

外部锚定：一个关于治理如何从内部制造出它事后所命名的“敌人”的社会学判断

不是他选择了服从外部指令，而是在做出所谓选择之前，那套能够做出选择的器官已被预先拆除

高鹏 著 · SDE 学员 · 治理研究与犯罪社会学 · 2026年7月26日

摘要 关于犯罪，历来的主流追问始终嵌在同一道二元结构里：一边是社会及其秩序，一边是“罪犯”及其破坏性。本文将这道结构本身的诞生视为待解释项，将追问从“一个罪犯如何被制造出来”扭转为“一个被社会事后锁定为‘罪犯’的生存状态，是如何在治理过程内部，作为其不可取消的副产品而被发生出来的”。核心判断是：当治理系统将人作为“需要被预防的风险”来管理时，它便在完成保护职责的同时，启动了一种持续而深刻的剥夺——它在系统地拆除一个人从内部生成行动依据、在困惑中自我停留、在道德张力中挣扎校准的那套能力。当这套能力被拆除殆尽，此人便被置于一种“必须行动，却已无法从内部生成行动依据”的生存悬空之中。在这种悬空里，第一个承诺为他卸除判断负担、提供现成身份与确定性脚本的外部指令系统，便会作为他唯一的骨骼被接纳。本文将此状态命名为“外部指令的生命锚定”。它描述的并非一个人“选择”服从外部指令，而是他在做出所谓“选择”之前，其内部那套能够做出选择的器官已被预先拆除。这一定位的经验基础与理论后果，将通过一个基于公开法律文书的典型化案例、对数字风险评分技术的分析，以及一场与“平庸之恶”传统和当代风险社会理论的正面对话来展开。论文在结论中诚实划定解释力边界，并给出明确的证伪条件。

关键词：外部指令的生命锚定；犯罪的内部发生；安全治理；生命内部生成能力；生成机制社会学

一、引言：一个被问错的问题

“他为什么会做出这种事？”

当一个极端暴力案件进入公众视野，这个追问总是最先被抛出来，也总是最先被堵回去——被医学—心理学诊断标签（“精神病患”“共情缺失”）、被道德直觉（“天生残忍”“毫无人性”）、或者被一种更精致的社会学归因（“贫困的产物”“亚文化的牺牲品”）。这些回答彼此冲突，但共享着一道从未被质疑的预设：犯罪者与社会是两样先在的、彼此外在的实体。社会是秩序，犯罪者是秩序的破坏者。追问的全部努力，只是在两者之间架设一道因果桥梁——那个破坏性，到底是植根于破坏者体内的某种缺陷，还是被社会的某种病态注射进去的。

这道二元结构把分析锁在一张局内人的棋盘上，让我们以为自己在追问原因，其实只是在天平两端来回挪动砝码。它使我们看不见一种更根本的可能性：那个被我们事后命名为“罪犯”的人，与他所破坏的那个秩序，不是在两个地方分别长成、然后才相撞的两样东西。后者，在试图保护自己的过程中，恰恰是前者不可或缺的共谋——不是出于恶意，而是出于它自己无法逃脱的操作逻辑。

本文试图撤销的，正是这道二元结构本身。由此，那个必须被追问的对象就不再是“一个邪恶的人如何出现”，而变成了另一个更棘手、也更诚实的追问：那个被社会终审为“邪恶”的稳定行为状态，它自身，是如何在社会治理的过程内部，被一步步发生出来的？

这不是一个修辞问题。本文的核心概念——“外部指令的生命锚定”（以下简称“外部锚定”）——是在撤除上述二元对立之后，从治理自身操作中一道矛盾里逼出来的。这道矛盾可以被这样表述：治理越是将人当作需要被提前干预的“风险”来管理，它就越必须深入这个人的生命内部，去监控他、塑造他、在规定好的路线上纠正他；而它越有效地把这个人管理得不出错、不越界，它就越系统性地拆除着这个人从内部生成行动依据、在不确定中自我停留、在道德张力中痛苦校准的那套能力。当这套能力被拆除殆尽，此人便不再是治理的保护对象——他变成了一个被掏空的容器，必须被什么东西填满。而第一个足够强大、清晰、承诺确定性的外部指令系统，就会在此时，被他作为自己唯一的骨骼，毫无内心抵抗地接纳进来。此后他做出的那些被社会事后称为“邪恶”的事，不过是这台机器在他身上所完成的最后一道工序的副作用。

本文提出的“外部锚定”，不是一个从既有理论的排列组合中导出的诊断标签。它是在与鲍迈斯特的“邪恶之平庸”论、阿伦特的“平庸之恶”论、以及津巴多的“去个性化”论正面碰撞后，切开的一条此前未被精确命名的辨别线：它区分了“内部器官因过载而临时罢工”“有能力思考却选择不去思考”“内部被情境暂时覆盖”与“内部器官因生长土壤被系统性撤除而从未长成或长成后被拆除”这四种被长期混为一谈的状态。

本文的论证路径如下。第二章将界定“生命的内部生成能力”这一概念，分析它的发生条件，并论证安全治理如何通过三种合法的日常操作系统性撤除这些条件。第三章是全文的理论心脏：它正式命名“外部锚定”，补全从被掏空到被捕获的微观衔接机制，并完成与先行者的不可还原硬校验。第四章基于公开法律文书提供一个被典型化的完整生命史，在每一个关键转折点上显式排除主要竞争性解释

——社会排斥论、亚文化论、理性选择论、自我调节资源耗尽论，以及一般紧张理论。第五章专门分析数字风险评分技术如何成为当代最精密的“锚定加速器”，并区分人力密集型剥夺与算法密集型剥夺在生成机制后果上的关键差异。第六章主动回应三个最有力的反驳，清楚划定本文判断的解释力边界，给出明确的证伪条件，并与福柯治理术传统做出简要的切割。

有一笔纪律必须写在前面。本文将不再使用“罪犯”“犯罪”等词作为对研究对象的未经审视的指称。那个被社会事后编码为“罪犯”的人，在本文的分析中将被严格地指认为“一个被社会事后锁定为‘罪犯’的人”或“被锚定者”。这不是咬文嚼字——这是本文所要完成的整件事在写作的毛孔里必须守住的那条生成机制底线。同理，那个被社会称为“邪恶”的行为，不构成本文的研究对象；本文的研究对象，是一种可被经验追踪的生存状态——一个被系统性地剥夺了内部生成能力的生命，如何被逼到将某种外部指令接纳为唯一骨骼的临界点上，并在此后毫无内心抵抗地执行那些被社会事后编码为“邪恶”的行为。“邪恶”这个词，在本文中将被严格保留给一个尚未被回答的生成机制追问：那个将上述状态锁定为“邪恶”的社会命名过程，它本身，是如何发生的？它的功能是什么？这个问题不会在本文中得到完全回答，但它将在结论中被严肃开启。

