断，还是由解释者在事后做出的具体填充？）是不可争论的——它是事实，只需要记录。

惩罚边界的程序归属是本文的控制变量。对于任何一次惩罚边界的位移（即一个此前不被认为落入该条文的行为被纳入惩罚范围），本文问的不是“这个纳入在语义上是否合理”，而是“做出惩罚决定的那一个实质判断，是立法机关在条文公布时就已完成并公开标注的，还是司法机关在事后通过建构性解释做出的”。前者是罪刑法定在运作，后者是罪刑法定的仪式在运作——条文还在，法条的字面没变，但惩罚边界已经被移动了，而移动者不需要对此单独负责。

归因性：本文从博弈论的“代价分配”逻辑中借入“归因性”这个概念。在当前制度中，惩罚边界的每一次位移——通过立法弹性化出厂（立法者）、通过公诉罪名选择弹性（检察官）、通过判决中扩展解释（法官）、通过学术事后合理化（学者）、通过舆论定价反馈（公众）——都是一种“外溢性行为”：行为者享受了位移带来的当下收益（解决眼前问题、提高绩效、发表论文、满足公众期待），而位移对罪刑法定原则造成的远期代价——法律文本约束力的信用贬损、惩罚边界决定权在制度上的漂移——由整个法律共同体在未来分摊。

“归因性”就是把此前外部化的代价重新内化给行为者。归因性程序的核心设计原则是：在每一个惩罚边界位移的产出节点上，设置一个程序环节，让这次位移被记录、被公开、被归因到发起它的制度节点的具体行为者身上，并且这个归因会在下一次该节点启动同类操作之前产生即时的、不可转嫁的约束成本。

解释性扩张：本文用“解释性扩张”而不是“类推”或“目的解释”来指称惩罚边界通过解释途径向外延伸的任何操作。理由是：在当代刑事司法实践中，公然的类推极少发生（罪刑法定的子规则“禁止类推”使任何被识别为类推的判决会被撤销），绝大部分惩罚边界位移是以扩张解释——被辩护为“仍在文义可能的范围之内”——之名完成的。解释性扩张的特殊隐蔽性就在于：它永远可以声称自己“没有超出文义射程”——因为“文义射程”本身是一个可由解释者自行判定的弹性空间。因此，以“是否属于类推”作为判别惩罚边界位移的标准，在操作上是一条死胡同——解释者永远可以说“这仍在文义之内”。本文改用“程序归属”替代“语义归类”作为判别标准，回避了这个死胡同。

3.2 博弈论的代价不对称模型

解释性扩张在制度层面的持续性，可以用一个简单的博弈论模型来解释。

设一个疑难案件，公诉人面临一个行为：它在日常语言中不被自然想到为该法条所覆盖，但它与被法条覆盖的行为在某个维度上具有相似的法益侵害性。公诉人面临两个选择。

选择A：放弃以该法条起诉，以“法无明文规定”为由不予追诉（或选择与行为更匹配的轻罪罪名）。选择A的个体代价——对公诉人本人而言——包括：来自被害方的舆论压力（“为什么放着坏人不管？”）、来自上级的不作为质疑（“为什么不作为？”）、以及自身的绩效考核受损（有案不诉是不作为指标的一部分）。选择A的系统收益——对整个法律共同体而言——是：维护了“惩罚边界由条文决定”这一集体信仰的一小份信用。

选择B：以该法条（或其兜底条款）提起公诉，由法院在审判阶段完成对法条的“扩张解释”。选择B的个体收益——对公诉人本人而言——包括：完成了对明显有害行为的追诉（在舆论中是加分项）、满足了上级的作

为期待、维护了自身的绩效考核指标。选择B的系统代价——对整个法律共同体而言——是：对“惩罚边界由条文决定”这一集体信仰的一次微小松解。

在这个成本-收益结构中，选择A的全部收益归于集体，全部代价由个体承担；选择B的全部收益归于个体，全部代价归于集体。这不是一个“违法”与“合法”的选择（两者都是合法的），这是一个“自己承担代价”与“把代价转嫁给集体”之间的选择。在这个选择面前，理性个体的占优策略是B——除非这个个体有极强的个人道德信念愿意独自承受选择A的全部代价。

这个模型有四个重要的实证推论。推论一：当一个法域的立法供给与惩罚需求之间的缺口越大，选择A的社会代价就越大（因为更多“法无明文”的有害行为会被纳入犹豫地带），选择B的激励就越强，解释性扩张的频率就越高。推论二：当一个法域的司法系统的民意回应压力越强（如绩效考核中“社会效果”占权重高、媒体监督强度高），选择A的个体代价越大，选择B的激励越强。推论三：当一个法域中对解释权力的程序约束越弱，“选择B”的博弈成本（如被上级法院撤销判决、被宪法审查机关否决）越低，选择B的激励越强。推论四：在以上三个条件同时呈现的法域，解释性扩张不仅是个体理性行为，而且是制度层面的占优均衡——即使所有解释者都知道解释性扩张对罪刑法定原则在远期的伤害，他们各自的理性行为仍然会持续产出这种伤害。因为伤害是分摊的，而代价是集中的。

这个模型解释了为什么“呼吁克制”从不起作用——呼吁试图改变选择者的偏好（“请你更有原则”），但没有改变博弈结构（“你为原则付出的代价仍由你一个人承担”）。要使选择A成为理性行为，唯一的途径是将博弈规则改变为：让选择B的代价也由行为者本人承担——即，通过程序性制度设计，将每一次扩张从“代价分散于共同体的不可归因行为”转变为“代价由行为者即时承担的、被记录了归属的行为”。这就是归因性基础设施的基本逻辑。

3.3 法律文本的四圈层语义结构

在进入归因性制度设计之前，还需要一个概念工具来操作化“解释进入了哪个区域”：法律文本的词义不仅有传统的“核心-边缘”两个层次（那是哈特以来的标准区分），在实际的刑事解释操作中，词义可以更精确地划分为四个圈层。

第一圈层，核心语义圈：法条用语的日常或专业意义毫无疑问覆盖的行为类型。如刑法第二百六十三条抢劫罪中的“以暴力、胁迫或者其他方法抢劫公私财物”中的“暴力”“胁迫”——行为人在现场用刀架在被害人脖子上强行夺取财物。这一圈层不需要解释，也不产生罪刑法定的问题。

