学员专栏 · 黄倩盈

代理如何成为暴力：高利害评价对不可代理之物的结构性排除

为什么半个世纪以来所有“找更精确代理品”的评价改革都失败了，而人们仍认为出路是找更精确的代理品？本文说：问题不在代理不精确，而在代理这个动作本身——把不可代理的成长过程强制转译为可比较证据，这一动作即是暴力。

作者 黄倩盈 · SDE 学员 · 约 2.6 万字 · 发表于2026年7月22日

这篇文章的思想创新

“代理的暴力”：在高利害分配条件下，将不可代理的内在认知生成过程强制转译为可比较的外部证据，这一动作本身即构成结构性侵害——不管代理品是分数、等级还是叙事档案。它通过“锁—切—叙事”的互锁增益运转：利害压力倒逼评分精确化，精确代理为利害计较提供更锐利的着力点，“可比较必公正”的叙事又不断为这一循环赋予道德正当性。。

导读 · 300字

这篇文章的价值与意义 —— 它能解决你的什么问题

理论问题：它撤销了从坎贝尔定律到布迪厄象征暴力、从鲍尔审计社会到比斯塔可测量性批判所共享的预设——默认成长可被代理、问题只在代理不精确／代理人被腐化／代理品被当成目的。本轮再与古德哈特定律、斯科特可读性正面切割：古德哈特的暴力在代理品被博弈、斯科特的暴力在简化丢失，而代理的暴力主张——强制转译动作本身即侵害，哪怕代理品完美、不可博弈、信息无损。

现实问题：它把改革的方向从“找更精确的代理品”整个翻过来：既然伤害在代理关系本身，那么在高利害条件下无论把代理品做得多精确，被排除的东西都不会回来——它给出的不是又一套代理方案，而是对“可比较必公正”这一叙事合法性的釜底抽薪。

自我精神问题：它把评价里那种说不出的被冒犯感命名了——当你最在乎、最不可代理的那部分成长被折算成一个可比较的分数或档案时，被侵害的不是精度，是“它拒绝被提前定价”这件事本身。

摘 要

本文追问一个被教育评价批判史系统绕开的问题：为什么半个世纪以来所有试图替代代理品的改革都失败了，而人们仍认为“找更精确的代理品”是出路？通过对中国高考评分细则演进史与课堂微观现场的发生学追踪，本文揭示：既有批判——从坎贝尔定律到布迪厄的象征暴力、从鲍尔的审计社会到比斯塔的可测量性批判——共享一个未经审查的预设，即默认成长是可以被代理的，问题只出在代理不精确、代理人被腐化、或代理品被当作了目的本身。本文撤销这一预设，论证：在高利害分配条件下，将不可代理的内在认知生成过程强制转译为可比较的外部证据，这一动作本身即构成结构性侵害——不管代理品是分数、等级还是叙事档案。此种侵害的运转通过“锁—切—叙事”的互锁增益结构实现：利害压力倒逼评分精确化，精确代理为利害计较提供更锐利的着力点，而“可比较必公正”的叙事合法性则不断为这一循环赋予道德正当性。本文进而论证，“代理的暴力”这一命名不能被既有的异化理论、象征暴力、指标暴政、默会知识传统或物化批判所1:1替换——它切开的区分面是批判代理关系本身的暴力性，而非某一种代理品的缺陷。最后，论文正面回应“无可比较证据如何公正分配”的反驳，给出“代理的暴力”的证伪条件，并指明其理论边界。

关键词：代理的暴力；不可代理；互锁增益；高利害评价；发生学

一、引论：当换了所有的尺子，侵害仍在发生

过去半个世纪，教育评价批判积累了一套成熟的语汇。坎贝尔定律警告，当一个指标被用于社会决策时，它将被扭曲并腐化其原本所欲测量的过程。Muller的“指标暴政”展示了量化指标如何从衡量工具异化为目标本身。福柯将考试分析为规训权力通过区分—分类—等级化来塑造主体的经典装置。布迪厄揭示了教育系统如何通过将特定阶层的文化资本伪装成普遍标准来实施象征暴力。鲍尔追踪了“审计爆炸”如何让可比较的指标殖民了专业判断。比斯塔则明确区分了教育中“可测量”与“不可测量”的维度，指出对测量的过度强调正在侵蚀教育的主体化功能。在中国语境中，“唯分数论”“应试教育”“评价窄化学习”等批评已持续至少三十年。

这些批判各自不可替代。但它们共同面临一个令人不安的事实：半个世纪的批判，换来的是评价系统的进一步精细化而非松动。中国高考从最初的教师整体判分，演进到分步采分、采分点逐项赋分、标准答案逐字锁定的精确评分体系；综合素质评价从理念到试点，最终在落地过程中催生了“社会实践材料润色”“艺术素养陈述优化”等收费服务；档案袋评价、成长叙事、过程性评价这些被寄予厚望的“软尺子”，每换一个，市面上就出现一套新的应试培训。换了所有的尺子，现象只是换了面孔继续运转。

这迫使本文追问一个被系统性地绕开的问题：为什么所有试图替换代理品的改革最终都走向同样的结局，而人们仍然认为“找更精确、更软、更过程取向的代理品”是出路？为什么每一项新评价工具在未被异化之前都曾被歌颂为“终于可以把真实成长纳入评价了”，而在被异化之后又被指责为“还是没能逃脱应试的宿命”——这场循环为什么从未停过？

本文的回答是：因为既有批判从未质疑过“代理本身”。坎贝尔、Muller、布迪厄、鲍尔、比斯塔以及半个世纪以来的评价改良运动，都在批判代理过程的各种问题——代理不精确、代理人被利益捕获、代理品被等同于目的本身——但没有任何一方质疑过：在高利害条件下，把内在的、不可比较的、不可被第三方验证的认知生成过程强制转译为可比较的外部证据，这个动作本身，是否就已经是一种结构性侵害？

本文将论证：是的，这个动作本身就是侵害。不是代理品“变质”之后才开始侵害，也不是代理过程“被污染”之后才出现问题——而是，代理在结构原点上就已经在系统性地排除那些无法进入代理管道的生成形态。本文称这一结构为“代理的暴力”：当一个人不可代理的内在发生——笨拙的摸索、漫长的自我修正、在不确定中推进的能力——被强制置于“必须产出可比较证据以争取分配机会”的位置上时，这种强制转译本身，就已经在结构性地排斥那些无法被转译的生成形态，不管代理品换得多么“软”、多么“多元”、多么“过程取向”。

这个判断与既有批判的根本差异在于：它不是在批判“某一种特定的代理品不够好”，而是在追问“代理这个关系本身是否就是一种不断再生产排除的结构”。

为了把这个判断的边界刻画清晰，有必要在论证展开之前，先区分几个极容易被混淆、但实际上与“代理的暴力”处于不同层面的批判传统。

首先是它与异化命题的区别。马克思传统的“异化”概念揭示了一种经典的侵害结构：人所创造的东西反过来成为人的枷锁。在教育语境中，“学生创造了分数，却被分数奴役”听起来像是对代理暴力的完美概括。但二者之间有一个不可化约的区分面：异化预设了一个“异化前的本真状态”——在马克思那里是“自由自觉的劳动”，在教育批判的异化叙事里往往隐含着“评价介入之前的纯真学习”。而异化的政治诉求也因此指向“消除异化、回归本真”。代理的暴力拒绝这个预设。它不假定存在过一个没有代理的、纯真的教育黄金时代，也不把出路定位为“回到那个时代”。它要追问的是另一件事：只要高利害分配要求产出可比较的证据，暴力的结构就会启动——而这不是因为某个时代背离了本真，而是因为“可比较的证据”这个格式本身就对某些生成形态构成了结构性的排除门槛。因此，代理的暴力不是异化理论的教育版——它取消了对“回归本真”的承诺，转而在承认暴力的结构性这一前提下，寻找局部的抵抗空间。

其次是它与布迪厄象征暴力的区别。布迪厄的象征暴力揭示了权力如何通过将特定阶层的分类体系强加于行动者、并使其被误认为“自然的”或“公正的”，来再生产社会不平等。本文在后续论证中将会与布迪厄正面交锋，但这里需要先标出一个差异：布迪厄的暴力定位在“分类体系的社会偏向性”上——伤害来自标准本身是偏的。代理的暴力则将其定位更前一步：即使分类体系是完美的，即使评价标准真的“公正地”代表了某种普遍的教育目标，只要代理动作本身启动了——把不可比较的内在发生强制转译为可比较的证据——结构性的排除就已经发生。换句话说，布迪厄追问的是“谁的标准被伪装成了普遍标准”；代理的暴力追问的是“即使标准是普遍且公正的，把不可代理之物强行推上比较的天平，这件事本身是否合理”。二者的差异在于：象征暴力可以在理论上通过“让标准真正公平”来缓解（虽然布迪厄本人对此持悲观态度），而代理的暴力即使面对最公平的标准，也仍然构成结构性排除——因为排除不在标准的偏向性里，而在代理的格式强制性里。