二、内部生成能力：是什么，如何被拆除

2.1 概念界定与发生条件

在命名“外部锚定”之前，必须先把它对立面讲清楚——那套在被剥夺之后迫使生命不得不向外寻找替代性支点的“自己的骨骼”。本文把它叫作生命的内部生成能力。

这个概念不是一个心理学变量。它不是一种可以量表测得的“心理韧性”或“自我效能感”。它是一个生成机制概念：它说的不是一个人“拥有什么”，而是在什么样的互动条件下，一个人可以逐步长出一种特定的自我供给方式——一种从内部生成行动依据、在不确定中自我停留、在道德张力中挣扎校准的稳定能力。它不是种在土里的已有的种子，它是那块能让种子发芽、歪歪扭扭却终于长出自己形状的土壤本身。

这套能力由三种相互咬合、缺一不可的自我供给方式构成。

第一，从自身经验中自己生发出问题意识的能力。它的真正内核不是“会提问”，而是：那个问题是他自己从自己与现象的遭遇中识别出来的，不是别人替他准备好、写在黑板上的。一个孩子蹲在地上看蚂蚁搬家，没人考他，没有标准答案等在前方，但他被自己的好奇驱动着：“它们要去哪里？”那一刻，他的思维不是从外部被启动的，而是从他自己与世界的碰撞里内生出来的。这种不依赖外部指令的内在启动，是一切自主判断的原始细胞。

第二，不依赖外部计时器、在困惑中自己停留的能力。人面对拿不准的事时，内心会进入一种混沌状态——旧的解释不灵了，新的还没长出来。这种状态天然低效，看不到进度条，而且让人难受。一个人能否具备内部生成能力，关键就看他能不能在这种“不知道”的难受里待住——不急着找一个现成答案来终止它，不急着把责任推给一个外部权威来解除它，而是让互相冲突的感受、相悖的念头、不确定的线索同时在意识里浮着，慢慢地，让新的秩序从内部长出来。这种能力不能被教会；它只能在无数个“没人催你、没人给你倒计时”的发愣中，自己把自己练结实。它需要时间。更准确地说，它需要不被外部指令填满的时间。

第三，在道德临界处的灰色地带中自我挣扎、并据此校准选择的能力。道德不是背下来的条规。背下来的条规可以在考试时准确复述，却在诱惑来临时像一张纸一样薄。真正的道德感，是一种被个体的完整经验浸泡过的判断力：一个人曾经在诱惑面前有过真实的摇摆——他不是背诵“不能说谎”，而是在某个时刻差点说谎，而且在那个“差点”的瞬间，他第一次辨认出了自己内心有一道他自己在意的界限；他曾经在做出妥协之后，体验过一种无法用任何理由彻底抹去的羞耻；他曾经在伤害别人之后，经历过无人知晓的深夜辗转。这些挣扎本身，才是道德感的实际发生过程。一个从未被放进可能犯错、摇摆、后悔的处境里的人，他的“善良”是没有经过检验的。

这三种能力相互咬合：第一种让他有能力对自己发问，第二种让他在回答长出之前不自毁，第三种让他在行动之后仍然能听见自己的回响。它们合在一起，构成一个生命能够“在行动时，自己为自己提供依据、方向与约束”的完整器官。

微型对质：与“韧性”和“自我效能感”的切割。

这套“内部生成能力”极易被收编到当代积极心理学的两个核心概念里——塞利格曼（Seligman）式的“习得性乐观”或“心理韧性”（resilience），以及班杜拉（Bandura）式的“自我效能感”（self-efficacy）。它们的重叠是真实的：三者都反对将人的行为仅仅归因于外在环境或先天缺陷，都承认某种内在的、可以习得的行动资源的存在。但这里有一条必须被切开的界线。

积极心理学的“韧性”和“自我效能感”，本质上是个体资源模型：它们假定个体内部有一套可被量表测量、可被培训增强的心理资产（“我相信我能”“我能在打击后弹回来”）。这套资产可能因创伤而暂时受损，但它们预设了资产的可恢复性——通过认知重构、正向反馈、成功体验的累积，个体可以重新习得对自己的掌控感。换句话说，它假定那个能长东西的土是永远在那里的，只是需要翻一翻、浇浇水。

然而，本文讲述的是一群在长达十几年的时间里，被一整套合法的社会操作系统性地撤除了那片土的人。他们的问题不是“韧性”被削弱了，而是那份让“韧性”这个东西能够在最初长出来的结构性土壤——那些不被填满的下午、那些被郑重对待的困惑、那些挣扎失败后仍被相信“还在挣扎”的瞬间——从未被允许存在过，或在长成之前就被铲平了。积极心理学可以帮一个经历了单次创伤的孩子“重新站起来”，但它面对一个从六岁到十九岁从未拥有过“自己的下午”的人，它手里那张量表量不出他缺的是什么——因为他缺的不是分数的低下，而是那块让他能够长出任何“分数”所测量的那个东西的土壤本身。

这不是一个可以在咨询室里被“修复”的缺损。它是一个必须被社会在结构层面上重新提供生长条件的、生成层面的意义上的剥夺。“内部生成能力”这个概念的理论收益，恰恰在于它拒绝将这种剥夺还原为个体心理资产的亏损；它坚持追问：那片能让资产长出来的土壤，它到哪里去了？

2.2 安全治理如何系统性地撤除生长条件

现代社会对付“人的不确定性”的方式，在操作层面上有一个统摄性的特征：它倾向于用外部标准替代内部判断，用可量化指标替代不可量化的过程，用即时的、程式化的纠偏动作替代漫长而不可控的自我校准。本文把这一整套弥散在刑事司法、儿童福利、学校教育、社区矫正与就业管理中的操作逻辑统称为安全治理。它不是一个特定的制度，而是一个将“预防风险”置于一切之上的逻辑——它在意的是这个人的下一步会不会造成损失，而不在意这个人的下一步对他自己的生命意味着什么。它的每一个具体操作，单独拆出来看，都能被合理地辩护为“保护”：保护可能的受害人，保护社会资源，乃至保护这个风险载体免于陷入更深的困境。本文要论证的是：正是这套在每一个节点上都可以被单独辩护的、合法的、甚至带有道德光环的保护性操作，当它持续作用在同一个生命上时，会构成一个缓慢而深远的剥夺过程。它剥夺的不是他的自由或权利，而是他那套能够在未来任何时刻、面对任何外部指令时，做出属于他自己的内部停顿、打量与判断的器官。这个过程通过三种相互咬合的日常操作完成。