第二圈层，语义扩展圈：行为不直接落在核心语义中，但通过一次自然的语义延伸（如一次常规隐喻、一次常识类比），一个理性人在事前能合理预见该行为可能被该条文覆盖。如将“虚拟物品”解释为盗窃罪的“财物”——在2000年代早期这需要通过一次类比完成（虚拟物品与有体物在“经济价值”“可被占有”上的同构），但这一延伸在今天的数字社会中已被事前合理预见。就罪刑法定而言，第二圈层的进入是耐受的——因为可预见性这个底线未被击穿。

第三圈层，语义构造圈：行为与法条的核心语义之间需要两次以上的隐喻、类比、或体系性建构才能建立关联，且在行为发生时理性人无法合理预见该条文会覆盖该行为。这是罪刑法定的实质性失血区——在这一圈层的每一次进入，都不是对立法者已做出的判断的“发现”，而是对立法者从未做出的判断的“填补”。关键问题在于，在现有法学工具中，没有一个独立的审查机制能够分辨“这是第二圈层的自然延伸”与“这是第三圈层的建构性进入”。这个分辨被完全留给了审理法官或合议庭的自由心证——而这个合议庭正是在同一个判决中做出惩罚决定的同一主体。当裁判者同时也是裁判自己的语义跨度的唯一审查者时，“审查”在制度上是不存在的。

第四圈层，语义断裂圈：行为与法条文字之间完全不存在任何语义上的可建构关联，惩罚只能由完全独立于条文的外部理由（如“社会危害性极其严重，不容不罚”）来独立支撑。在当代成文法刑事司法中，公然进入第四圈层的操作极罕出现——因为那是明目张胆的类推，会被上诉法院撤销。但第四圈层的功能边界不应被误解为“安全的底线”：当代解释性扩张的绝大多数操作集中在第三圈层，而第三圈层的每一次进入都离第四圈层只有一层的距离。

区分四个圈层的工程意义在于：它将形式解释论与实质解释论之争从一场没有终点的哲学辩论，转化为一个可以被独立审查的程序事件。形式解释论与实质解释论的根本分界线，不在第一圈层（因为那里没有争议）或第四圈层（因为那里没有辩护），而恰恰在于第二与第三圈层之间——在于“这次进入是自然的语义延伸还是建构性的语义填补”。在现有制度中，这个问题是谁也说服不了谁的——解释者说“这是自然的延伸”，批评者说“这是伪装的填补”，两者都拿不出独立于自身解释立场的判准。四圈层语义结构的方案是：将这个“分辨”从一种实体层面的争论（“这到底是延伸还是填补？”）转变为一种程序层面的触发事件（“本次解释涉及了第三圈层的进入吗？如果是，启动独立审查”）。这个转换是三层归因性（判决阶段）的制度设计的语义学基础。

3.4 跨学科理论调用与有效性说明

本文的核心理论框架调用了两个远端学科的分析工具——博弈论（用于模型化制度行为者的策略选择）和认知语言学（用于操作化语义距离的识别）。在此有必要显式说明，为什么这两个跨学科调用在本议题上是有效的，以及它们的有效边界在哪里。

博弈论在法学分析中的有效性，建立在“制度行为者是理性的”这一基本假设之上：他们在给定的制度约束和信息条件下，会选择使自身利益最大化（或成本最小化）的行为。这个假设在“法官是否追求‘使自身成本最小化’”这个点上可争议——因为法官在理想类型中应是中立的纠纷裁决者，不为任何一方的利益所动。但本文所分析的博弈中的“成本”不是金钱收益或职务晋升——它是“当下判决被舆论反弹的强度”“起诉被上级质疑的可能性”“自身判决被上诉法院撤销的概率”这类与法官的职业声誉和制度生存相关的成本。任何一个法官——无论多么正直——在面对一个疑难案件时，都会对这些成本做出判断，这些判断不可能不对他的裁量产生影响。因此，博弈论在此处的有效边界是：它能够模型化制度行为者在面临“两种合法选择”之间的成本权衡及其均衡输出，但它不能够模型化制度行为者在个人道德信念驱动下愿意承受多大程度的个人代价去选择不占优的策略。后者属于个体心理学的范畴，不在博弈论的有效覆盖范围内。

认知语言学中的语义距离分析（原型理论、范畴化、概念隐喻理论）在法学解释中的有效性，建立在“法律文本的语言是社会共享的公共语义资源”这一假设之上。法律条文的用词不是立法者个人的私人语言——它在制定时就预设了“一个理性的普通公民或法律专业人士在阅读时能够理解的通常含义”。因此，当一次解释将条文推向一个需要两次以上隐喻或类比建构才能到达的行为类型时，这种“推”是可被语言学的语义距离分析从外部进行测量和标记的——不需要依赖对解释者内心意图的猜测。这套语言学工具在本文议题上的有效边界是：它能够提供一个独立于争论双方的第三方参照系，来标记“解释在核心语义上延伸了多远”，为“第三圈层进入”的程序触发提供客观化判据；但它不能够（也不应该试图）替代法律判断——“语义距离多远算太远”最终是一个规范性判断，不是一个语言学家的实验室测量。语言学提供的是那张测量出来的“语义距离地图”，法律共同体在这张地图上去决定“哪条等高线是不可逾越的边界”——这个分工使语言学调用不具有越界的风险。

4 五个齿轮：惩罚边界的分工瓦解机制

4.1 概览：一个合作-瓦解场域

罪刑法定原则在制度实践中不是一个单一的行为者或制度节点在守护的“圣火”。它是被五个行为群体——立法者、检察官、法官、刑法学者、公共舆论——在各自不同的制度位置上，以各自不同的操作方式，共同支撑起来的一个动态均衡。任何一个群体单方面放弃支撑，都不足以导致整个均衡的崩溃；但当又一个齿轮先后停转时，均衡就会在无声中被瓦解。

这套瓦解机制的隐蔽之处在于：这五个齿轮的每一次停转，都可以被合理解释为“在特定情境下的正当选择”。立法者把条文写得弹性化，是因为“要给司法留出应对社会变迁的空间”——这在立法技术上被认为是成熟的。检察官在兜底条款和精确罪名之间选择前者，是因为“前者的举证负担更轻、胜诉率更高”——这在检察职业伦理中非但不是错误，反而是“高效办案”的体现。法官对“文义射程”做出包容性解释，是因为“若不惩罚此行为将造成明显的处罚漏洞”——这是实质正义在个案中的本能反应。学者为已生效的司法解释提供融贯的理论论证，是因为“这是学术对制度实践的智识服务”——这是法学作为实践学科的自觉。公众