第三，它与鲍尔“审计爆炸”批判的区别。鲍尔令人信服地展示了审计社会如何通过制造可比较的绩效指标来殖民那些原本依赖专业判断的领域。代理的暴力追问的是一个鲍尔没有触及的问题：那些被指标殖民的“专业判断”本身——比如教师对学生理解水平的整体判断——难道不是另一种代理吗？鲍尔的叙事隐含了一个前提：在审计爆炸之前，专业判断是自由的、不受可比较指标约束的。代理的暴力要指出的是：教师的“整体判分”——那些在八十年代早期被允许的“酌情给分”——虽然在技术水平上不同于后来的精确采分点，但在结构上，它同样是代理：同样是拿一个分数来代表一个学生的理解水平。审计爆炸让代理更精确、更不可逃避，但代理本身在审计爆炸之前就已经在那里了。所以症结不在“审计”的过度扩张，而在代理关系本身——在高利害条件下，专业判断不是审计的受害者，而是代理暴力链条上的第一个节点。

最后，它与比斯塔“可测量性批判”的区别。比斯塔区分了教育的资格化、社会化和主体化三个维度，并警告对前两个维度的过度强调正在挤压主体化的空间。这一洞见与本文的核心关切高度接近。但比斯塔的着力点是“平衡”——他呼吁评价系统更多地关注那些不可测量的维度。代理的暴力则追问：在高利害分配的结构中，这种“平衡呼吁”为什么总是失败？因为高利害条件天然地会使所有能量向可代理方向倾斜——不是教育者不懂得主体化维度的价值，而是他们承受不起“学生在主体化过程中暂时拿不出高分”的后果。因此，代理的暴力不是在呼吁“更平衡的评价”，而是在揭示“平衡呼吁”为什么在高利害条件下会反复失败的动力学原因。

综合以上，代理的暴力不能被任何一种既有概念1:1替换。它不是异化（取消了回归本真的承诺），不是象征暴力（批判的不是标准偏向性而是格式强制性），不是审计爆炸（追溯到比审计更深的代理关系），也不是可测量性批判（追问的是平衡呼吁为何必然失败）。它切开的

区分面是：在高利害条件下，把不可代理的内在发生强制转译为可比较的外部证据——不依赖于代理品是否精确、代理人是否被腐化、标准是否偏向某一阶层——这一动作本身就已经在系统性地排除那些无法被格式化的生成形态。接下来的任务，是让这个判断从具体的制度和现场中自己长出来，而不是作为一套预制的坐标被贴在材料上。

二、一根链条的两端：考试大纲里的“锁”与补习班里的“切”

要理解代理暴力如何运转，不能从“定义”入手，而必须从一条完整的因果链入手。这条链的上游在考试院的命题细则里，下游在补习机构的“踩分”课堂上，中游是教师的专业判断被逐步殖民的日常实践。它们互相供能、互相辩护，共同构成了一台“谁也不能先停”的自我强化引擎。本节的任务，就是把这根链条从上游到下游完整地拉一遍。

1. 上游：利害压力如何倒逼评分精确化

上世纪八十年代早期，中国高考的评分保留着相当的弹性。1985年高考物理卷的阅卷细则可以为证。对于一道证明题，评分标准列出的不是“采分点列表”，而是一段叙述性的判分指引——“能正确分析物理过程，列出基本方程，推理过程正确，结论正确，给满分；基本方程列对但计算有误，酌情扣1-2分；推理方向对但关键步骤缺失，给一半分”。这种指引给了阅卷教师相当大的裁量空间：一位阅卷教师面对某份答卷时，他要做出的不是一个“勾选采分点”的动作，而是一种整体性的判断——大方向对不对？逻辑对不对？表述清不清楚？他把这几条合在一起，给出一个在某个区间内的分数。不同阅卷教师之间会有偏差——同份卷子甲老师给8分乙老师给6分——在当时，这种偏差尚可容忍，因为当时的利害区分还相对粗糙：“考上”和“考不上”之间隔着一道分数线，分数线上下的具体分差没有那么多人盯着。

变化从九十年代末开始加速。这里有一个容易被错过的悖论：高等教育的规模扩张本应稀释选拔压力，但毕业生就业市场的分层反而把这种压力推向了更精确的区间。1999年高校扩招之后，大学入学机会整体增加，但与此同时，重点大学与普通院校之间的毕业生起薪差距开始拉大。据教育部所属研究机构1999年至2005年的全国高校毕业生就业调查，985院校毕业生的平均起薪在此期间与普通本科院校的差距持续扩大。更致命的是，这个差距不是仅在“大学”与“非大学”之间出现，而是在大学内部不同层级之间精细分化：一个985院校的学生哪怕学的是就业面最窄的专业，其“学校出身”仍能在很多用人单位的初筛中获得优先；而一个普通一本的学生即使专业更“实用”，也可能在简历关就被过滤掉。“多考一分超过一操场人”在这个背景下成为社会共识：一分之差，就可能意味着完全不同的人生命运。

在这种压力下，评分标准的“粗”就不再只是一个技术瑕疵——它变成了一种社会不公的靶子。进入新世纪，多省相继曝出高考语文卷或文综卷的评分争议：多名考生在查分后发现，自己对某道题的作答与标准答案“大意相同但表述不同”，被不同程度扣分；更有考生指出，同一道题在不同阅卷点的评分偏差达到3分之多。舆论的集中追问是同一个：如果我的孩子因为阅卷老师的主观差异丢掉了几分，从而掉出了本该属于他的位置，这个损失谁来承担？

评价机构的回应是评分细则的持续收紧。具体机制为：每一次争议暴露了评分细则中某个“主观空间”，下一次命题之后，这个空间就会被拆解成更细的格栅。到2000年代末，多数省份的高考评分标准已经从“叙述性指引”演进为“逐词核对式清单”：一道6分的简答题不再有一个“酌情给0-2分”的区间，而是被拆成三个2分的采分点，每个采分点对应标准答案中一句固定的表述；学生写到了这个句子就给分，没写到就不得分，哪怕他写了十句与此相关但不完全一致的表述。到这一步，一个根本性的置换已经完成：“这个学生对一个学科问题的实际理解水平”——那个内在的、不可直接比较的状态——在评价系统内部被替换为“该生作答与标准答案采分点列表的吻合程度”。前者是一个生成过程，后者是一个可被精确比较的产品。这个置换一旦完成，“提高评价质量”就从“更准确地了解学生学到了什么”被偷偷改写为“生产更精确的代理品”。

2. 下游：精准代理如何被利用，又为何倒逼更精准

补习市场在这个置换中扮演了一个被严重低估的角色。大约从2000年代中期开始，中国各大城市的补习机构中出现了一类新的课程形态：不是教学生理解数学，而是专门训练学生“踩得分点”。这类课程直接以高考评分细则为教材——机构教师逐题拆解历年真题的标准答案，告诉学生“写到这个词就得分，没写到就不得分”，训练学生放弃任何超出采分点之外的自主展开，把全部精力押在“精确吻合标准答案表述”上。

南方某省会城市一位在教培机构任教多年的退休教师，在他自费印行的教学回忆录中描述了这种教学的逻辑：“我们逐题研究历年高考的标准答案，总结出每一种题型对应的固定话术。学生对斜率的概念可能完全不懂，但只要他记住了‘设一次函数y=kx+b，代入已知点坐标得方程组’这段开场白，就有分。到后来，我们甚至编出了一套‘失分预警指南’：哪些动词在阅卷细则里是采分关键词（比如‘令’‘故’‘即’‘联立解得’），哪些表述方式会被判无效（用口语替代数学符号、跳步骤、缺少‘由题意得’的帽子）。学生背的不是数学，是阅卷系统。”

这种教学形态的可怕之处不在于它在“应试”——它比一般意义上的应试走得更远。它不是在训练学生“掌握知识以应对考试”，而是在训

练“把自己塑造成评分细则所能识别的那个可比较物”。学生在这个过程中不再是一个学习的主体，而是一个自我代理的代理品——他学会的不是数学，而是如何在评价系统的眼中“看起来像掌握了数学”。

在此回看那个“一操场人”的利害压力对学生形成的深层冲击：在那样的白热化竞争中，学生每一分钟都在被迫掂量“做什么最划算”。投入一小时去真正理解斜率的本质，可能最终在新情境中多拿一道难题的分；但用同一小时去背诵三个题型的采分点话术，能保证在标准题型上少丢六分。后者确定性更高、见效更快、反馈更即时。而且，补习机构持续不断地向学生和家长传递一条潜台词——所谓理解，不过是熟练到“能在规定时间内写出所有采分点”罢了。

更值得追问的是：外部的“踩分研究”反过来也在倒逼考试系统进一步收紧评分标准。培训机构在拆解题目的过程中不断暴露出评分细则的“灰色地带”——比如某个题目的采分点措辞不够确切，或者换一种句式同样正确却被原细则判错。这些灰色地带一旦在考后被特定的争议案例引爆，阅卷部门在下一次命题与阅卷中就会进一步压缩解释空间、增加标准答案的锁定程度：“只要多解释一句就更精确？那就干脆把标准答案写死。”外部补习机构因此变成了评价精确化引擎的间接推进器：它们每一次“破解”都是在往评分系统身上扎更细的刺，而被扎的评价系统回应的是长出更坚硬的壳——更硬的标准答案、更严格的采分点、更少的阅卷自主性。更重要的是，长壳的过程对补习机构是有利的——每一次“壳更新”都意味着去年的“踩分秘籍”部分失效，也就为它们提供了新一轮授课素材和“最新考情分析”的卖点。代理的精确化不是孤立发生的，它是在学校、补习机构、考生家庭的合力之中被反复加码、再生产出来的。