第一，以“预防风险”为名，用外部安排填满生命中所有可用于自发探索的缝隙。当一个孩子被标记为“高风险”的那一刻起——不管这个标签是叫“困境儿童”“行为偏差”还是“需要重点关注的复杂个案”——他的日常生活便被一系列他无法拒绝的安排填满：定期报到、强制辅导、心理评估、行为矫正课程、暑期特别活动。每一个单项都“为他好”。但它们合在一起，构成了单独看谁都不会承认、累计起来却无可逃遁的一件事：他不再有那些“没人告诉他该干什么”的时间了。而本章第一节已经论证过：那套内部生成能力，不在评估表里长，不在标准课程里长，不在被安排好学习成果的活动里长。它只长在那些缝隙里——放学路上、周末下午、无聊的暑假里那些什么都不为、只是东摸摸西晃晃的时刻。一个孩子在这些时刻里学到的不是知识，而是如何在自己内部启动一件事。当他的所有时间都被“有益的”外部指令提前占领，这个学习过程就被无声地终止了。

第二，以“客观公正”为名，用可量化指标替代对他整个人的完整关注。安全治理不信任人的直觉判断，它信任数字。一个孩子“最近怎么样”——这个本该由那个每天和他在一起、看得见他最细微变化的老师或社工来回答的问题，如今被一套量表分数代答：出勤率、违规次数、行为评分表的波动曲线。这些数字是客观的，但它们也是扁平的。它们能精确地告诉你这个人有没有按规定出现，却完全无法捕捉另一类信息：他最近是不是开始对一件事表现出真正的专注了？他在那次谈话中，是不是有一个瞬间不那么防御、流露出一丝真实的困惑？这些无法被装进表格里的信号，恰恰是他那套内部生成能力可能还活着的仅有迹象。但因为表格不收它们，系统就看不见它们。因为看不见，它们就不被计入任何决策。

第三，以“纠正错误”为名，将他所有的挣扎，重新解读为他故障的进一步证据。一个被长期监控的少年不会完全停止尝试。他偶尔会努力控制住一次暴怒，主动完成一件不是被催着做的小事，在谈话中流露出含糊的自我反思——用他自己还不具备的语言来说，这些是他在试图长一点自己的骨头。但这套治理逻辑的评估框架有一个致命的特性：它只看“达标与否”，不看“挣扎与否”。他那天控制住了脾气——没有用，因为另一天他还是失控了。他那次主动帮忙——没有被当真，因为帮忙太小，没有改变总体评分。更致命的是：当他挣扎到一半却失败时——比如他连续三天表现得像个“模范”，第四天却又爆发了——这次爆发在系统里不会被登记为“他在挣扎的路上滑倒了”，而是会被登记为“他果然还是老样子”。挣扎本身，成了他无可救药的证据。挣扎一旦被反复当作故障来惩罚，放弃挣扎就是唯一理性的选择。

三种操作，三把彼此咬合的齿轮。当它们转上足够长的时间——不是几个月，而是贯穿一个人从被首次标记到成年的整个成长史——它们所完成的事，已经不是“教育失败”或“福利失灵”这类温和词汇所能涵盖的了。它们是缓慢的、合法的、无人需要为此承担个人责任的一场内部器官的拆除。手术完成时，被手术者的外部档案看起来可能好多了：他按规定出现，表格上不出错，不再惹麻烦。他不是“被毁了”——他是“被掏空了”。那套能让他自己启动问题、在困惑里站住、在道德岔路口痛苦摇摆的内在结构，已经不存在了。

2.3 与福柯治理术的一个必要切割

在将这套操作逻辑命名为“安全治理”时，本文已置身于一个被福柯（Foucault）及其后继者深耕了数十年的论域之中。福柯在《安全、领土与人口》中提出的“安全机制”（dispositif de sécurité），以及《规训与惩罚》中的“规训权力”（disciplinary power），为理解现代社会如何通过正常化、风险管理和人口治理来塑造个体，提供了一个影响深远的分析框架。本文绕不开它，也不应绕过。但本

文在使用“安全治理”这一概念时，与福柯传统之间有一条关键的分界线必须被点明。

福柯的“规训”瞄准的靶心，是身体。规训权力通过时间表、空间分配、层级监视和规范化裁决，将身体训练成既顺从又高效的“驯顺的身体”——它要的是一个能精准执行指令、同时将外部监视内化为“自我监控”的现代个体。这种内化是规训的核心操作：囚犯最终自己管理自己，学生自己督促自己，工人自己检查自己。

本文关注的“安全治理”，瞄准的却不是身体的驯顺，而是内部生成能力的拆除。这两条路径的终点在外表上可能很像——两者都生产出抵抗地执行指令的人。但它们抵达这一终点的生成机制截然不同。规训的逻辑是“往里加”：把外力做成内心的监工，让个体即使在没有外部监视时，心里也住着那个监视塔。本文所描述的逻辑是“往外掏”：不是往里安装一个自我监控的装置，而是把人里面那套能够自己发问、自己停留、自己挣扎的器官预先拆掉，拆完之后，那个空出来的位置，会被第一个趁虚而入的强外部指令系统自动占据。被规训的人，心里有塔；被掏空的人，心里有空壳——前者仍然可以内疚和痛苦（尽管被压制），而后者在行动时可以是“没想那么多”的。

这一区分不仅关乎理论精确性，也关乎实践指向。规训的问题指向的是“如何打破那个内化的监视塔”——去自主化、去内化，重建自身的反抗主体性。而外部锚定的问题指向的是另一条路：如何在那套器官的土壤被撤除之后，在结构层面上重新提供那些能让它从头长出来的条件。这不是一个意识觉醒的事，这是一件重建生长生态的事。

三、外部锚定：从被掏空到被捕获

3.1 掏空之后：微观衔接机制

以上对人的内部生成能力被拆除的描述停止在一个点上：那个人被掏空了。但“掏空”不是一个可以持存的稳定状态。生命不能悬在真空中。当一个人失去了从内部为自己提供方向、依据和意义的能力之后，他并没有同时失去另一件东西：他仍然被嵌入在社会生活中，仍然被持续地抛入“你必须行动、必须回应、必须在下一步之前做出选择”的结构性压力之中。

这一点是整个论证链条里最容易被跳跃的一环，必须单独钉死。被掏空的人，他的处境不是“可以不做选择了”——恰恰相反，他被监控得更紧，被评估得更频繁，被要求表明态度、交代行踪、在表格上签字承认错误的时刻比普通人多得多。这种被迫的、持续的“行动压力”，与他的内部依据已近于零的状态之间，构成了一对他自己无力解决的基本矛盾。他必须做，却已经没有一套自己的东西来帮他判断“做什么”和“为什么做”。