舆论对待扩张解释的判决给予正面评价（“抓得好！”），而对宣告无罪的判决（“法律有漏洞”）给予负面评价——这是因为公众不具备区分“立法惩罚决定”与“司法惩罚决定”的专业信息。

每一个选择在各自的局部里都说得通。但把这五个选择串在一起，它们在系统层面完成了一件谁也没有公开同意、但谁都在日常操作中推进的事：惩罚边界的决定性权力从立法端漂移到了司法端，而这个漂移不产生任何可以被追溯的制度责任——因为五年、十年后回头看时，你找不到任何一个“应当对此负责”的单一行为者。这就是罪刑法定最隐蔽也最致命的退化形式：不是被推翻，而是被分工消解。

4.2 立法齿轮：弹性化出厂

立法者是惩罚边界的第一齿轮。在理想类型中，这个齿轮的运作方式是：立法机关在制定一项刑事禁令时，就已经完成了对“哪些行为应当受到此项禁令的惩罚”的实质性判断，并将这个判断明确地表述在条文的文字中。在这个理想状态中，惩罚边界的决定在立法阶段就已经完成——后续的司法只是在“执行”这个已完成的决定。

然而现实中的立法操作，在上述理想状态之外，还存在着另一种常态模式：立法者在制定条文时，有意为司法解释预留弹性空间。这种预留不需要采取“显式的授权”（如“其他严重扰乱市场秩序的行为”这种典型的兜底条款），它也可以采取“隐式的留白”——使用一个在日常语言中具有较大语义弹性的规范概念

（如“情节严重”“数额较大”“后果严重”），然后在立法的过程中不给出明确的上限说明。从立法技术上看，这种“留白”有正当的功能理由——社会变迁太快，条文写得太死会导致频繁修法（修法的政治成本远高于解释）。但它的无意后果是：立法者在“出厂”时就已经把惩罚边界的画线权移交了一部分给司法端。这个移交在出厂时是不可见的——条文在那里，字面清楚，没有任何一行写着“此边界由法院在个案中确定”。但二十年后回头看，你会发现立法者从没有就“这条文的惩罚边界止于哪一类行为”做出过一个可以被外部援引的、可被事后再追溯的事先判断。

4.3 检察齿轮：罪名选择的自由裁量空间

检察官是惩罚边界的第二齿轮，也是最不可见的一个齿轮。这个齿轮的操作方式不是通过“解释法律”（那是法院的工作），而是通过“选择罪名”——在提起公诉时，检察官拥有从刑法分则中选取一个或多个罪名来指控被告的自由裁量权。

这个裁量权的关键在于兜底条款的存在。刑法分则中存在着两类“弹性罪名”——如非法经营罪（特别是其第四项“其他严重扰乱市场秩序的非法经营行为”）、寻衅滋事罪、以危险方法危害公共安全罪等——这类罪名的条文本身未详细列举其覆盖的行为类型，而是在列举了几个特定类型后，附加一个“其他……”的兜底短语。当一个行为在现行精确罪名下难以被完整覆盖、但又具有明显的社会危害性时，检察官面临一个选择：是按精确罪名的构成要件逐一检验、承认其中某个环节不成立、从而考虑更轻的罪名（或不起诉）？还是直接用兜底条款——其模糊性恰好意味着它不需通过那么严格的构成要件检验？在当前的制度激励下，后者的个体成本显著低于前者——兜底条款的举证负担更轻、胜诉率更高、政治正确性更强。而前者——承认精确罪名不完全覆盖、选择更轻罪名或不起诉——会招致来自被害方的抗议、来自上级的质疑、和来自公众舆论的负面评价。

最关键的是，这个罪名选择——惩罚边界的第一个程序入口——是完全不受外部审查的。公诉权在罪名选择阶段的自由裁量，不受任何独立机构在事前或事中的审查；它只在判决阶段受到法院的终局裁判——如果法院认为起诉罪名不当，法院可以变更罪名。但从惩罚边界漂移的角度看，判决阶段的“变更罪名”已经太晚了：法院面临着“要么宣告无罪（个体代价极高）、要么用较重的兜底条款定罪、要么用一个很不匹配的轻罪定罪（个体代价也高）”的三难困境。这种三难困境被有经验地预设于检察官的罪名选择环节中——这正是检察齿轮在惩罚边界位移中的力量所在：它在起点上选择了一条让后续齿轮（法院）很难拒绝的路径。

4.4 审判齿轮：分层博弈中的抵抗意志消解

法官是惩罚边界的第三齿轮，也是最被寄予厚望的齿轮——因为罪刑法定的传统叙事倾向于把“司法独立”和“法官忠于法律”当作原则的最后一道防线。

但当代法院系统面临着一个分层化的博弈结构，不同层级的法院在解释性扩张中面临不同的成本-收益计算。基层法院是“代价完全集中、博弈空间为零”的典型场景——基层法官面对一个疑难案件时，他的职业生存考量（判决被上诉法院撤销的概率、被舆论负面曝光的风险、被当事方投诉的可能）远大于他对“远期的法律文本信用”的维护动力。在他的决策时间窗口内，扩张解释是占优策略——它满足了对“法律应惩罚明显有害行为”的公众直觉，因而也最小化了他个人的职业风险。最高法院（及各省高级人民法院）则面对不同的博弈——它更关心制度合法性的维持、学界的评价、国际比较中本国司法产品的声誉。因此最高法院有能力产出体系化的、融贯的论述，为扩张解释提供“这依然在法律的文义与体系范围之内”的叙事。

这种分层博弈的产出是：基层法院提供了扩张解释的原始素材（个案的“不得不如此”操作），最高法院将这些原始素材淬炼为体系化的、带有权权威性的司法解释或指导性案例——而这些解释或案例在下一次基层法院遇到同类案件时，就变成“前人已经通过的桥”，不再需要重新承受博弈的压力。审判齿轮的抵抗意志就这样被分层消耗了——不是在违法的土壤中被腐蚀，而是在合法的分层体系中被“向上转移”——基层法院把疑难和责任上交给最高法院，最高法院用体系化论述来吸收疑难和责任，但在这个上移的过程中，没有任何一个层级做过“这个行为真的是立法者当初画定的惩罚边界内的吗”这种原点性的追问。

4.5 学术齿轮：从边界守夜人到事后合理化供应商

刑法学者的制度角色，在罪刑法定的理想叙事中，应该是一个“边界守夜人”——当一项刑事司法解释或判决将惩罚边界向外推移时，学者的任务是识别这次推移是否击穿了罪刑法定的底线，并且（如果是）公开指出这个问题。