3. 中游：教师如何从判断者退化为代理人

如果把眼光只放在考试院和补习班两端，容易产生一种误解：代理的暴力只发生在“制度制定者”和“应试掠夺者”那里，而学校和教师是被迫卷入的无辜者。但这个图景缺了最关键的中间层——教师的专业判断是如何在日复一日的评分压力下，从“酌情给分”退化为“采分点核对”的。

在世纪之交评卷技术尚未完全电子化时，高考质检组的核心任务是抽查同一份卷子不同评卷教师之间的给分偏差。抽检警报一旦被触发，阅卷组长的标准动作不是去追问“哪个教师对学生的整体理解把握更到位”，而是去比较“谁的给分更贴近统一采分点”。这个现场逻辑极其单纯：精准吻合细则 = 少惹麻烦，自由裁量 = 万一被抽查到偏差解释不清。阅卷教师因此被训练成了一种特定的判断者——不是判断“这个学生理解得怎样”，而是判断“这份答卷距离采分点列表还有多少个空没填上”。

这个逻辑迅速从考后阅卷向前渗透到了日常教学。一位在山东一所重点高中执教二十余年的语文教师回忆说，九十年代她给学生作文打分时，“看的是整体——思路好不好、有没有自己的想法、语言能不能撑住这些想法”；但到了2005年前后，阅卷组在省质检培训中反复强调的是“扣题关键词”“结构完整度”“每段的格式要求”——“你判作文变成了一件事：把考卷往评分细则的格子里填”。她逐渐发现，自己在课堂上讲作文评改时，也开始不自觉地使用阅卷系统的话术，“因为你没办法——学生问你‘老师我这个作文能拿多少分’，你如果说‘你写得有想法但格式不太合规’，他就会觉得你在说空话。你必须告诉他，这篇作文在哪个采分点上可能丢分，他才会觉得你是负责任的。但你一说采分点，你就在教他‘格式’而不是‘写作’。”

这不是说这位教师丧失了教育信念。而是在高度利害相关的评价架构中，“帮学生拿到那几分”的诉求以一种不容拒绝的方式，挤占了“让学生学会真正写作”的空间。而且这种挤占在教师群体内部是有传染性的：一旦某几位教师率先开始在讲评课上嵌入精准的采分点攻略，并且他们的班级在考后拿到了更高的平均分，其他教师——无论理念如何——都会被家长、学生甚至学校管理层推到同一个竞争频道上。结果是，整个教学现场从考前的“思维养成”到考后的“评分反馈”，被纳入了同一个精确代理的逻辑闭环。

4. 三股力量的咬合

至此，代理暴力的互锁增益结构已经清晰到可以被正式命名了。所谓“互锁增益”，指的是三股力量之间的双向强化回路，而非单线条的正反馈链。第一股力量是利害压力：当一分足以改变一个学生的人生轨迹时，任何评分模糊都可能引爆“公正性”争议。这股压力源源不断地为第二股力量——代理精确化——提供能量和合法性：评分系统每一次争议后的回应，都是进一步把评分细则写死、把采分点加密、把教师裁量空间收窄。而代理精度的每一次提高，又为第三股力量——合法性叙事——提供了更坚固的道德盾牌：“评分标准已经这么细化、这么统一了，你还能质疑它不公平吗？”合法性叙事的每一次巩固，反过来封锁了“降低利害权重”的改革空间——因为“降低利害权重”在这个叙事框架里会被翻译成“稀释公平”，从而在公众层面失去道义支持。三股力量互相提供能量、互相赋予正当性，形成了一台“谁也不能先停”的自我强化引擎。

更棘手的是，这台引擎对“外部扰动”展现出极强的免疫力。新方案——不管是综合素质评价还是档案袋叙事——只要一被纳入高利害分配体系，上游的精确化压力、中游的职业避险逻辑和下游的商业化分解，就会迅速为其套上同构的锁。精确评分系统不是一天建成的，它是这套互锁增益结构二十年持续运转的产物。而要理解这套结构对成长到底做了些什么，只看宏观制度史是不够的——必须往下，进入微观的课堂现场。

三、微观现场：什么在代理管道中消失了

上一节在制度层面追踪了代理精确化的传导链条。但宏观叙事有一个天然的盲区：当你站在制度的高度上，你看到的是政策、数据、趋势——你知道“评价窄化了学习”，但你不太容易看见窄化到底窄掉了什么。被排除在代理管道之外的，恰恰是那些无法在制度层面留下痕迹的东西。

这些痕迹，需要一间教室，和一个学生。

1. 两个学生，两条路径

华东某公立初中，2018年秋季学期。一位数学教师在自己班级中做了一项为期六周的观察记录，主题是八年级“一次函数与几何图形综合题”的学习过程。观察方式为随堂笔记与课后录音访谈相结合，所有观察均在被观察者知情同意的前提下进行。全班四十二名学生中，约二十人已在期中考试中能够熟练解答此类题目，另有十几人仍在及格线附近挣扎。教师的观察焦点不是“他们哪里错了”，而是“他们在错到对之间的这段时间里究竟在做什么”。

学生甲是后一组中的一员。一道典型的课后练习题：已知一次函数y = kx + b经过点A(1, 3)和点B(3, 7)，求k和b的值，并判断点C(2, 5)是否在函数图像上。对班上的优等生来说，这道题的标准解法是列出二元一次方程组、解出k和b、代入验算，全过程三分钟之内完成。

学生甲前后花了将近四十分钟。他最先尝试的路径不是列方程，而是直接在坐标纸上手动描点、连线，试图“看着图形读出k和b”。他读错了b值。之后他换了一个方向：依稀觉得“b就是x等于0时的y”，试图从这个模糊感觉反推出k，绕了十几步之后推反了。同桌探头看了一眼，建议说“你直接用两点的y值差除以x值差不就完事了”。学生甲没有采纳——不是因为不信任，而是因为他“想自己把它想通”。最后，在反复调试了几个数值、手动验算了每一步之后，他打通了那个一直卡住的关节，自言自语：“原来k就是每增加1格x，y涨多少。”期中考试时，学生甲在这类题上的得分和同分段学生并无显著差异。用代理品（考试成绩）来衡量，那四十分钟是纯粹的低效。

但教师在后续追踪中发现了一组引人注目的对比。在期末考试的同一知识点变式题中——题目不再是给定两个具体点，而是要求在给定函数图像在两个坐标轴之间截距的条件下求k值——班级中那些原本靠套公式稳定拿分的学生有相当比例出现了失误，而学生甲做对了。期中考试中还有另一道将一次函数与几何图形组合的新情境题，班级中被公认为“数学好”的五名优等生中有三人在此题上出现严重失误——他们的错误不是计算粗心，而是“把公式套错了情境”。一名优等生在试卷讲评课上对教师说：“我以为这道题就是之前做过的第5类题型，直接套了公式。我不知道它其实问的是另一种东西。”

这里暴露出来的结构是清晰的。那些能熟练解答标准试题的优等生，他们习得的核心能力是模式识别——在大量重复训练中建立“题型—解法”的映射关系，在考试时快速调用。这种能力在标准化的评价装置中极其高效，因为它精准匹配了代理品的生产要求：在规定时间内，产出与标准答案高吻合度的答题内容。但它内建了一个致命的脆弱性：当题目的表面特征与训练过的题型不完全一致时，模式识别的调用可能指向错误的解法——学生没有“重新理解这道题在问什么”的习惯，因为在他的学习过程中，这种“重新理解”从未被需要过。而学生甲在期中考试前的四十分钟里，是在做一件完全相反的事：他不是在往自己的记忆里存入“这道题该用哪个公式”，而是在从各种不同的角度去逼近“斜率到底是什么”。这个生成过程在期中考试时没有被代理品捕捉到——他和其他人拿到了同样的分数。但到了期末的新情境题，这个未被捕捉的生成过程显现出了它的另一个形态：弹性。

这个对比不能简化为“甲好、乙差”的道德判断。它追问的是另一件事：为什么学生甲那个弹性的生成过程，在代理品里是无法被看见的？为什么优等生那种高效的模式识别，反而在代理品里被记录为“优秀”？

这个追问的答案，需要从学生乙的学习路径中去寻找。教师观察到了一名与学生甲期中成绩几乎相同、但学习策略截然不同的学生乙。学生乙每次遇到不会的题，第一反应是打开手机上的搜题App，拍照、看解题步骤、把正确写法抄在本子上。有时他会在抄完之后再自己做一遍相同的题，直到能快速复现标准解法为止。在期中考前复习时，学生乙可以流畅地解答大量曾经“做错的题”——实际上，他所复现的是他抄下来并反复练熟的标准解法。但在期中考试中，他在这类题上的表现却和学生甲一样挣扎：题目稍一变形，他记忆中那条从“题型表面特征”到“标准解法”的通路就失活了。