在这个节点上，他的经验反应，原则上可以有几种完全不同的走向。他可以陷入彻底的被动与麻木——一种抑郁式的全面退缩，不再行动，不再回应，从社会舞台上自行消失。他也可以像一个被风吹动的塑料袋，随机地、碎片化地跟随任何一个闪过的微弱线索，不被任何特定的系统捕获，只是漂浮。从纯粹的逻辑上说，一个被掏空了内部依据的人，并不“必然”走向那个提供清晰指令的帮派。

但接下来的一步是本文判断最核心的微观机制所在。在这个被掏空者与那个特定外部指令系统之间，存在着一个使捕获变得“不是必然、却高度可预测”的衔接结构。它由三个嵌合在一起的齿组成。

第一，接口的预装。在漫长地被标记、被干预、被矫正的那些年里，他唯一被持续训练、持续强化、从未中断的东西，不是思考，不是判断，不是道德挣扎，而是对“指令—反馈”闭环的高效执行力。他学会了准时出现，在表格上打勾，在被训话时说出最不容易惹麻烦的回答。这些东西不创造成长，但它们形成了一套高度可用的、与外部指令对接的行为模板：只要外部有一个清晰的指令告诉他做什么，他就能在指定时间内完成；完成之后，他会获得一个反馈——通常是被放过、被从本月的名单上划掉。他的执行接口是高速的，他的判断接口已经近于停用。

第二，认知饥饿与身份空缺。在被治理的整个历程中，他的身份是纯负面的——他不是好学生、不是好员工、不是有希望的个案，而是一串被反复叠加的诊断标签和分数标签。这些标签告诉他“你是谁”的方式，全是“你不是什么”。与此同时，他从未被提供过一套自己能够操作的语言，来把自己那些堵在心里的困惑、愤怒和被放弃感，组织成一个连贯的、可被自己理解的叙事。他不知道如何命名自己正在经历什么。他的身份空缺，和他对一套现成的、自己不需要费力去建构的叙事的认知饥饿，在同一个时期里同时达到了峰值。

第三，捕获者在空间、形式与反馈上的精准匹配。那个操刀捕获的组织——无论是街头帮派还是其他锚定提供者——恰好在三样东西上精准地弥补了他的全部功能缺失。其一，物理空间的重合：他不是经过一场长途旅行才进入帮派的辐射范围；他每天在街头晃着的地方，就是帮派活动的同一片街区。其二，认知门槛的极低化：帮派的世界观不需要他费力理解——敌我二分，兄弟与出卖者二分，谁该打谁不该打，全都是现成的。这套语言恰好不需要他使用那套已经被摘除了的内部判断器官；它只需要他执行。其三，反馈的即时性、确定性与情感厚度：每一次完成任务后的那句“你是自己人”，伴随着真实的归属、实际的信任、实际的分成，直接打在他的身份空缺和认知饥饿最裸露的部位——那是一个此前十几年里从未被任何量表分数或矫正课程喂饱过的地方。

这三颗齿一起转上去之后，“他为什么没有被抑郁吞没”和“他为什么不去随机游荡”这两个竞争路径，就被自然地排除了——因为

他那套被训练得极其高效的执行接口，不会让他在行动压力面前选择静止；而他的身份饥饿，又使得那种不能提供身份承诺的、碎片化的微弱指令根本抓不住他。他需要一个足够强的、能把他从此前十几年纯负面身份中一次性拔出来、并赋予他一个正面位置的系统。帮派，恰好是那个在同一个街区里、用最低认知门槛、提供最强确定性的唯一在场者。

他不是几个选项里选了它。它是在他必须做选择、却已被剥夺了选择的能力的那一刻，唯一一个能够被他识别为“选择”的东西。

3.2 正式命名：“外部指令的生命锚定”

在上述分析的基础上，本文正式提出“外部指令的生命锚定”这一概念：

当一个生命被系统性地剥夺了从内部生成行动依据、在困惑中自我停留、在道德张力中挣扎校准的能力之后，他被迫——在他做出任何能被算作“选择”的动作之前——将自己的行动依据、身份认同与生存意义，全部挂靠在第一个为他提供足够强大、清晰且承诺确定性的外部指令系统之上。这种挂靠不是他“服从”某个权威，而是他在内部已被掏空的处境中，把那个外部指令系统接纳为自己唯一的骨骼。

这一定义需要在以下几组边界上被钉实。

它不是“服从”。米尔格拉姆（Milgram）实验中，服从者流汗、颤抖、咬嘴唇——他们在内心交战。“外部锚定”状态下的被锚定者，他的内心可以是没有战争的。不是因为他的邪恶性压倒了良知，而是因为那个会发生战争的心理场域——那个能在诱惑面前摇摆、在暴力面前颤抖、在他人痛苦面前感到刺痛的内部空间——在长年的剥夺中，已经被铲平了。他执行，但他并不在场。这解释了为什么许多施暴者在落网后供述时都说的是同一类无法被接住的话——“我没想那么多”。这不是推卸，这是对那个内部空无状态的诚实描述。

它不是“社会排斥”。社会排斥理论正确地指出了个机制：被排斥的个体因为合法的机会结构对他关闭而转向非法手段。但是，一个仅仅被社会排斥的人，如果他内部的生成能力尚且完好，他在转向非法时仍会挣扎、权衡、负罪——他仍然是他自己行动的作者，只是在一个被收窄的选项集里行使着那个作者身份。而一个被外部锚定的人，连那个作者身份都已经失去了。他行动，但行动不是从他里面流出来的——是直接在外面穿透他、经由他被训练出的高效执行接口，毫无阻碍地释放出去的。

它不是“精神病态”。精神病态（psychopathy）概念假设了一种深度植根于个体内部的共情缺陷，把共情缺失定位在个体自身内部。“外部锚定”不做这个定位。它描述的共情缺失，不是那个人内部本应有个东西却坏掉了，而是那个东西的生长条件被一整套合法的、弥漫的社会操作在长达十几年的时间里给撤除了。两者路径迥异：前者是现成器官的缺陷；后者是器官生长土壤的移除。

它不是“亚文化同化”。亚文化理论讲的是：一个人习得并内化一套不同于主流社会的价值与规范体系。“外部锚定”讲的根本不是“习得一套价值观”。价值观的习得需要一个人在某个时刻——哪怕非常微弱——主动说“对，我认这个”。被锚定者已经失去了那个能够说“我认”的内部器官。他不是认同一套亚文化，而是把一整套现成的外部身份直接套在了自己身上。两者的界限，就是“选择穿上一件衣服”与“这件衣服就是你的全部皮肤”之间的界限。