但在当代刑法学术评价体系中，学者面临的激励结构正在系统性地将其角色从“事前审查者”扭转为“事后合理化提供者”。这背后是三股推力。其一，事后合理化的论文在发表市场上更具竞争力——一项已经生效的、产生广泛影响的司法解释或判决，天然构成了一个“重要学术议题”，为它提供一个融贯的体系化论证，是学术期刊偏好的“与实务对话”的选题类型。其二，反面论证——指出一项司法解释超越了立法边界——的论文，面临着远更高的论证负担：它不仅要论证“边界在哪”这个本身就极困难的问题，还可能被审稿人以“纯属学者个人意见、对实践无指导意义”为由拒稿。其三，在学术评价中，“事后合理化”类论文因其“与制度的建设性互动”而在特定评价体系中被赋予更高的实践价值权重。

这个三股推力的综合效应是：大多数刑法学者自觉或不自觉地将自己的学术资源配置在“为已经完成的解释性扩张提供论证资源”上，而不是“在解释性扩张尚在草案阶段时完成对其的文本约束审查”上。学术齿轮因此从罪刑法定的第五个守夜人，退化为第五个“合法的但不必为自己的合法化服务承担任何责任”的扩张共谋人——当一个扩张解释出现时，它会被学界论证为“体系融贯的”，这个论证会在下一次类似案件中被司法机关直接援引，扩张解释就此从“个案中是不得不如此的操作”升格为“经学界充分论证的稳定解释”。而学者不需要为这种“升格”产生的远期后果——惩罚边界的进一步漂移——承担任何可被追溯的责任。

4.6 舆论齿轮：盲人定价

公共舆论是五个齿轮中唯一一个不能被制度设计直接约束的齿轮——它不是制度行为者，而是无数分散的个体对刑事司法的情感反应集合。但它在博弈场域中扮演了一个看不见的、但刚性的“价格信号”角色。

舆论齿轮在惩罚边界位移中的运作机制是：它给两种司法行为标上相反的价格。宣告无罪（选择A）被标为负价——舆论将之解读为“法律有漏洞”或“法院不作为”。扩张解释并定罪（选择B）被标为正价——舆论将之解读为“法院打击犯罪有力”或“法律跟上了时代”。这个价格信号在公众认知中是完全自洽的——公众不具备（也不能被合理期待具备）区分“立法者事先做出的惩罚决定”与“司法者事后做出的惩罚决定”的专业能力。当他们看到一个人在交易中收受明显不合理的“交易价格”（实际是贿赂），他们无法理解为什么“不能定受贿罪”。他们的正义直觉是一元的——该罚的就要罚，条文怎么绕着弯也不能成为不罚的理由。

但问题在于，这个正义直觉的定价效应，精确地契合了前四个齿轮中每一个的博弈激励：立法者知道弹性化出厂不会被舆论惩罚（舆论看不懂立法文本的弹性），检察官知道用兜底条款起诉会被舆论奖励（舆论只知道被告人被抓了），法官知道扩张解释定罪会被舆论奖励（舆论只觉得正义被伸张了），学者知道事后合理化

论文会被实务界和舆论共同欢迎（舆论不知道这篇论文是在为解释性扩张提供事后合理化，只看到“学者说这法律解释是对的”）。

舆论齿轮不是恶意的——它只是“盲”的。但在一个每个齿轮都在根据价格信号调整自身行为的博弈场域中，一个盲的定价者比一个恶意的攻击者更危险：因为恶意可以被识别，盲不能被说服。

5 惩罚边界的多层可归因性：四层制度设计

5.1 从局部修补到跨齿轮可归因性

上述五个齿轮的分工瓦解机制揭示了一个关键的制度事实：罪刑法定的修复，不能靠向任何一个单一齿轮呼吁“回归文本约束”（因为每个齿轮面临的博弈结构决定了“背离约束”才是占优策略），也不能靠向终端齿轮施加“事后否决”（因为事后否决只能触及第五齿轮的产品，且否决的成本已经大到了使否决不可能启动）。

修复的唯一路径是：建立一套能穿透五个齿轮的程序性归因机制，使每一个齿轮在启动惩罚边界位移的产出节点上，就即时产生个体化的约束成本。这套机制不是要在五个齿轮上各安装一个独立的“刹车片”（那样的话，被按下的齿轮会将扭矩传递给相邻齿轮，使修补被烧开），而是要将同一个归因逻辑——每次惩罚边界位移都必须被记录、被归因、被即时定价——贯通地实施在四个关键的制度产出节点上。

这四个节点分别对应四层归因性：立法的语义上限说明（第一层）、公诉的罪名选择归因（第二层）、判决对第三圈层语义构造的独立程序（第三层）、学术评价的区分性激励（第四层）。以下分述每一层。

5.2 第一层归因性：立法语义上限说明

制度设计：在每一项涉及刑事惩罚边界的立法草案提交表决前，立法机关须附随一份“语义上限说明书”，以自然语言明确回答三个问题：（1）本条文的惩罚边界止于哪一类行为？（2）条文文字与止于此类行为之间的对应关系是什么？（3）条文文字在何种解读下会超出此边界？“语义上限”不是对条文含义的穷尽性定义（那是不可能的，也是不可欲的），而是对“立法者此次做出的实质性判断”的诚实标记——它告诉未来的解释者：我做过判断的是这一块区域，我没有做过判断的是这一块区域之外。

制度逻辑：将“惩罚边界在哪”的事先判断从立法者不透明的内心状态拖入一个可被外部审查的文本标记，使第一齿轮的每一次“弹性化出厂”都被记录、被归因、并承受未来被援引作为审查依据的约束成本。当一个未来案件中的解释者试图将惩罚边界推至语义上限说明书所画定的边界之外时，被追诉方可以援引该说明书作为“立法者在制定时已明确将此类行为排除在外”的证据。

成/败判据：未来五年内，至少一个法域出现了在立法文件中附带语义上限说明的首次制度实践，且该说明在后续的司法解释中被援引为审查依据。最可能的先发区域是德国——该国联邦宪法法院长期坚持“立法者必须完成实质性利益权衡”的判准，语义上限说明仅是将此判准从宪法审查环节前移至立法环节的程序化延伸。