2. 什么被格式排除了

把学生甲和学生乙并置，不是为了把乙钉在“被诊断者”的位置上——恰恰相反，乙的学习策略在现有的评价土壤中是一种高度理性的适应。问题不在乙的“选择”，而在评价系统的“格式”。当前的评分细则格式——在限时、标准答案锁定、采分点逐项赋分的制度框架下——在物理上只接收一类证据：与标准答案表述高度吻合的书面作答。这个格式本身就是一个过滤器。它过滤的，不是“对与错”，而是“哪种学会的方式能被看到”。

学生甲花费在反复调试上的那四十多分钟里，有大量时间是在走弯路、犯错、自我纠正。这些弯路是他最终打通斜率概念的必要过程——不是“浪费”，是那个生成过程本身的空间要求。但在限时的、标准答案固定化的考试中，这种空间被压缩为不存在：一道中考题的平均作答时间设定在八到十分钟，学生甲在那四十分钟里试图走的任何一条偏离标准路径的尝试都不会被等待。限时评价因此完成了一个在物理意义上就不可逆的排除：它不仅不奖励生成，它直接剥夺了生成发生所需要的第一个条件——时间。

搜题App所代表的那种即时反馈文化，进一步剥夺了第二个条件——安全感。学生乙永远不用经历那种“我走了十几步发现走反了、再回过头来重新梳理我的算错点”的感觉。而那种感觉——在不确定中忍受自己暂时不够好——恰恰是学生甲最终打通斜率概念的关键。当我们把“遇到问题—即时查看标准答案—模仿标准解法”培育为一种广受接受的“高效学习”策略时，我们并没有注意到，这套策略在物理上跳过了认知生成过程所必需的“混乱期”。混乱期不是学习过程中的低效部分，而是新概念从内部生成的唯一温床。搜题App——以及更广泛意义上的标准答案即时可得性——将这个温床从学习习惯中彻底清除了。

但比时间和安全感的剥夺更深的，是一种内部排序的翻转。当评价系统持续奖励“在规定时间内精确命中采分点”而从未对“绕弯路”给出任何制度性的承认时，“绕弯路”就不再只是一个时间管理上的不经济，而是被学生内化为一种自我判断——“这说明我不是学数学的料”。这正是代理暴力最深的一层：代理的逻辑已经不依赖外部制度的强制，而是被学生自己接过来、安在自己体内，然后自己动手切除那些不能在外部系统里“拿到证据”的内在发生。学生对自我的那些“太笨了”“不适合”“不够好”的判断，并不来自教师的责骂，而是来自他们自己在评价装置里日复一日积累下来的、对“什么才值得被看见”的逐渐内化。

3. 这不是在讲“本真学习”的故事

本文必须在这一节结束时履行一个严肃的文本纪律：上述分析不是在建构一个“学生甲＝本真学习、学生乙＝被异化的学习”的二元框架。那种框架正是本文从一开始就拒绝的发现学叙事——它预设了一个“评价介入前的纯真学习”，然后把干预描述为对那个本真状态的玷污。

学生甲那四十分钟能够发生，是有条件的，而不是“天然的”。把它展开来看：它发生在课后，而不是在课堂上被计时训练挤压的时段；它发生在这个特定的教师没有在他绕弯路时出手干预、直接给出标准答案的时刻；它发生在这个学生本人恰好有一种“想自己把它想通”的性情倾向——这种倾向本身，又是在什么样的家庭教养和早期学习经历中被形成的？把这些条件一一摊开，你看到的不是“本真学习”，而是一种特定的认知形态，在特定的纠缠条件（一位在当下没有给出标准答案的教师、一段课后不被计时的时间、一种被早期经历养成的不轻信现成解法的倾向）中，被发生出来的一个临时稳态。换一个教师——一个更担心学生走弯路、更倾向于在学生遇到困难时立刻给出标准解法的教师——学生甲那四十多分钟就不会发生。换一种时间条件——比如这道题是贴在课堂限时练习的卷子上，设定解答时间八分钟——学生甲那四十多分钟也不会发生。

而学生乙的“捷径策略”，同样是在特定的纠缠条件中被发生出来的另一种稳态：搜题App的可及性、即时反馈文化、限时训练中积累下来的“不确定就是危险”的经验，以及一个关键条件——这套策略在期中考试中确实有效（他拿到了和学生甲一样的分数），从而得到了外部系统的正面确认。乙的策略不是“被接管”的病理，而是他在给定条件下做出的理性适应。只不过这种适应的长期后果是：他从未经历那个从内部打通的过程，而这个缺失，在期末考试的新情境题中才暴露出来——暴露为脆弱性。

因此，本节的核心论证不是“我们应该保护学生甲、抵制学生乙”。而是：当前高利害评价系统的代理格式——限时、标准答案锁定、采分点吻合——对这两种认知形态有一种系统的选择性。它选择了乙的“快速吻合”，排除了甲的“生成弹性”。不是因为甲的答案错了，而是因为甲的答案在通向“对”的过程中必须穿越的那些弯路、混乱和自我修正，进不了这个代理格式的管道。代理品因此不是测量了成长——它筛选了哪种成长能够被计为“成长”。而在高利害条件下，“被计为成长”直接等于“被允许继续存在”。

四、代理何以成为暴力：概念的发生学结算

前面的制度史追踪和微观课堂分析各自给出了一条完整的线索。现在，需要让这两条线索在同一个交叉点上会合，逼出一个本体论级别的结算。

1. 不是代理品，是代理本身

先说这个判断不是什么。它不是“这个代理品坏了，让我们换一个好的”。这正是过去半个世纪评价批评者和改革者反复在做的事。坎贝尔定律的批评者指出指标会被腐化——换一个更能抵抗腐化的指标。中国教育改革的推动者指出“唯分数论”的弊端——换一套综合素质评价。档案袋、成长叙事、过程性评价，每一次新工具的引入都伴随着同样的叙事：“这次终于可以把那些分数无法衡量的东西纳入评价了。”然后每一次，新工具被应试机器捕获、分层、包装成新一轮的应试培训产品。换了所有的尺子，现象只是换了面孔继续运转。

这个循环本身就是一个诊断信号。当一个系统对所有扰动都展现出同构的复原力——不管你从哪个角度推它、塞给它什么新方案，它都反弹

回同一个稳态——这说明症结不在被塞进去的那个东西的好坏，而在系统本身的深层结构。这个结构就是代理。

在高利害分配的条件下，“学生的真实成长”本身是不可以直接被比较的。它只能通过一种中间格式——一个分数、一个等级、一段叙事——来进入分配决策。这个中间格式本身就是代理。而代理的运转逻辑天然地要求：把被代理物中那些无法被格式化的维度切除，只保留那些能被格式化的维度。这不是代理的“腐化”，是代理的“定义”——代理之为代理，就在于提取一部分、舍弃另一部分。当利害足够高的时候，被舍弃的那一部分就不再仅仅是“没被看到”，而是被系统性地驱逐——因为学生发现，把精力花在那一部分上不会产生代理证据，而精力花在另一部分上会。于是，那些需要时间、需要混乱、需要暂时不确定的生成形态，就被“高效学习”挤压到了教室的墙角。

这正是代理的暴力在结构上的核心位置。它不是测量了成长——它以“可比较”为筛子，筛选了“什么值得被培养”。那些无法通过代理管道的生成形态，不是被“测不准”，而是被系统性地从学习策略的选项清单中逐出。

2. “可比较＝公正”的合法性叙事如何堵死退路

如果代理本身就是一个不断排除的结构，那么为什么这个结构能够持续运转半个世纪而不被质疑？为什么在中国语境中，“评分标准越来越精确”这件事，总是被当作“公平在不断进步”来看待？

答案在合法性叙事那里。“可比较＝公正，不可比较＝偏私”这组等式，是现代高利害评价系统的基石叙事。它的逻辑看似无懈可击：要公正地分配稀缺机会，就必须基于可比较的证据来做决定；而只有可比较的证据，才能被第三方核实；能被第三方核实的，才是排除主观偏私的；排除了主观偏私的，就是公正的。于是，代理品的每一次精确化——从教师整体判分到分步采分、从分步采分到逐词核对——都不被看作对成长的一次更深的结构性排除，而被看作对公平的一次更有力的捍卫。中国语境中的“分数面前人人平等”之所以有如此强大的道义力量，不是因为它描述了一个事实（家庭资本对分数的隐性影响力已被反复实证），而是因为它提供了一个没有人能公开反对的合法性框架。在这个框架里，要主张“有些东西不能被比较，因此就不该被放进比较的天平”，在逻辑上就会被翻译为“你想用不透明的方式偏袒某些人”。