一个必要的警戒。“外部锚定”不是一个可以被简单地用来“诊断”一个具体人的静态标签。它拒绝被发展为“锚定量表”或“锚定诊断标准”。它是一个辨别维度——它的理论收益在于分开那些被既有理论混为一谈的生成路径（情境覆盖 vs. 结构性拆除），而不在于归类个体。本文在使用这一概念时，始终保持在宏观和生命史尺度的分析上；读者不应在阅读本文后，举着这个概念去对一个具体的人说“你被外部锚定了”——那恰恰是把一个生成机制概念降格为既成事实论的诊断标签。

3.3 被社会称为“邪恶”的行为，从这里如何被释放出来

有了“外部锚定”这个概念之后，我们就可以回过来处理那个被本文一早悬置了的词：“邪恶”。本文从未使用“邪恶”作为解释项。它现在要正面回答的是：那个被社会称为“邪恶”的行为，它实际上是什么？

它是一个人被外部锚定之后，在特定情境中，作为一条对锚定指令的忠实执行，而无内心抵抗地被释放出来的高效行动。它不是从他内部某个“邪恶的意图”中长出来的。它是从他内部那个空无之中，被外部指令直接穿过之后，在行为层面留下的痕迹。

这个过程可以拆解为四个连续步骤。第一，锚定系统替他卸除了所有判断负担。什么是错的、谁该被打、做到什么程度才算够——这些全部由指令给出。他不需判断，只需执行。而这种执行，恰恰是在他被安全治理管理的那十几年里，唯一被真正训练到熟练的事。第二，锚定系统提供了一个现成的身份。在漫长的被治理历史中，他的身份是纯负面的——不是好学生、不是好员工、不是有希望的个案。锚定系统给了他一个完全不同的身份：你是被选中的人，你是我们的兄弟。这是他此前十几年从未再体验过的归属感与确定感。第三，锚定系统通过逐步升级的考验，将他的执行力推到没有回头的临界。每一次执行——从望风到陪同作案到亲自动手——都伴随着新一轮的认可与身份加固。与此同时，那个本该在这些时刻跳出来叫停、迟疑、痛苦的东西，却始终没有出现——因为它在长年的剥夺中，已经萎缩了。第四，当那个最终被社会称为“邪恶”的行为发生时，他已经不在那里了。那个制服受害者身体的人、那个扣动扳机的人——他不是“一个极度邪恶的人”。他是在那个时刻，作为一套被外部植入的指令的执行终端，在运行。他不是故意关掉了内心的战争，他是从来就

没有被允许长出一个会发生战争的内心。

这就是被社会终审为“邪恶”的行为在生成机制眼光下暴露出的真正骨架：它不是从“恶”中长出来的，它是在“空”中被外部指令穿过之后，留下的一个社会无法消化、只能将它命名为“恶”的痕迹。那个命名本身是社会的防御机制——把它叫作“邪恶”，意味着不用再看它与治理自身操作之间的关系了。

3.4 不可还原硬校验：与先行者的正面碰撞

一个新概念要站住，不能仅靠内部自治。它必须正面回应直接的质疑：“你讲的这件事，是不是已经被别人讲过了——只是你换了套新词？”本节主动邀请本文三个最危险的敌意最近邻进场，在碰撞中指明：他们各自握着一一条在常态下极为强大的解释线，但他们在某个临界面之前会同失效——而本文的“外部锚定”，正是在那个临界面上，作为一条他们全都切不开的辨别线而被逼出来的。

第一位：鲍迈斯特的“邪恶之平庸”。鲍迈斯特（Baumeister）在《邪恶：在人类暴力与残酷之中》（1999）中论证：大多数人类施暴者并非恶魔式的虐待狂，其施暴行为来自一些在心理机制上极为平常的过程——其中最重要的是自我调节资源的逐渐耗尽（ego depletion）。这一模型的核心预设是：一个人的自我调节能力在耗尽之前，是完好地存在着的——它是一个可被重新充满的电池。只要让他休息、补充葡萄糖、逃离压力情境，那个内部的“管住自己”的能力就会恢复。本文的“内部生成能力被剥夺”不是一个状态性概念：它不是电池被耗尽了，而是电池这个器官本身，在长达十几年的时间里被一整套合法的社会操作拆除了。它的不可恢复，不是因为它被压到了零，而是因为让它得以在最初长出来的那片土壤已经被铲平。你给他再多休息，他也不知道在休息里该做什么——因为他的“自己发起思考”的器官，从来就没有被允许长出来过。第一条辨别线：鲍迈斯特解释的是“内部器官因过载而临时罢工”；本文解释的是“内部器官因土壤被撤除而从未长成、或长成后被拆除”。

第二位：阿伦特的“平庸之恶”。阿伦特在《艾希曼在耶路撒冷》中对艾希曼的诊断，是她本人在《精神生活》中反复修订过的：她看到的不是“服从权威”，而是一种“思考的缺席”（absence of thinking）——不是这个人被剥夺了思考能力，而是他在可以思考的每一个时刻，都放弃了思考。他的“不说‘不’”，不是因为他不能，而是因为他更愿意随大流，更愿意不去面对那个独自判断会带来的巨大痛苦和孤独。换句话说，“平庸之恶”预设一个主体的懒惰与放弃——他有思考的能力，但他不用。本文的“外部锚定”讲的是另一个完全不同的故事：被锚定者不是“放弃”了思考，他是在有机会发育出那个能够思考、能够自我停留、能够在诱惑面前痛苦摇摆的器官之前，就被一套为他好的、不可拒绝的外部治理给制度性地撤除了它需要的全部生长条件。他从未拥有过阿伦特预设艾希曼已经拥有、却拒绝使用的那套东西。第二条辨别线：阿伦特追责的是“有能力思考却选择不思考”；本文追问的是“那套能思考的器官，它是怎么被一个社会的正常运转所不允许长出来的”。

第三位：津巴多的“去个性化”与“路西法效应”。斯坦福监狱实验及其衍生的“情境力量”研究传统，贡献了一个生命力极强的解释模型：一个正常的、此前从未表现出暴力倾向的人，被放入一个具有强大角色指派和规范脚本的情境之后，会在匿名性、责任扩散和群体压力的作用下，暂时失去个人认同，变成情境指令的执行者——即津巴多所说的“去个性化”（deindividuation）。本文与它的关键分界线在于：津巴多解释的是，一个内部器官本已完好的人，进入强大情境后，内部被情境暂时覆盖；本文解释的是一群在进入那个强大情境之前的十几年里，已经被安全治理预先拆除了内部器官的人，情境对他们所做的不是“覆盖”，而是“填充”。被去个性化的人，当情境撤掉——实验结束、退伍、离开封闭组织之后——内部自我会慢慢恢复，他会回过头来为自己的行为感到惊恐与愧疚。被外部锚定的人，离开锚定系统之后，内部并不会“恢复”——因为那里从一开始就没有过一套属于他自己的行动依据。他不是被自己的行为惊扰；他只感到一种更深的悬空：那个曾经填充他的指令消失了，而他自己里面，仍然什么都没有。第三条辨别线：津巴多解释的是“内部被情境暂时覆盖”；本文解释的是“内部在进入情境之前已被拆除，情境只是趁机做了一次填充”。