先验失败模式与修复路径：失败模式一——语义上限说明书沦为立法理由书的另一个名称，以循环表述替代真实的语义边界画线。修复：将语义上限的审查权赋予一个独立于立法过程的技术审查机构，该机构有权以“语义上限未明确画出边界”为由要求立法草案重新附具。失败模式二——语义上限说明书自身在后续司法中被解释扩张，法院将“止于A类行为”解读为“包含A类及其结构类似行为”。修复：在语义上限说明书中加入“反扩张条款”——明确声明“本说明的语义上限禁止在解释中通过类比、目的性扩张或体系解释将惩罚边界扩展至本说明书未指明的行为类型”，违反该条款构成上诉或违宪审查的独立事由。失败模式三——语义上限说明书过于严格，导致立法频繁修法以扩张边界。修复：这不是失败模式，这是罪刑法定自身的结构要求——当一种新危害是文本真实读不出的，它就应该通过修法而非解释进行惩罚。如修法过于频繁造成立法负担，修复路径不是放宽语义上限，而是建立快速修法通道，将特定类型的刑事化提案纳入快速立法议程。

5.3 第二层归因性：公诉兜底条款选择之归因说明

制度设计：在检察官以兜底条款提起公诉的任何案件中，起诉书须附加“精确罪名不可用性说明”——逐项解释为什么刑法分则中已存的精确罪名不足以覆盖本案行为，而必须动用弹性兜底条款。这份说明不是对兜

底条款本身的解释（那是法院的工作），而是对“公诉权在罪名选择环节为什么放弃了精确罪名、选择了弹性罪名”的程序性日记。

制度逻辑：将公诉权的罪名选择弹性——这是五个齿轮中最不可见、最不受外部审查、但对惩罚边界的影响最直接的环节——首次纳入程序性归因约束，使每一次弹性选择都必须以书面形式公开回答“精确罪名真的不能用吗”这个问题。这个答案一旦被记录，就在未来的上诉、再审或司法审查中成为可供追溯的归因节点。

成/败判据：未来五到十年内，至少一个法域在检察起诉规则中明确引入此项义务，且该义务的实施与以下两个趋势的停止或逆转同时发生——（1）公诉书中兜底条款引用占比的逐年攀升停止；（2）法院在审查“精确罪名不可用性说明”后，出现了首次以“精确罪名实质可用、公诉未经充分归因说明”为由要求检察院变更起诉罪名的裁定。

先验失败模式与修复路径：失败模式一——精确罪名不可用性说明沦为格式化文本，不提供逐项比对。修复：将说明义务的操作化判准明确写入起诉规则——说明必须包含“现有精确罪名清单”“本案行为与每个精确罪名构成要件的逐项比对”“比对后确认不符合的具体项”，缺一则视为程序瑕疵，被告人可据此提起中间上诉。失败模式二——检察官通过选择性起诉规避说明义务：在精确罪名可用的情形下选择性添加兜底条款以获取量刑谈判筹码。修复：将“起诉罪名选择”纳入审前程序的司法审查范围——被告方有权在正式庭审前向预审法官申请审查起诉罪名的适当性，预审法官有权要求检方就特定罪名的选择提供归因说明。失败模式三——说明义务导致刑事诉讼效率下降。修复：这不是代价，这是成本的内化——当前体制中效率的高企恰恰是因为罪名选择的成本被完全外部化。如果制度运行初期确实出现滥用中间上诉拖延诉讼的情形，修复是设定程序性门槛（如预审法官的初步审查须在四十八小时内做出是否受理的裁定），而非放弃归因义务本身。

5.4 第三层归因性：判决对第三圈层语义构造的独立程序启动

制度设计：当法院在任何刑事判决中将一项行为纳入第三圈层（即该行为与被解释法条的核心语义之间需要二次以上的隐喻、类比或体系性建构才能建立关联）时，法院不得在判决书中自行完成这一关联的论证并在此基础上直接判罚。法院必须先启动一个独立的程序环节：（1）法院在判决草案中明确标注“本案解释涉及语义构造圈的进入”；（2）将该标注及待决的语义关联问题移交给一个独立于本案合议庭的审查组织；（3）该独立审查组织在限定时间内做出裁定——确认进入或驳回进入。

制度逻辑：将“第三圈层语义构造”从法官内心的自由心证空间拖入一个可见的、程序化的、由独立审查组织完成的裁定空间，使每一次在文本沉默处“替文本开口”的行为都必须穿过一道额外的程序窄门——这道窄门的钥匙不在审理法官手中，在独立审查组织手中。这是五个齿轮中唯一可以切断“从解释到判决”直接通道的环节。

成/败判据：未来五年内，至少一个法域在其刑事司法制度中首次出现了这样一种程序——它不是事后宪法审查对某个司法解释的偶发否决，而是将“第三圈层进入”识别为一个需要触发独立程序的事件，该程序的启动在判决生效之前、由审理法院主动发起、由独立审查组织完成裁定。最可能的先发区域是德国（该国联邦最高法院已有在特定疑难案件中启动“大合议庭”作出先例性裁定的制度，可在此基础上扩展其启动条件）或英格兰和威尔士（上诉法院在量刑指南中已有对“偏离常规”的强制审查，可将同类逻辑移植到“偏离核心语义”的审查）。

先验失败模式与修复路径：失败模式一——法院系统性地压低第三圈层识别的门槛，将大量实质进入第三圈层的解释标注为第二圈层以规避独立程序的触发。修复：将第三圈层的识别判准从法院自行判断改为双方可争执的程序事项——被告方有权在法庭辩论阶段提出“本案解释已进入第三圈层、应启动独立程序”的动议，法院必须以书面裁定回应（可上诉）。同时建立“第三圈层识别指南”——参照语言学中的语义距离分析方法，给出从核心语义到待决解释需要经过几层隐喻或类比的形式化判准，为被告方动议和上诉法院审查提供可参照的技术工具。失败模式二——独立审查组织在事实上等同于原审合议庭的“橡皮图章”。修复：独立审查组织的构成规则必须明确排除本案合议庭成员及与本案有利害关系者；更激进的方案是将第三圈层进入的审查权全部上提至上一级法院。失败模式三——独立程序启动过度频繁，导致刑事诉讼效率下降。修复：设定启动门槛——只有当待决解释涉及法定刑三年以上有期徒刑的案件时，第三圈层的独立程序才强制启动；轻罪案件

继续由审理法官自行裁量。更重要的是，即使只覆盖重罪，这一制度也将在每年数千起案件中强制执行“解释纪律的第三圈层自锁”，其产生的信号效应将倒逼检察机关在第二层归因性中更审慎地选择罪名。