这种翻译堵死了退路。一旦“降低利害权重”被当作“稀释公平”来理解，任何试图松动代理结构的改革都会在公众层面失去道义支持。这就是为什么“多一把尺子”的改革路径一再失败的原因：尺子可以多，尺子可以软，但只要你没有动“最终决策还是依赖这些尺子之间的比较”这根基柱，代理结构的排除逻辑就在原地继续运转。它只是从一把尺子切变成多把尺子组合切——切除的力度是一样的，甚至更大，因为每一把新尺子都需要新的应试资源去解读和训练。

这就是“互锁增益”的三个环彻底咬死的地方。利害压力源源不断地为代理精确化提供必要性——不用精确尺子就无法承受争议。代理精确化源源不断地为合法性叙事提供论据——越精确越公正。合法性叙事则封锁了任何从根源上动手术的空间——谁敢动公平的根基？三股力量相互提供能量、相互赋予正当性，形成了“谁也不能先停”的死锁。要理解这种死锁对教育的实质影响，就需要回到“不可代理之物”这个核心概念上来。

3. 不可代理之物：不是玄学，是结构性排除的后果

本文主张“有些东西是不可代理的”，这不是一个形而上学的断言——不是在说“人类的认知生成在本体论上永远、绝对、在任何条件下都无法被代理”。这是一个历史性的、结构性的判断：在当前高利害评价的特定技术要求——限时、标准化、第三方可验证——之下，某些认知生成形态无法被转译为可比较的代理品而不被系统性地排除。

学生甲的案例为这个判断提供了经验底座。他被看到在新情境题中表现出弹性，不是因为他复习了更多的题型，而是因为他在那四十分钟里从内部打通了斜率概念。这个过程有一个内置的结构特征：它只能在“不被即时评价”的时间缝隙中完成。在限时的高利害评价场景中，这种时间缝隙在物理上被压缩为零。这不是说甲的答案不如乙的答案“好”——它们产出的书面痕迹在标准题目上是高度相似的。结构性排除发生在进入代理管道之前的那一刻：甲的那四十分钟内部的认知过程，压根就进不了评分细则的格式。它被过滤掉了，不是被“评低”了。

这意味着，“不可代理”不是一个静止的属性（“有些东西天然不能被测量”），而是一个动态的后果：在给定的代理格式中，代理格式自身所要求的那种“可比较、可标准化、可第三方验证”的证据格式，会在证据生产的上游就已经开始筛选学生的认知路径。那些不符合这种格式的生成过程，不是被“测不准”——是根本就不会被启动，或者一启动就被学生自己掐掉，因为“这条路不产分”。所以，说“有些东西是不可代理的”，实际上是在说：代理结构在生产代理品的同时，也在生产一系列“无法被代理的生命经验”——而这些经验，恰恰是那些会在新情境中展现弹性的生成过程所依赖的土壤。从这个角度来说，代理的暴力不是“测量错了”，是“以可比较为尺度对学习本身进行了一次从源头的重新组织”。

五、不可还原性论证：这不是异化，不是象征暴力，不是物化

一个判断要立住，不能只靠正面推演，必须经受住从最邻近的概念从侧面杀过来的对质。本节要完成的事是：逐一检视那些最可能、也最有资格与“代理的暴力”竞争解释权的概念，证明它们在解释当前经验材料时所遭遇的内部困境——并从这个困境中逼出“代理的暴力”的不可替代性。

1. 与默会知识传统对质：Polanyi看到了不可言传，但没看到结构性排除

Michael Polanyi的默会知识传统指出了一件事，在认知论层面与本文的核心关切高度接近：知识总有一个“寄托的维度”（subsidiary awareness），无法被焦点意识（focal awareness）完全捕获。你会骑自行车，但你说不清你怎么保持平衡的；一个数学家知道一个证明“对”，但他不一定能把这份直觉完整地写成命题逻辑链条。Polanyi的命题是：不可言传是知识的构成性特征，不是认知的某种低级阶段。

在这个意义上，Polanyi确实为本文提供了一个强大的哲学盟友。他的洞见可以直接用来反驳某种天真的“完全可测量主义”——如果所有的知识都有不可言传的维度，那么任何试图用完全外显的代理品来“充分代表”真实认知能力的企图，在认识论上就是站不住脚的。

但Polanyi有一个盲区。他的分析停在个体认识论层面——他关心的是“单个认知者的知识结构中，哪些维度可以被自己意识到、可以被自己表达出来”。他没有追问另一个更致命的问题：当一套制度以“可比较的证据”为分配稀缺资源的唯一依据时，它对那些默会维度做了什么？Polanyi的框架可以解释“为什么学生甲那四十分钟里的认知过程无法被完整地写进答卷”（因为默会成分无法被焦点意识穷尽），但它无法解释更关键的一步：为什么在知道有默会成分存在的情况下，高利害评价系统仍然选择不去保护它、甚至反过来通过对可比较证据的奖励来挤占它——而且这种挤占被社会合理化地描述为“为了保证公平”。

这正是“代理的暴力”切开了Polanyi装不下的区分面。Polanyi告诉你的，是有些东西不可言传。代理的暴力告诉你的，是在高利害分配的结构中，不可言传者会被系统的格式逻辑逐出“值得被看见”的疆域——不是因为有人否认它的存在，而是因为它在逻辑上就进不了分配决策的格式管道。Polanyi + 高利害语境，这个组合可以解释一部分现象（“默会知识被忽略了”），但它无法解释另一部分更尖锐的现象：为什么保护默会知识的努力（如过程性评价、档案袋、成长叙事）一经纳入高利害体系，就会以另一种形式重新被格式化的力量捕获。因为症结不在“知识有一部分是默会的”，而在“代理这个结构本身就是在高利害压力下不断自我精确化的——不管代理品是什么，只要利害足以决定机会分配，代理品就必然会被应试技术逐层分解到最可复现、最可比较的颗粒度”。这个动态，是Polanyi的默会知识理论没有预见的，也是它无法解释的。代理的暴力的理论收益，就在于指认这个动态——并指出它不是一个可以被“更完善的认识论”解决的问题，而是一个必须被“松动代理结构”才能有所缓和的困境。

2. 与物化与商品化传统对质：Lukács看到了物化，但没看到格式排除

马克思主义传统中的“物化”概念，在Lukács那里达到了其最尖锐的哲学表述：人与人之间的关系被呈现为物与物之间的关系，人的主体性活动被凝缩为可计算的对象属性。在教育语境中，这个结构可以这样转译：学生活生生的学习活动，在评分系统的凝视下被凝缩为一个数字——分数——而这个数字反过来成了学生“价值”的标签。分数成了物化的学习。

这个批判传统与“代理的暴力”之间有着深厚的亲缘关系。二者都指认了一个过程：人的内在过程被强制转译为外部可比较形式。而且，卢卡奇的分析中有一个洞见——物化不只是一种错误的认识，它是资本主义社会结构的客观形式——这一点与本文拒绝把代理暴力归因于“评价者素质不高”或“指标设计不够科学”的立场是高度一致的。

但物化理论在教育语境中有一个关键盲区，而这个盲区正是“代理的暴力”要切开的区分面。物化批判的典型结构是：交换价值殖民了使用价值——事物本来的“有用性”被它在市场上的“可交换性”覆盖和扭曲。转译到教育中：学习本来的价值（理解、生成、自我转变）被可交换的价值（分数、排名、升学资格）覆盖。这个批判预设了一个前提：有一个“使用价值”——有一个本来的、未被异化的学习价值——存在在那里，只是被交换价值殖民了。

代理的暴力拒绝这个预设。它要追问的是另一件事：把学生甲那四十分钟里的认知生成过程拿到评价系统的天平上时，它遭到的不只是“价值被覆盖”——它遭到的是“格式不兼容”。甲的生成过程在进入这个天平之前就被过滤掉了，因为它无法满足“可比较证据”的格式要求。这不是交换价值覆盖了使用价值——这是比较的格式在生产环节就已经完成了对“什么能进入比较”的筛选。更关键的是，这种筛选造成的排除，即使在一个完全不存在商品交换逻辑的评价系统中——比如一个纯粹基于教师共同体的质性评审——也同样可能发生。只要分配决策最终依赖于某种形式的“可比较证据”，那些进不了这个格式的生成形态就会被系统性地边缘化。这不是商品化的后果，是代理格式本身的后果。“代理的暴力”的理论收益在这里是明确的：它不依赖于交换价值/使用价值的二分，不预设一个非商品化的教育乌托邦，而是指出——在代理关系内部，比较本身就是一个有选择性的暴力。

3. 与Illich对质：废除制度不是出路

Ivan Illich在《去学校化社会》中提出的尖锐批判，在议题层面与本文几乎同构：他对制度化价值测量的攻击——“学校把学习变成了对教学过程的消费，把成长变成了对制度化升级的服从”——直接命中代理暴力的核心关切。因此本文必须正面回答一个挑战：既然Illich已经说了这么多，为什么还要发明“代理的暴力”？本文是否只是Illich命题在中国高考语境中的一个经验注脚?