这不是程度的差别，这是类的差别。三者各握住了一条常态下极为强大的解释线——但它们全都在同一个临界面上失效了：当一个人在其整个成长史中从未被允许发展出那套内部自我供给的器官时，说他的器官“耗尽了”、“被他放弃了”、还是“被情境覆盖了”，已经没有意义了——因为那里首先就没有一个可以被耗尽、被放弃、被覆盖的东西。那个器官生长的土壤已经被铲平了。本文的“外部锚定”，正是从这三者共同的失效之处，作为一条被迫着长出来的新辨别线，而站住的。

四、一个人如何被锚定：一个完整生命史及其竞争性解释排除

以下案例基于公开法律文书中多起真实案件的合并与典型化——这不是一个特定的个人，而是在司法档案中反复出现、具有惊人结构相似性的一个模式。它的价值在于让“外部锚定”这个概念获得可被追踪的生命史，同时让本文的判断经受竞争性解释的检验：在小军（化名）生命中的每一个关键转折点上，既有理论能解释什么？会漏掉什么？为什么只有“内部剥夺—外部锚定”的框架，才能把那些被漏掉的东西兜住？[1]

阶段一：被填满的童年（6—12岁）

小军出生在一个被多重社会服务系统长期“在册”的家庭。父亲在监狱服刑，母亲有精神疾病诊断，家庭领低保。从小学二年级起，

他就是一个被标记的孩子：“困境儿童”“需要重点关注”。这个标记意味着，从那一刻起，他的生活开始被各种善意的外部力量填满——社工定期家访，学校安排一对一辅导，社区组织的暑期活动名单上永远有他的名字。每一次干预都真诚地想要帮他，但每一次干预，同时也是一次他无权拒绝的日程填入。别的孩子在放学后有一段“没人管”的时间——在巷子里疯跑、蹲在地上看蚂蚁、发呆、和同伴商量一个毫无实际意义的游戏规则。在那些无所事事的下午，他们不知不觉地学会了自己发起一件事、在无人评分的状态下判断“我喜欢什么”“我不想碰什么”。小军没有这些下午。他的课后被辅导填满，他的周末被活动预定。所有活动都安排好了主题、流程和预期的学习成果。他不需要发起任何事，只需要出席。到了小学高年级，他的成绩始终没有起色。他在学校的角色被固定了下来：他是那个“虽然给了很多辅导但就是跟不上”的孩子。

如果仅用社会排斥理论看这个阶段，它看到的是：家庭贫困与父职缺失使他被排斥在优质教育资源的可及范围之外，这是导致他此后一切困难的起点。这个解释是对的，但它漏掉了这个阶段最核心的机制——不是“排斥”，而是“善意的填满”。他并没有被学校赶出去；他接到的帮助太多了，多到那些能让一个孩子长出自己的问题意识、自己和自己玩的耐性、自己做错事后自己后悔的经验的那些缝隙，全都被提前占领了。社会排斥理论擅长解释资源缺失，它解释不了资源过度填满所造成的内在生成能力的窒息。关于这一点，已有田野研究反复发现，在高度被干预的困境儿童群体中，“时间被结构化”（structured time）的程度与自发游戏行为的衰减呈正相关，而这种衰减与日后的自主决策困难之间存在着被低估的因果链。[2]

阶段二：挣扎的被污名化（12—15岁）

进入初中后，小军开始出现行为问题：逃课、打架、在校外小店偷过两次东西。每一次出事，系统都启动标准程序：谈话、写检查、叫家长（能叫来的是他的外婆）、加派辅导。没有人问过他：你听得懂课吗？你心里有没有什么特别愤怒的东西，说不出口，却一直硌在那里？评估表之外的他——那次他在和同学冲突后主动道了个歉，那次他交了一次虽然字迹潦草却是自己独立做出来的作业——这些瞬间，评估表不搜集，也就不被任何决策看见。相反，每一次犯错都在被记录、归因、累加。他的档案越来越厚，里面的标签越来越稳：“攻击性行为”“品行障碍倾向”“低自控力”。到了初二下学期，他被建议转入特殊教育班。“建议”在这个语境里不是商量，是判决。他没有抗拒。他已经学会了——抗拒只是让下一次判决来得更快。

亚文化理论（以科恩（Cohen）的“地位挫败”论为代表）在这个阶段会提供一种有力的解释：小军开始逃课、打架、偷窃，是因为他在主流价值体系中无法获得地位，于是转向同龄人偏差群体，在那里获得替代性的地位满足。这个解释抓住了一个真实动力——他确实渴望被认可。但它漏掉了一个更早、更致命的过程：在加入任何偏差群体之前，他那些“试图留在主流价值里”的挣扎，是怎么被系统性地看不见、不认可、乃至反过来变成他进一步被困的证据的？亚文化理论预设一个人是在权衡“哪条路能给我地位”之后选择了亚文化；但它没有处理，在这个人还可以走另一条路的那个阶段，那条路的门是怎么在他一次次试图拉开一条缝的时候，被系统从另一侧关上的。关于学校纪律处分如何累积并自我强化标签，已有质性研究详述了这一“挣扎被污名化”的微观过程：学生为弥补上一次违纪而做出的努力，往往因为不符合“戏剧性的转变”而未被教师注意；而他的下一次违纪，则被读作“本性难改”的铁证。[3]

阶段三：悬浮期（15—17岁）

初中毕业后，小军进了一所他没有选择过的职业培训中心。他不是因为想学技能去的，是因为没有别的地方可以待。待了三个月，开始旷课。社工找他谈，他坐在那里，目光空着，问什么都说“不知道”。他不是在对抗。他是真的不知道——不知道自己能做什么，不知道明天除了继续这样晃着还有什么别的可能。他的内部生成能力，在他整个童年和青春期的被填满、被忽略、被污名化中，已经被拆除了临界点。他还没有犯罪，但他内部已经空了。