5.5 第四层归因性：学术评价的区分性激励

制度设计：在刑法学学术评价体系（期刊审稿、职称评审、研究生培养）中，为“事前文本约束审查”与“事后合理化”两类学术产出赋予不同的学术价值权重。具体操作：（1）学术期刊在审稿环节区分“事前审查”类论文——即在起诉时为时间节点，反思性地检验当时法条的核心语义是否确实无法以常规解释覆盖本案行为——与“事后合理化”类论文；（2）学术职务评审中，候选人提交的代表作清单须包含至少一篇“事前审查”类论文；（3）法学院在刑法学研究生培养中，将“文本约束审查写作训练”列为必修模块——给定一个已生效的扩张性解释，学生被要求为辩方写一份“该解释在起诉时点下的文义上限无法覆盖本案行为”的法律意见书。

制度逻辑：学术齿轮是五个齿轮中最慢、但也最深的齿轮——它在短时段内为每一项扩张解释提供“学界论证”的合法化资源，在长时段内塑造一代法律人的专业直觉。区分性激励的工程目标，是加速这种直觉从“论证融贯就是好”转向“能在事发时点诚实地辨认出文本约束的边界、并以这个边界为起点来建构论证就是好”的转向，并将它从“少数独立学者的个人操守”升级为“学术评价制度的固有规则”。

成/败判据：未来五到十年内，至少两个主流法域的具有影响力的刑法学学术期刊在其评审规则中主动引入了区分性审稿标准，且至少一个法域的刑法学研究生培养方案中引入了“文本约束审查写作训练”作为正式课程模块。

先验失败模式与修复路径：失败模式一——区分性激励沦为形式化操作。修复：将“事前审查”标签的操作判准公开化——论文必须明确列出本案行为被公诉的时点、在该时点下逐项检验了所有可用的精确罪名、并给出当时法院应如何裁定的具体法律分析。这三个判准是公开的审稿单的一部分，任何读者都可拿着这些判准去核对论文是否真正完成了“事前审查”。失败模式二——区分性激励导致学术产出结构失衡。修复：区分性激励的目标不是用新的极端取代旧的极端，而是在两者之间建立一种此前不存在的学术竞争力的均势；当两类论文发表比接近一比一时，激励即可退出。失败模式三——与国际接轨的压力。修复：这恰恰是一个历史窗口——中国刑法学在全球刑法解释学中拥有最大规模的解释扩张经验数据库，如果率先将这一区分制度化，它将不是一个“与国际脱轨”的地方标准，而是一个“基于最丰富经验提出的可供其他法域借鉴”的先发标准。

6 可证伪性检验：排除关键竞争解释

6.1 最强的竞争解释

以上提出的“惩罚边界多层可归因性”框架面临的最强竞争解释是：罪刑法定的实效恶化，其实只有一个根本原因——在某些法域中，司法系统缺乏源自政治结构的独立性。在那些法官会因为“宣告无罪”而遭到实际职业报复（调岗、降级、免职）的法域，任何程序性归因机制都会被“服从上级意志”这一更强的激励所覆盖。也就是说，本文提出的四层归因性基础设施，在缺乏基本的法治保障的法域中，不过是一组无法发动的制度装饰——因为独立审查组织不可能独立，学术评价不可能不被政治引导，检察官的罪名选择归因会被“这是上面的意思”一语消解。

这是一个需要在研究设计层面被认真对待的竞争解释。如果它成立，那么本理论的有效范围就被压缩至那些已经具备一定法治基础的法域——在这些法域中归因性程序多少还有点用；而在最需要罪刑法定保障的法域中，归因性程序不过是徒增制度成本的空转。

6.2 将竞争解释转化为可控制变量的研究设计

本文对这一竞争解释的回应不是“宣告它不成立”（那需要实证数据，而当前还没有），而是设计一个能够在未来的跨国实证研究中排除它的控制方案。

(1) 样本选择：选取六到八个法域，在“法治基础设施成熟度”上分层抽样——选取三至四个在司法独立指标（可使用世界正义工程的法治指数中的“司法独立”子指标）上得分较高的法域，和相同数量的在司法独立指标上得分较低的法域。在每一层内，再根据是否具备某一归因性程序，将法域分为“有归因性程序”和“无归因性程序”两组。

(2) 因变量：惩罚边界的逐年漂移幅度——操作化为“兜底条款引用占比”和“语义收缩指数（SCI）的均值变化”两个可编码指标。

(3) 控制变量：除了司法独立指标外，还需控制每个法域的刑事立法频次（排除“立法供给缺口差异是唯一驱动”的竞争解释）、公众对刑事司法的满意度（排除“舆论压力差异是唯一驱动”）、以及法系传统（排除“在大陆法系中解释性扩张本就更高”的竞争解释）。

(4) 证伪判准与干预方向的改变：

- 如果本理论成立：在司法独立指标相同的情况下，“有归因性程序”组法域的惩罚边界漂移幅度显著低于“无归因性程序”组——也就是说，归因性程序在控制住“法治基础”这个常量后，仍具有独立的约束效应。这意味着归因性程序即使在司法独立度较低的法域中也部分有效——因为它不依赖于法官的政治勇气，而依赖于程序节点（罪名选择归因、第三圈层独立审查）的结构性设置。

- 如果竞争解释成立（本理论被证伪）：“有归因性程序”与“无归因性程序”两组之间的惩罚边界漂移幅度，在统一控制司法独立指标后，没有显著差异。也就是说，归因性程序在司法独立度较低的法域中确实是一组空转的装饰品，它的任何效果都被“政治服从”这个变量吸收了。

- 干预方向的分叉：如果本理论被证伪，则对罪刑法定的修复努力必须从“程序性归因设计”推至“更根本的法治基础设施保障”——包括但不限于法官终身制、司法预算独立、外部审查机构的宪法保障等政治结构的变革。这些变革超出了刑法学内部的制度设计能力。反之，如果本理论成立，则刑法学共同体拥有了一个不依赖根本政治变革而仅依赖程序创新的修复工具——归因性基础设施的有效性在“给定现有司法独立水平”的条件下可以被识别、被测量、被复制。这是一个重要的实践分歧：它决定了刑法学在罪刑法定修复上应该把有限的制度资源投向哪里。