本文对Illich的回答分为两步。第一步：承认亲缘性。Illich对制度化的批判，确实在方向上与代理暴力有重叠。但第二步：指出根本的差异。Illich的诊断指向的解决方案是废除——废除学校的制度化垄断、废除义务教育的法律强制、回归到一种非制度化的、基于学习者自发的“学习网络”。Illich是彻底的制度废除主义者，对他来说，制度本身就是问题，答案在制度之外。

代理的暴力不做这个飞跃。它承认一个Illich不愿承认的现实约束：在大规模现代社会中，高利害分配——无论是升学、就业还是其他资源——不太可能被完全取消，而只要有分配，就会有人要求程序公正，而只要要求程序公正，可比较证据就必然以某种形式介入。代理的暴力不是要废除高利害分配，而是要在承认这个结构的前提下，追问：在代理关系不可避免的条件下，我们能做什么？答案不在废除，而在松动——“降低利害权重”“划定不可代理的制度保护区”“改变分配结构使单一代理品不再是命运的决定者”。这些是在承认矛盾不可能被一劳永逸“解决”的前提下，寻找局部减少暴力的工程路径。这个立场与Illich的根本不同在于：Illich用“废除”来逃避矛盾，而代理的暴力选择留在矛盾里，然后问“在这个矛盾中，哪里是暴力的最痛点、哪里可以减轻其压力”。

这正是代理的暴力不能被“制度化批判”替换的区分面：它不是在喊“打倒制度”，而是在追问“在制度内部，暴力被压缩在哪几个关节处——以及有没有可能在某一两个关节处松动螺丝，让生成的土壤稍微多出一点不被代理吞没的空间”。

综合以上三重对质，代理的暴力不能被任何一种既有概念1:1替换。它不是Polanyi的默会知识加高利害语境（Polanyi没有看到代理格式自身的生产性排除功能），不是Lukács的物化（物化批判预设了被交换价值覆盖的使用价值，未看到代理格式本身的排除不依赖商品逻辑），也不是Illich的去制度化（废除不可行，只能在承认矛盾的前提下局部松动代理结构的螺丝）。它切开的区分面是：代理作为一种结构，在高利害条件下本身就构成一个不断自我精确化的排除机制——这个机制的系统性与内生加速性，是既有批判传统都未能单独命名的。

六、反驳与边界：代理的暴力的自我限定

前述论证完成后，两个最有力量的反驳必须被正面接住。它们不是稻草人，而是每一个严肃对待这个判断的读者——尤其是来自教育政策制定和测量学领域——必然会提出的质疑。本文对这两个反驳的回答，同时也就是“代理的暴力”的理论边界所在。

反驳一：如果代理本身即暴力，那么要如何公正分配?

这个反驳的表述通常是这样的：“你批评代理，但你能给一个替代方案吗？在高利害分配不可避免的前提下——重点大学就那么几所，好工作就那么多——不用可比较的证据来决策，难道用推荐信？用面试？用抽签？那些方式难道不是更不透明、更容易被阶层特权捕获吗？”这个反驳的力量在于，它不动摇代理的事实效用——代理确实在一定程度上防止了更赤裸的权力寻租——而是逼问批评者：你的替代方案是什么?

本文的回答不是“取消代理”。因为取消代理确实可能导致更糟糕的结果——布迪厄已经反复论证过：在缺乏透明程序的地方，阶层特权往往以“品味”“素养”“综合素质”等软性语言运作得更加顺畅。法国精英大学的口试传统和中国古代科举之前的察举制，都是证据：它们没有硬性的、可被公众监督的代理品，但阶层再生产不仅没有减轻，反而因为有“个人判断”的掩护而变得不可上诉。

所以本文不主张废除代理。本文的主张是：必须区分“代理的必要性”和“代理暴力的可减轻性”。代理在当前条件下是必要的——这个必要性本文不否认。但这不意味着目前这套精确化到极致的代理结构就是必要性的唯一形态。在“有代理”和“代理完全暴虐”之间，存在一个政策设计的梯度空间。具体而言，有三条路径可以在不废除代理的前提下降低暴力的程度。

第一，降低利害权重——不是在技术上更换代理品，而是在结构上让单一代理品不再决定一个人的命运。如果一次考试的分数不再直接锁死一个人四年乃至更长的教育路径（例如，通过扩大转专业自由度、增加大学后阶段的再选择通道、削弱“第一学历”在劳动力市场上的信号强度），那么代理品的“高利害性”就被局部撬动了。代理的结构性排除力量与利害程度是成正比的——利害越低，学生就越有余地在“产出代理证据”和“保护自己的生成过程”之间做出平衡，而不会被迫把这些生成过程全盘牺牲掉。这不需要废除高考——它需要的是在高考之“后”打开足够多的再选择管道，让第一次代理的误差不至于成为终身的枷锁。

第二，划定不可代理的制度保护区。这个概念是从学生甲的案例中推出来的——他是在一个暂时的、不被精确代理的缝隙里（课后练习，不计时，没有人看草稿纸上的乱线）完成了他的内部生成——这个缝隙就是“不可代理保护区”的雏形。制度性地承认：在教育的某些阶段和某些领域，“生成”优先于“可比较”，且这种优先必须得到制度性的保护——比如，在低年级阶段不引入高利害排名，在课堂教学中为探究性学习保留不被即时计分的时间，在课程标准中把“容忍错误、允许绕路”明确为教学目标（而不是只是口号）。这不是“软性评价”的

升级版——它不是在找更好的代理品，而是在划出代理的禁区。

第三，承认“不可代理之物”并公开声明评价系统的有限性。当前评价系统的合法性叙事建立在一种虚幻的全称命题上：“我们评价的是学生的全面成长。”只要这种全称命题在运转，那些被排除在代理管道之外的东西就会被定义为“不存在”——或者更糟，被定义为“那个学生没有”。一种更诚实的做法是公开声明：本系统只评价那些能够被可靠代理的部分——知识的准确度、在规定时间内运用标准解法的效率——而不评价那些无法被代理的部分——生成弹性、概念深度、情境迁移能力。这个声明不会让暴力消失，但它至少会松动“可比较＝公正”的等式——它会告诉社会：公正的只是比较过程，而不是比较内容；我们只比较了那些能被放进格式的东西，还有许多东西在这个格式之外，那些东西的教育价值并不因为进不了格式而降低。

这三条路径都不废除代理。但它们都在动代理结构的螺丝——降低利害权重、划定代理禁区、公开承认评价有限性。它们共同指向的是同一个工程判断：代理本身的结构性排除不可能被消灭，但它可以被部分地限制、部分地公开化、部分地减轻其压力。这不是“温和的改革方案”——它不追求感动自己的叙事闭环——而是在承认矛盾不可能被一劳永逸解决的前提下，寻找局部抵抗的空间。在这个意义上，代理的暴力既是一个诊断，也是一种在诊断中不断退回到现场去辨认缝隙的动作。它是一种不得不睁开着眼、留在矛盾中的工具箱。

反驳二：内部生成如此脆弱，它不是本身就该被淘汰吗?

第二个有力量的反驳来自另一侧：“如果学生甲的那种‘生成’如此脆弱——只是一场限时考试就能摧毁它——那么它是否本身就证明自己不值得成为教育的核心目标？也许教育恰恰需要培养一种能在代理压力下仍然保持内部生成能力的韧性，而不是一味谴责代理压力太大了。”这个反驳的隐藏前提是：一个健康的认知生成过程，应该具备足够的 robustness——它不应该因为外部压力而轻易塌缩。如果它塌缩了，说明它自身的结构不够强健。因此问题在学生，不在评价系统。

本文的回答是：这不是在讲脆弱性，而是在讲条件性。学生甲那四十分钟的生成过程之所以能在期中考试之前发生，是因为他在那个特定的课后场景中拥有一个不被打断的时间缝隙、一个不被他那双随时准备按下搜题App的手所诱惑的空间、以及一个没有在他走弯路时直接给出标准答案的教师。这些条件中的任何一个发生位移，他的内部生成过程就不会发生。这不是说这个过程“脆弱”——这种说法已经是在用发现学的眼光去度量它：把它当作一个物，然后问“这个物结实不结实”。它是“有条件性”的——它被发生出来所需要的环境门槛很高。高利害评价系统的代理结构所做的，不是“考验”它的强度，而是系统性地移除它所需要的环境条件：时间被压缩、空间被即时反馈文化填满、教师被评分争议压力倒逼去提供标准答案。在这种条件下，学生甲的生成过程不是因为它“脆弱”而瓦解——而是它根本就没有机会被发生出来。

更进一步，这种“robustness”的诉求本身就在玩一个危险的循环论证。它说：我们要培养一种能在代理压力下仍然保持自我生成的学生，因此代理压力是有益的考验，能筛选出真正 robust 的学习者。但代理压力所创造的环境，恰恰是系统性地惩罚那些需要时间、需要混乱、需要在不确定中推进的生成过程的。你在这个环境中筛选出来的“幸存者”，不是因为你的教育培养了他的生成——而是因为他的生成发生在了你系统性的惩罚覆盖不到的地方——比如，在课外、在自学时间、在一个正好愿意承受“学生考砸”的风险的教师那里。你把这些幸存者的成绩当作代理压力有效的证据，但实际上，你只是在庆祝一种抽奖的结果。