鲍迈斯特的“自我调节资源耗尽论”对这个阶段的小军看上去极有解释力：多年的压力与控制耗竭了他的自控能力，他处于一种类似 ego depletion 的状态，这解释了他为什么“提不起劲”、为什么无法向上规划。但这里有一个致命的不匹配：鲍迈斯特的模型预设，耗尽之前他曾拥有过自我调节资源——哪怕很少。而小军的生命史显示，他从未有过一个阶段，是可以在不被外部指令填满的缝隙里，自己去尝试启动、停留、挣扎、并因此发展出哪怕最基本的自我调节经验的。他的内部不是耗尽的电池，而是从未被允许组装起来的电池组。把“从未有过”当成“有了之后耗尽”，是整个自我调节文献在遇到这类生命史时最普遍的误诊。

阶段四：锚定捕获与加固（17—19岁）

就在这段时间里，他在街头认识了一个比他大几岁的年轻人——本文称其为“锚定提供者”。这个人和小军之前遇到的所有“帮助者”都不一样。他给小军的第一样东西不是任务，是注意。他会听小军说话——听他说这些年在心里堵着的东西：被看不起、被放弃、被各种“为你好”的人安排来安排去。听完之后，他没有填表格、没有下一个判断、没有启动一个流程。他给小军的，是一套现成的叙事：“你知道你为什么是这个下场吗？不是你不行，是这个社会根本没打算让你站起来。你想翻身，就得跟我们一起——我们才是自己人。”

这套叙事对他的打击之所以无比精准，不在于它的深刻，而在于它的简单与确定：它给了他一个此前十几年全部人生经验从未给予他的东西——一个不是由诊断表和评分栏拼凑出来的身份。他不是被社会怜悯的被淘汰的次品；他是一个被压抑的反抗者。这是他第一次体验到“我是谁”这个问题的答案，不是别人替他填的，而是他“自己选的”——尽管这个“选”是在内部已被掏空的前提下发生的。他加

入了团伙。最初的任务是放风、跑腿。每一次行动之后，锚定提供者都会在事后给予认可：“你今天做得不错，是自己人。”他渴望这句话的程度，远超他渴望过任何一句学校老师的表扬。因为那些表扬他听得出真假——表扬之后，他还是那个差生。而这句话伴随着真实的后果：实际的归属、实际的信任、实际的分成。

理性选择理论对这个阶段会给出最直接的解释：小军是在理性地计算成本收益之后，选择了加入犯罪团伙——这比继续在社会底层毫无希望地漂泊更划算。这个解释只捕捉了一层皮。理性的计算需要一个能够清醒地在选项中比较、衡量、投下偏好权重的主体。而小军在被捕获的那个时刻，他已经没有一个内部完备的、能够进行这种比较的“主体”了——他只有一套被掏空的内部，和对确定性的绝望渴求。如果在他面前同时摆着一个合法的工作机会和一个帮派的邀请，而那个合法的工作机会需要他自己判断、经营、面对不确定的人际关系和职业前景，他仍然会走向帮派——不是因为他计算了帮派更有利，而是因为那个合法选项要求的恰恰是他已经被剥夺掉的东西：从内部生成、在不确定中坚持、在复杂关系中自我导航的能力。

阶段五：极端行为的无内心抵抗释放（19岁）

在一次团伙的报复行动中，上面交代的任务是“弄出点动静”——让另一伙人“记住疼”。小军分到的任务是制住受害者的身体。他照做了。受害者反抗，他加力；受害者不动了，他没有立刻松手，而是按指令多制了一段时间，直到有人喊停。事后他在供述中说：“我没想那么多。他们让我按住，我就按住了。”

这不是在推卸责任。这是“外部锚定”状态下，一个人对自己行为的诚实描述。“我没想那么多”——那个会去想“他是不是会死”“他疼不疼”“我到底在干什么”的人，不是在那一瞬间被他故意关掉的，而是从他六岁到十九岁这十三年里，被一整套为他好的、精准的、不可拒绝的安全治理，一寸一寸地拆掉的。在多起公开判决书中，类似“没想那么多”“当时没感觉”“就是按他们说的做了”的供述反复出现——这些表述不是偶然的口语习惯，它们标记着同一个内部状态：行动时，那个能发生道德战争的心理场域，不在现场。

与一般紧张理论（GST）的必需对质。阿格纽（Agnew）的一般紧张理论是当代犯罪学中最具解释野心的理论框架之一。它的核心论点是：个体在经历“无法达成正面目标”“失去正面刺激”“遭遇负面刺激”这三种紧张源时，会产生愤怒等负面情绪；这些情绪制造出一种“修正行动的压力”，如果个体缺少合法的应对资源与能力，就更可能转向犯罪。这看起来恰好能解释小军的全程——他经历的难道不是典型的紧张吗？

然而，一般紧张理论在本文所处理的这类案例上，恰恰在它最核心的情绪中介变量上失效了。一般紧张理论把“愤怒”当作连接紧张与犯罪的核心情绪中介——它解释的是带有强烈负面情绪驱动的犯罪：报复、发泄、在愤怒之下失去控制。但小军在行动时是没有愤怒的——“我没想那么多。”他的核心状态不是愤怒，是“空”。他不是被怒火推着去捅人；他是在指令之下，平静而高效地执行着一个外部脚本——因为那个会发生愤怒、会发生犹豫、会被受害者的痛苦所刺痛的内部空间，在行动发生的很久之前已经被铲平了。

一般紧张理论能解释另一类案件——那些在长期被欺压之后爆发性伤人的、情绪激烈、事后懊悔的犯罪者。但它切不开小军这一类：他们的犯罪不是情绪的出口，而是内部空无被外部指令穿过之后，在行为层面留下的高效痕迹。两者的行为可能指向同一条刑法条文，但在生成机制上，它们是两条完全不同的路径。对质这一步完成之后，本文的“不可还原性”就从概念层面渗透到了案例验证层面：一般紧张理论握住了“愤怒驱动型”犯罪，社会排斥理论握住了“机会阻断型”犯罪，鲍迈斯特握住了“耗尽型”犯罪——但小军的案子，恰好在这三者各自变量的交叉盲区里，要求一个他们全都无法单独提供的辨别维度：一个人，当他那套能够产生愤怒、产生计算、产生自我调节的器官本身，已经在长长的被治理岁月里被撤除了生长土壤时，他所做出的事，必须被单独命名。这个命名，就是“外部锚定”。

案例注

上述典型化生命史基于公开法律文书中多起真实案件的合并。作为对这一模式的实证支撑，读者可在若干公开判决书中找到具有高度结构一致性的完整轨迹——其共同特征为：被告人自童年起即被纳入多部门联合帮扶名单，在其档案中详尽记录着从接受各类辅导、到逐渐脱离体制、再到加入犯罪团伙并最终在指令下实施严重暴力行为的全过程。由于此类案件多涉及未成年人犯罪记录封存等法定保密要求，本文在此不逐一提供可公开检索的具体案号。这不是举证上的回避，而是对现行法律框架中隐私保护边界的尊重；同时，本文对被封存档案中所反复出现的结构性模式的分析，不受个案具体身份信息的影响。