6.3 什么结果会推翻本理论，并指向不同的干预方向

除上述竞争解释外，本理论面临的其他可证伪条件包括以下三个。

第一，如果一个法域实施了第三层归因性程序（第三圈层独立审查），但在实施三年后，其SCI指数的均值与实施前无显著差异，则本理论在“第三圈层独立程序可有效约束解释性扩张”这一核心假设上被削弱甚至推翻。导致这一结果的深层原因需要进一步检验——最可能的原因是第二圈层和第三圈层的识别在操作中被系统性压低了（即上述失败模式一），这意味着问题的症结不在归因性程序的设计逻辑，而在独立审查组织的“独立意志”本身——而这又把问题推回了“司法独立”的总变量。

第二，如果一个法域实施了第二层归因性程序（公诉归因性说明），但其兜底条款引用占比在实施三年后继续攀升，则本理论在“罪名选择归因可约束检察齿轮”这一假设上被证伪。导致这一结果的深层原因可能是：检察官在“精确罪名不可用性说明”中诚实地完成了逐项比对，但比对的结果仍然指向“精确罪名确实不够用”——这意味着问题的症结不在检察裁量权的滥用，而在立法的惩罚边界本身在网络时代面临根本性的“词义崩塌”。如果是后者，归因性程序的策略目标就从“约束解释”转向“加速修法”——但罪刑法定原则本身将被推向一个更深的两难。

第三，如果在未来五到十年内，全球主要法域均未出现任何归因性程序的制度化实验，则本理论的“概念合法化”和“必然发生”两个时序假设被证伪——这意味着本理论严重低估了制度的惯性，或严重高估了法学

共同体（包括本文在内）从“诊断问题”到“推动制度化实验”的行动能力。这一证伪不推翻本理论在诊断上的有效性，但推翻本理论在实践上（“这个方案是可被实际实施的”）的隐含假设。

7 结论

7.1 核心发现

本文通过六个论证步骤，推进了罪刑法定原则在当代面临的退化机制的诊断与修复方案。

第一，罪刑法定当前面临的根本威胁不是条文被公然违反，而是惩罚边界的决定权在日常操作中被五个制度齿轮分工消解——立法者弹性化出厂、检察官利用罪名选择弹性、法官在分层博弈中丧失抵抗意志、学者从边界守夜人退化为事后合理化供应商、舆论充当盲人定价者。这个过程不需要任何违宪行为，它只需要每个齿轮在自己的岗位上做自己认为正确的分内事。

第二，这种退化机制的底层驱动力可以用博弈论的代价不对称模型解释——宣告无罪的全部收益归于集体（维护了法律文本信用的远期价值），全部代价集中由决策者个体承担（舆论反弹、绩效受损）；扩张解释的全部收益归于个体，全部代价由整个法律共同体在未来分摊。这个结构让扩张解释成为任何理性决策者的占优策略。

第三，修复这一退化的唯一路径不是呼吁任何人“变得更高尚”，而是改变博弈规则——将每一次惩罚边界位移的代价，从不可归因的“由共同体分摊”状态，拖入“由发起位移的制度节点当下即承担”的可归因状态。这个思路被操作化为四层程序性制度：立法的语义上限说明、公诉的罪名选择归因说明、判决对第三圈层语义构造的独立程序启动、学术评价的区分性激励。

第四，这四层归因性不是五组独立的局部刹车片，而是同一个归因逻辑——每一次惩罚边界位移都必须被记录、被公开归因到发起它的制度节点、并在该节点下一次启动同类操作前产生即时的约束成本——在四个关键产出节点上的贯通实施。

第五，四层归因性连同本文累积的诊断工具（语义收缩指数、四圈层语义结构、五齿轮停转序列），可以被编织进一个命名的研究纲领——惩罚边界归因性基础设施研究纲领（AIPB），含五项主干问题群、四条方法论传统、六个独立的跨国实证研究项目，以及一个分三阶段、可证伪的预测框架。

第六，本文明确设定了可证伪条件——控制“法治基础设施成熟度”变量后，归因性程序是否仍具有独立的约束效应；若归因性程序的引入未能在时间序列上产生显著的惩罚边界漂移减缓，则本理论的实践有效性被推翻，干预方向需推至更根本的法治基础设施保障。

7.2 理论贡献

本文对法学文献的主要贡献在于以下三点。

第一，将罪刑法定问题的分析框架从“解释方法争论”推入“制度博弈诊断”。形式解释论与实质解释论之争在过去二十年里产出了大量有价值的成果，但它在结构上无法回答“为什么明知扩张有害还是选择扩张”这个问题——因为这个问题不是解释哲学问题，而是博弈均衡问题。本文首次将代价不对称的博弈论分析系统性地移植到罪刑法定议题上，并由此推导出归因性制度设计的逻辑必然性。

第二，将罪刑法定的修复路径从“呼吁”和“事后否决”推入“前端程序性归因”。既有文献对罪刑法定的修复建议，要么是呼吁解释者更克制（这无法改变博弈均衡），要么是建议建立更强大的宪法审查机构（这只能碰到第五个齿轮的终端产品）。本文首次提出了可以穿透五个齿轮的、在每一个齿轮的产出节点上即时产生约束成本的一套连贯的程序性制度设计，并给出了每一层设计的操作步骤和失败模式与修复路径。

第三，将一个已经有数百年历史的原则性问题，推入了可以被跨国实证检验的阶段。AIPB研究纲领的核心贡献，不在于它提出了什么答案，而在于它把罪刑法定的实效从一个“只能被感受、不能被测量”的模糊状态，转化为了可以被编码、被统计、被跨国比对的实证问题——惩罚边界在哪里漂移、以什么速度漂移、归因性程序对漂移速度有没有抑制效应——这些问题现在有了可操作的变量定义和编码方案。它不需要等待任何人“承认”罪刑法定在退化——它只需要把数据跑出来，让数据说话。

7.3 实践与政策含义

对立法机关而言：（1）在制定新刑事禁令时，考虑附加“语义上限说明书”作为草案的组成部分；（2）在修法程序中建立快速通道，以降低立法面对新危害类型时的修法门槛，从而减少将修法压力转移给司法解释的激励。

对检察机关而言：考虑在内部起诉规则中引入“兜底条款归因说明”义务——在司法行政层面先行试点，不待立法强制；这同时也能为检察机关在面对舆论压力时提供一个制度性的回应工具（“我们已经逐项检验了精确罪名，确实不能用”）。

对法院系统而言：考虑在审判管理创新中试点“第三圈层语义审查”——先在重罪案件中试行，作为法院内部的审判监督机制（而非外部审查），积累运行经验后再评估是否需制度化。