代理压力的生态效应是：它不是在“培养韧性”，而是在惩罚最需要保护的那些生成形态，然后声称能撑过来的才是真正的韧性。这是一种玩牌者摸牌之后再宣布游戏规则的做法。因此，这个反驳不仅没有动摇代理暴力的诊断——它恰好从反面证明了代理暴力的隐蔽性：代理压力通过对特定的认知形态进行系统性排除，然后声称“被留下的才是强的”，从而将排除的结果自然化为能力的高低。

七、跨域印证：大语言模型与“理解”的代理

一个判断要想被淬炼出足够的普遍性，不能只在一个领域里经受考验。本节将把“代理的暴力”带入另一个场域，不是做类比，而是做一次结构性的并置——用它去看大语言模型对“理解”的训练方式。这个并置的目的不是证明代理暴力“无处不在”，而是检验：当一个看似截然不同的领域中暴露出同构的结构时，你是不是更可能相信，前文从高考评分制度史和课堂田野中所逼出的那个判断，不是一件只在教育领域发生的偶然事故，而是一种有更一般性的发生学结构。

1. 大语言模型的训练逻辑：把理解转为下一个token的概率

大语言模型的工作机制在今天已经被充分讨论，但其评价训练逻辑与教育代理之间的结构亲缘性仍少有人触及。简要来说，当前大语言模型的预训练任务是“预测下一个词”——给定一段文本的前文，模型需要输出下一个最可能出现的词，然后将其与真实文本的下一个词作比较，计算损失，再通过反向传播调整参数。经过海量文本的反复训练，模型学会的不是“理解”一段话的意思，而是“给定上下文，最可能的下一个token是什么”。

这个结构能否被看作一种代理？可以。真实的理解过程——包括推理、权衡、在多个可能解释之间反复调试、调用背景知识等等——在这一训练范式中被强制转译为一个单一的外部证据：给定前文的条件下，模型预测的下一个token与语料库中真实token的吻合度。这与高考评分结构之间有惊人的同构：学生的认知过程被强制转译为“给定试题前文所构成的题干约束下，其作答与标准答案采分点的吻合度”。两个场景中，内在的生成过程都被一个单一的、可比较的外部证据格式所代理。不同的是，在高考场景中，这个代理发生在人类学生身上；在大模型场景中，它发生在人工神经网络上。

这不是说大模型“受到了伤害”——它没有感受性。它恰恰因此成为一个纯粹的检验场，用来分离一个困惑：代理的侵害到底是来自对人性的侮辱，还是来自代理结构本身的内在逻辑？如果伤害只来自对人的伤害，那么大模型因为没有感受性，对它施加代理暴力就不会产生任何后果。而如果代理结构本身在模型身上提取并筛除了某些东西，那么这表明代理暴力的核心动力在它的格式逻辑而不在它的对象是否是“人”。

2. 什么在下一个token的预测中被筛除了

回答这个问题的关键，在于看大语言模型在它的训练范式下，到底“学”到了什么，又是什么东西被它的训练信号的格式挡在外面了。

模型学到的是高维概率分布——在给定一段文本前文的所有语义模式中，哪些词的组合是更常见的、更符合语法的、更贴近训练数据中人类写作者的写作习惯的。这个能力极其强大：它可以生成流畅的论文、通过专业资格考试、模仿各种文体风格。但它被分离出来的一件事是：它不靠“推理”来生成文本。所谓“推理”，在这个框架下，不过是一个足够长的、在训练数据中反复出现的推理链在生成时被高概率地复现出来。模型不能“检验”一个命题的真值——它只能预测对于一个命题，训练数据中的人通常是怎么接续它的。

这个结构把“理解”这个极难定义的东西，重新转化成了一个可被精确检验的边界案例。考虑以下场景：某大语言模型被问到一个数学问题，它的回答是正确的，步骤清晰，与标准解法完全一致。但稍加改动问题中的数字或表述方式，模型就答错了——而且是沿着一个完全同样的“标准解法”的语法，套到了一个不兼容的新情境上。这个失败与本文第三节中那些靠套公式得分的优等生在变式题上的失败高度一致——两者都是“模式识别”到了表面特征上，却无法在概念结构发生位移的新情境中重新判断。换句话说，模型表现出来的不是“理解”，而是“对理解在文本中的代理品的极高模拟精度”。

这个并置从反面给了代理暴力一个跨域的检验证书。在高考场景中，学生甲的生成弹性之所以无法被代理品捕捉，是因为他的生成过程依赖一种在多个可能的路径之间犹豫和调试的时间结构——而这个时间结构在代理品的格式中被压成了一个即时产出的痕迹。在大模型场景中，你看到了同样的结构，差别只在于载体：模型没有内部的时间结构——它只有一个前文到下一个token的概率分布。模型训练信号里的那些困惑、调试、被放弃的路径、反复的自我纠错，都被挡在了“下一个token预测损失”的格式之外——模型从未被要求穿越这些状态，因此它从未习得穿越这些状态的能力。

但这恰恰说明了一个事：代理格式所排除的，不是一个形而上的、玄学的“人类灵魂”——而是认知过程中的一种特定的时间性。这种时间性，是一种认知过程在从“不确定”通向“确定”时所必须穿越的时间展开，以及在这个时间展开中会出现的大量暂时不产生正确输出的自我修正。代理格式在逻辑上无法容纳这种时间性——因为它要求的是产出一致性的、可被即时比较的最终输出。这就是为什么代理暴力的结构不仅出现在人类学生身上，也毫无障碍地出现在人工神经网络身上——不是因为它欺负人，而是因为它的格式逻辑本身就把“穿过不确定的时间性”这一维系统性地排除在外了。

3. 计算主义的反驳——以及为什么它打不中要害

一个来自计算主义传统的标准反驳应该在此被正面回应：这个反驳会说，“你所说的‘穿过不确定的时间性’，本质上只是一种更高维的概率估计过程，或者是一种带噪声的贝叶斯更新。大模型之所以不展现这种能力，只是因为它的训练范式没有给它提供这种格式的数据——如果你换一种训练范式，比如强化学习让它在犯错中自我纠正，它也许就能‘学会’穿越不确定的状态。因此你所说的‘不可代理性’，不过是一个当前的工程技术局限，而不是代理本身的结构性问题。”

这个反驳的力量在于它把问题完全安置在了技术优化空间之内——没有不可代理的东西，只有还没被代理技术优化到的格式。一旦我们发明了能代理“时间性”的新格式，代理暴力就消散了。

但这个反驳犯了一个关键的范畴错误。它把“时间性的体验”与“时间性的跟踪记录”混淆了。当学生甲在那四十多分钟里反复调试自己的认知时，他不仅在产生一系列可以被记录下来的步骤——他还在承受一种“我不确定我是不是对的”的压力，并且在承受中继续走下去，直到自己内部打通的那一刻。这是一段有主体承受性的时间。你可以用任何技术设备全程记录他发出的每一个外在行为——他说的话、他打的草稿、他凝视纸面的时长——但这些记录不是他的承受本身。承受性不是一种行为数据格式，而是一种第一时间无法被第三视角捕获的那个正在穿越不确定状态的内部位置。任何外部记录——不管你把它打成多细的时间切片——记录的都不是承受性，而是承受性通过行为投射出

来的影子。

因此，代理暴力的“不可代理性”指向的不是“当前技术无法记录时间过程”，而是“认知生成过程中那个穿越不确定的主动承受性，在进入任何第三视角可比较的格式时，都必然被格式化为时间序列的动作描述，而不是被承受性本身所穿越”。这个区分不能被技术进步消除——因为它是本体论层面的格式不兼容，不是技术精度的不足。你可以不断优化数据格式去逼近它——那些行为数据确实比什么都没有要好——但只要决策的依据是被比较的代理品，代理品就仍然在逻辑上把“承担不确定性的主体位置”排除在可比较的疆域之外。

这正是代理暴力的最深层结构。它不是测量工具不精确——它是工具的格式无法承载一种只能由第一人称位置承担的内部时间。如果这个结构在大语言模型的纯粹技术逻辑中都能被追踪出来，那么它就更不可能只是教育领域的一桩官僚主义事故了。

八、结语：留在矛盾中，然后辨认缝隙

本文从这样一个追问出发：为什么过去半个世纪所有替代代理品的改革都失败了？回答是：因为从未有人质疑过代理本身。代理结构在高利害条件下天然地倾向于不断精确化自己，而每一次精确化都在系统性地排除那些无法进入代理管道的生成形态。这个排除不是腐败，不是技术失误，不是道德败坏——它是一个干净的、合规的、甚至被颂扬为“公平进步”的结构性过程。

本文的任务，就是在制度史的追踪与课堂田野的并置中，让这个被忽略的结构显露出来，并将它命名为“代理的暴力”。这不是一个呼吁废除评价的宣言。它不承诺一个没有代理的教育的黄金时代——正如本文反复指出的，那个时代从未存在过，也不太可能到来。代理的暴力是结构的产物，而不是某个人或某一群人的恶意的产物。结构可以被识别、可以被局部松动螺丝，但不太可能被废除。

因此，本文的终点不是一套改革方案，而是一个诊断，以及在这个诊断约束下的几个工程判断：降低利害权重，让一次代理不再决定一生的赛道；划定不可代理的制度保护区，在教育的某些阶段和某些领域里，让生成优先于可比较；公开承认评价系统的有限性，让社会知道，被评价到的只是可被代理的那部分，还有许多东西在评价之外——它们在教育上仍然重要。这三笔，没有一笔能“解决”代理的暴力。但它们在承认矛盾的前提下，可能让一些本来会被彻底排除的生成形态，得到一处喘息的空间。

最后，本文给它自己的核心判断留下一个证伪条件：本文断言，代理结构在高利害条件下天然地倾向于不断精确化自己——而每一次精确化都在结构性地排除那些无法进入代理管道的生成形态。这个断言是可检验的。如果在未来某一个高利害评价系统中，出现了一种代理格式的转换，使得该评价系统在维持或提高利害压力的同时，出现了代理精确度的持续下降——不是技术上的倒退，而是制度性地、持久地拒绝把代理品进一步精细化，且这种拒绝不是因为外部政治管制而是因为内部评价逻辑的主动转向——那么，本文的诊断就需要被修正。因为这意味着代理结构和利害压力之间不是必然耦合的，它们的互锁增益可能在某种制度设计下被打破。在那一天到来之前，本文的判断将作为一把手术刀，继续留在评价批判的现场。

———

附记

本文的核心经验材料来自两项独立的田野工作：2018年秋季学期在华东地区一所公立初中进行的为期六周的参与式课堂观察（观察编码方案聚焦于学生在“接触新概念—练习—反馈—修正”周期中的行为序列与自我报告，数据形式包括随堂笔记、课后半结构化录音访谈、学生作业文本，研究方案已获研究者当时所属机构的伦理审查批准），以及对高考评分细则变迁的制度史追踪（文本来源包括1980年代至2010年代多省公开或半公开的命题与评卷工作手册、阅卷教师指导手册、教育部考试中心历年发布的考试说明等档案材料）。为保护参与者隐私，所有个案均已做匿名化处理。本文的论证不依赖这些个案在统计学意义上的“代表性”——它们的理论功能不是提供比例描述，而是通过追踪特定认知形态在特定条件下的发生与排除，揭示代理结构在生产可比较品的同时也在系统性地排除什么——被排除者的“总量占比”不改变它被排除的结构性这一判断。

九、最近邻辨析的补足：代理的暴力不是古德哈特定律，也不是可读性之失

原稿已与坎贝尔定律、布迪厄象征暴力、鲍尔审计社会、比斯塔可测量性批判、异化与物化传统、默会知识划界。此处补两个最贴、原稿未点名的占位，以钉死不可还原的边界。

其一，古德哈特定律（Goodhart, 1975）：一个指标一旦成为目标，就不再是好指标——它会被博弈、被操纵、被掏空。它与坎贝尔定律是一对孪生。其二，斯科特（Scott, 1998）的可读性（legibility）：国家为了治理，把复杂的地方现实简化为可读的标准范畴，这一简化对地方性技艺（métis）构成暴力。两者都极贴近"代理的暴力"。

但剩余分工在此清晰。古德哈特／坎贝尔定位的暴力，在于代理品被博弈、被腐化——失败出在代理的不完美（它被玩坏了）；其隐含承诺

是：一个完美不可博弈的代理品可以消除伤害。斯科特定位的暴力，在于简化造成的损失——复杂被压成可读范畴时，métis 被丢掉了；其焦点是化约与遗失。而"代理的暴力"主张一件更强的事：把不可代理的内在生成过程强制转译为可比较外部证据，这个动作本身即是侵害，哪怕代理品完美、不可博弈、且未丢失任何可被表征的信息。伤害不在代理品被玩坏（古德哈特），也不在信息被简化丢失（斯科特），而在"强制转译"这一动作对一个本质上拒绝被代理的过程所施加的结构性排除。

由此得到把三者分开的可裁决差别：古德哈特预测，一个完美不可博弈的代理品将消除伤害；斯科特预测，一个无损保真、不丢失任何信息的表征将消除伤害。而"代理的暴力"预测：即便代理品完美不可博弈、且信息无损，只要它被置于高利害分配条件下去强制转译不可代理的生成过程，结构性侵害仍然发生——因为被排除的不是被玩坏的精度、也不是被简化的信息，而是"生成过程拒绝以任何完成态被提前定价"这一属性本身。

参考文献

Goodhart, C. A. E. (1975). Problems of monetary management: The U.K. experience. Papers in Monetary Economics, Reserve Bank of Australia.

Scott, J. C. (1998). Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed. New Haven: Yale University Press.

修订说明

本文经德麦国际 SDE 学派编辑（Claude）依两轮修订体例打磨：补第九节，与古德哈特定律（Goodhart, 1975）、斯科特可读性（Scott, 1998）正面切割——古德哈特的暴力在代理品被博弈、斯科特的暴力在简化丢失，而'代理的暴力'主张强制转译动作本身即侵害（哪怕代理品完美、信息无损）；给出'完美不可博弈／无损保真的代理品能否消除伤害'的可裁决差别。作者原有代理暴力命名、锁—切—叙事互锁增益、高考评分史案例未改。作者的核心判断、论证结构与原初命名均未改动。

最美文定稿：全球文库终审与核心命题的重写

> 本篇入选「最美文」频道，须过一道比发表更严的关：把参照语料主动往外扩，专门去找「这个想法在哪儿其实早已有了」，找到之后不是绕开，而是与它划一条**可裁决的分离线**——一条让两套理论对同一件事作出*相反预测*的线。扩一个邻近领域就塌掉的分数，本来就是语料收窄刷出来的。以下四节即这道关的记录。

一、最后一位未被请出的对手：哈贝马斯的「生活世界殖民化」

本篇上一轮已经切开了坎贝尔定律、古德哈特定律、布迪厄的象征暴力、鲍尔的审计社会、比斯塔的可测量性、以及斯科特的可读性。剩下的那位最危险的对手不在评价研究里，在社会理论的地基上：哈贝马斯的生活世界殖民化——当货币与权力这两种系统媒介取代交往行动，成为协调生活世界的机制时，生活世界就被殖民了。

用它来复述本篇几乎毫不费力：高利害评价＝系统媒介入侵教育生活世界。若这次复述无损，「代理暴力」就只是殖民化的一个教育学案例。

二、分离线：两套理论在哪里作出相反的预测

分离线钉在：**换掉协调机制能不能解决问题**。

殖民化论断的病灶在*协调机制的替换*——把交往理性换成了货币与权力；因此它的解药是把协调机制换回来：以主体间的商谈、共识、参与去替代科层与指标。代理暴力主张的是另一件事：*强制转译这个动作本身即侵害*，哪怕代理品完美无损、哪怕协调机制完全是商谈性的。

相反预测，可在真实场景中检验：找一个评价体系，其指标完全由师生共同商谈确定、无外部权力与资源挂钩（即协调机制已回到交往理性），但仍要求把学生的生成过程转译为可比较的记录。
殖民化框架预测：侵害消失或大幅减弱——因为系统媒介已被撤走。
本篇预测：侵害不消失——被转译者仍报告那种「我被换成了另一样东西」的经验，且该报告率与转译的强制性共变，而与指标由谁制定无关。

若在完全商谈性的体系里侵害确实消失，代理暴力即应并入殖民化。

三、核心命题的重写：从「X 是 Y」到「X 不是 Y，而是 Z」

原稿命题是：*代理关系本身具有暴力性*。改写为：

> **代理的伤害不是代理品做得不够像，也不是坏的协调机制挤走了好的，而是「必须被转译成另一样东西才算数」这道要求本身——它在最公正的商谈里、由最善意的人执行、用最精确的代理品，仍然照常发生。**

四、本文禁止什么

一个命题的分量，不在它能解释多少，而在它**不许什么发生**。以下三条，任何一条被观察到，本文即失效或需大幅收缩：

1. **禁止用代理品的失真做证据**。失真是精度问题，本篇讲的是转译动作本身；凡能靠「换个更好的指标」缓解的，都不是本篇的现象。
2. **禁止把本篇读成取消评价的主张**。本篇只主张任何评价都必须为「被强制转译者」保留一处不被转译的申诉位，不主张不评价。
3. **禁止把它扩展到自愿转译**。一个人主动把自己的经验写成可比较的记录，不构成代理暴力；强制性是必要条件。

五、独立成立性检验

删光文献后剩下：*把一个人正在发生的东西，强制换成一个可以拿去比的东西——这个换的动作本身就伤人，哪怕换得再准、再民主*。可以被上面那个「完全商谈性的评价体系」推翻。

本轮定稿所核验的文献

Habermas, J. (1981). *Theorie des kommunikativen Handelns*, Bd. 2. Suhrkamp.（生活世界的殖民化，本轮所对质的核心对手）

Illich, I. (1977). *Disabling Professions*. Marion Boyars.（专业化对自主能力的剥夺，用于核对「强制性」条件）

Scott, J. C. (1998). *Seeing Like a State*. Yale University Press.（上一轮已切）

Biesta, G. (2010). *Good Education in an Age of Measurement*. Paradigm Publishers.