对学术共同体而言：刑法学核心期刊的编辑部可以主动发起对“区分性审稿标准”的讨论——这不需要任何官方授权，只需要编辑部自身的学术判断；领先的编辑部可以以试点方式在特定专题（如“罪刑法定与解释边界”专题）中试行区分性审稿，积累数据后再评估。

7.4 研究局限

本文的主要局限体现在三个层面。

第一，实证支持的缺位。本文提出的博弈论代价不对称模型、五齿轮分工瓦解序列、以及四层归因性设计，均属于理论建构和制度设想，尚未经任何法域的实际制度实践验证。文中给出的成/败判据和预测框架，是未来检验的工具，而不是已有实证的结果。

第二，归因性程序的“独立审查”终极假设。四层归因性中的每一层在提出修复路径时，都默认了一个“独立程序发起机制”的存在——如独立的语义审查专员（第一层）、独立的预审法官审查（第二层）、独立于原审合议庭的审查组织（第三层）、学术期刊的编辑部独立审稿权（第四层）。在那些政治结构中不存在独立程序发起机制的法域，归因性基础设施的建设可能必须从更根本的层面——如司法独立的结构性保障或行政审查机构的外部化——启动。本文未对这一“终极假设”展开充分讨论。这是AIPB作为研究纲领需要在第一个十年的实证研究中回答的核心问题：归因性基础设施的有效性，是否在根本上取决于法治基础设施的成熟度？

第三，跨部门法迁移的未完成论证。本文的归因性框架是从刑法领域推导出来的，并将其核心概念（“惩罚边界”“归因性”“第三圈层语义构造”）锚定在刑法的特有结构上。本文虽在AIPB的项目清单中列入了跨部门法迁移的研究项目，但并未在本文中完成从刑法到行政法、税法、平台治理的完整迁移论证。行政法的“程序正当”是否在结构上等同于刑法的“罪刑法定”？平台治理中的“社区规范”是否面临与刑法条文完全同构的“从宣传性标准到执行性标准的漂移”？这些问题需要独立的专题研究来回答。

7.5 未来研究方向

基于AIPB研究纲领的核心问题群，以下五个研究项目可在未来五到十年内被分别启动。

项目一：《立法语义上限说明书制度的跨国比较与制度实验提案》。核心问题：哪些政治结构和立法程序条件支持或阻碍语义上限说明书的引入？采用反向事实模拟法——在五到六个法域选取争议刑法条文，由独立法学专家组为其补写“语义上限说明书”，并追踪该说明书对后续司法援引该条文的影响。

项目二：《公诉罪名选择权程序约束的跨国实证——比较法视角》。核心问题：不同法域中检方在兜底条款与精确罪名之间的选择频率分布如何？程序约束的缺失是否与兜底条款的超比例引用相关？选取日本（已有“公诉权滥用”判例积累）、意大利（有“法定法官原则”传统）、中国（兜底条款使用频率较高的成文法体系）三个法域进行深描比较，并逐年编码每个法域中公诉罪名选择的类型分布。

项目三：《判决中“第三圈层”进入的识别与独立程序启动——一项模拟审判实验》。核心问题：第三圈层语义构造的识别在多大程度上可以被语言学的语义距离分析工具化？独立审查组织的制度设计（上提至上一级法院、同级异体审查、新设专门机构）在实证上哪一种更可能同时满足“审查独立性”与“诉讼效率”的双

重要求？采用模拟审判实验法——同一组疑难案例分两组审判（一组启动独立程序、另一组走常规程序），比较两组判决在惩罚边界稳定性上的差异。

项目四：《刑法学术评价的归因性转向——区分“事前文本约束审查”与“事后合理化”的学术影响力效应测量》。核心问题：在核心期刊中引入区分性激励后，事前审查类论文的占比是否显著上升？事前审查类论文与事后合理化类论文在被后续司法判决援引的频率上是否有显著差异？采用自然实验法——选取在某一时间点主动改革审稿标准的核心期刊为实验组，未改革的同级别期刊为对照组，追踪五到十年内两组期刊的论文在司法援引中的表现。

项目五：《从刑法的归因性到行政法、税法、平台治理的归因性——跨部门法的“惩罚边界”概念转换与制度移植研究》。核心问题：“惩罚边界”在非刑法部门法中如何重新定义？刑法的四层归因性设计在向其他部门法迁移时，哪些层可直接移植、哪些层需要根本重构？选取行政法（行政处罚中的“过罚相当”原则的可审查性）、税法（税收规避的“实质重于形式”规则是否隐含了同样类型的惩罚边界司法漂移）、平台治理（搜索引擎与社交媒体平台的“社区规范”执法是否面临着与刑法完全同构的边界漂移问题）三个领域进行聚焦分析，提取跨部门法迁移的通用判准与特定化改造需求。

这五个项目在方法上共享博弈论制度分析加语义编码加比较法实证的交叉工具链，在经验路径上覆盖立法、检察、审判、学术评价、跨部门法五个不同的制度场域，可被单独启动，也可被组合进同一个比较法研究网络，形成学派级的积累效应。

罪刑法定的未来，不取决于任何一篇论文的论证是否被接受，而取决于这些研究项目——或同类项目——是否有人在未来十年内去做了。本文的核心判断，将会在做了这些实证之后被证实、被修正或被推翻。这其中的任何一种结果，都是学术的正常命运。唯一不在选项里的，是继续用两百年不变的概念工具，守着两百年不变的分析框架，等待一个不同于前两百年的结果。

参考文献

1. 张明楷. (2010). 《刑法分则的解释原理》. 北京: 中国人民大学出版社.
2. 张维迎. (2002). 《法律制度的信誉基础》. 《经济研究》, (1), 3-13.
3. 陈兴良. (2006). 形式解释论的再宣示. 《中国法学》, (2006年第3期).
4. 杨仁寿. (2019). 《法学方法论》. 台北: 三民书局.
5. 张明楷. (2012). 实质解释论的再提倡. 《中国法学》, (2010年第4期).
6. 陈兴良. (2012). 形式解释论与实质解释论: 事实与理念之展开. 《法制与社会发展》, (2011年第2期).
7. 劳东燕. (2015). 《风险社会中的刑法: 社会转型与刑法理论的变迁》. 北京: 北京大学出版社.

说明：本文的理论框架与案例素材源自本领域内持续性的学术讨论，引用文献涵盖了该议题的主要研究脉络。研究局限已在结论部分完整陈述。