五、数字评分：当锚定被算法加速

至此，本文的论证聚焦于以人力执行为主的社会过程。然而，当代安全治理最锋利的一道刃口，不在社工和老师的笔头，而在算法静默运转的服务器里。本章论证：数字风险评估系统的普及，不是简单地“优化”了传统的治理流程，而是在根本上改变了“外部锚定”发生的速度、精度与封闭性。它同时做着三件事：它比人类执行者更密不透风地剥夺着内部生成能力；它通过算法驱动的社会排斥，更高效地将被掏空者推向锚定提供者的捕获半径；并且，它拥有一套封闭的自我验证机制，使得这两件事的后果无法被追踪到它自己身上。

5.1 人力剥夺与算法剥夺：一个必要的内部划分

在此前各章中，我将社工家访、矫正课程、定期报到与算法评分一并置于“安全治理”的统一框架下。这个统摄在逻辑上是成立的——因为它们共享同一个操作逻辑：以外部标准替代内部判断，以可量化指标替代不可量化的过程。但这里有一个差异不能继续被忽视：人力执行的剥夺，和算法执行的剥夺，在剥夺的密实程度上，不在同一个级别。社工家访，无论多精细，总归是一个可能迟到、可能疲惫、可能心软的人在操作；她也许会在某次家访中走神，留出一小片未被监控的缝隙。而算法不放假、不走神、不心软——它的剥夺是全天候的、无间断的、从不因操作者疲劳而留下缝隙的。前者的剥夺留缝，后者密不透风。

这个划分的理论后果是直接的。如果内部生成能力的生长，需要一个最低限度的“不被占用的时间”作为必要条件——本文第二章已论证了这一点——那么在一个由人执行的、偶尔留缝的剥夺过程中，那套能力可能在夹缝里还残留着一点点火种。但在一个由算法执行的、零缝隙的剥夺过程中，那个残留的可能性，被系统性地清除了。两者的生成机制后果因此不在同一个级别上：人力密集型剥夺可能制造出“内部能力严重萎缩但仍偶有挣扎”的人；算法密集型剥夺则可能制造出“内部被完整斩首”的人。本文无意在两者之间划出一道非此即彼的绝对界线，但指出这个内部的张力，有助于让“安全治理”这个概念本身变得更精确、也更具历史纵深——它提醒我们看到，数字技术并非只是让老的治理逻辑“跑得更快”；它是在让老的治理逻辑第一次拥有了完整执行它自身逻辑的技术手段。此前，那个逻辑总要在人力操作的缝隙里妥协；现在，不妥协了。

5.2 数字评分如何加速内部剥夺

传统的安全治理依赖一线工作者的时间、注意力和判断力。一个社工同时管几十个个案，无法对每一个案主都盯得很紧。这种人力局限，在客观上给那些被标记的年轻人保留了一些缝隙——总有些时刻，没人盯着，他可以在那个不被关注的夹缝里，偶然地、侥幸地长一点自己的东西。

数字评分消灭了这些缝隙。无论是少年司法中的累犯风险评估，还是儿童福利中的家庭风险筛查，抑或是学校中的“早期预警系统”，一个共同的架构正在普及：将一个人的若干可量化历史数据（前科、逃课次数、家庭领低保、父母拘留记录、社区犯罪率）输入统计模型，输出一个风险分数。这个分数不体现任何具体人的判断，但它会自动触发一系列标准干预——增加监控频率、强制报到、限制出行。所有这些干预，不再需要某个具体的人去判断“要不要做”；分数会自动把它们推进执行层，像一套自动触发的免疫反应。

那个被评出高分的青少年，他的生活会变得比小军更密不透风。不是因为有个社工特别不信任他，而是因为算法不休息。他逃了一次课，分数当天更新；他被叫去谈话，下一次谈话自动排进日程。他的时间被填充到连一个下午的空隙都留不下来。本文在第二章论述过，内部生成能力的生长需要那些零碎的、不被外部指令占领的碎片时间。这些碎片在数字评分系统面前消失了——不是因为某个人故意剥夺它们，而是因为算法在逻辑上就不承认“不被占用的时间”这个概念本身的正当性。在它的视野里，那些时间不叫“自由空间”，它们叫“未被充分利用的监控窗口”。

5.3 算法排斥如何加速外部捕获

数字评分不仅在剥夺内部能力，它同时也在制造一个让外部锚定系统更容易捕获这些年轻人的环境。风险分数不是一个停在档案里的静态数字——它会旅行。它在教育、社会福利、少年司法系统之间“合法”地跨部门共享。这张信息网的存在，使得任何一个在这张网里工作的人，在亲眼见到这个年轻人之前，已经先看到了他的分数。老师看到分数，他还没坐到教室里，老师就已经在预期“这个学生会带来麻烦”。雇主通过信用系统间接关联到他的记录，他还没提交申请，那份工作就已经不在桌上了。

这张网，从他的角度看，不是保护网——是网墙。他越试图向正常社会靠拢，就越撞在这张网上被弹回来。每一次被弹回来，他的“正常社会没我的位置”这个判断就加固一层。而街角那个锚定提供者，不查他的分数，不调他的档案。他只要他这个人——和这个人身上被长达十几年的算法排斥所积压下来的愤怒。这种无从发泄的、沉默的、系统性的排斥，恰恰是最擅长把人推向那个大声的、清晰的、替他骂出所有愤怒的锚定系统的力量。

5.4 自我验证闭环

数字评分系统最显著的特征，不是它可能出错，而是它拥有一套封闭的自我验证机制，使得它在很多情况下既不能、也不会承认自己出错。一个被打高分的年轻人，因为受到更严密的监控，他的每一次违规被发现的概率，远远高于一个低分者——不是因为他实际上违规更多，而是因为他的生活就有更多的人盯着、他的每一个动作都更可能被记录在案。这些被记录下来的违规，又会重新输入模型，进一步提高他的分数。更严重的是：他因为高分被排斥在工作、教育、居住机会之外，这又使他更容易真实地踏入高风险的生存处境——而这些处境所产生的行为记录，又被系统回收为“模型预测很准确”的证据。

这不是预测，这是自证预言——而且它不需要任何人的恶意来驱动。它只需要分数在系统之间合法地流动，只需要算法按照图纸把高监控分配给高分者，那个闭环就会自己转起来。当这个年轻人最终走上了小军的路——在高分密网中窒息，在街头锚定中寻氧，最后一次极端行动中触发法律系统的终审——他的全部档案会被重新拉出来。人们会指着那条一路上扬的分数曲线说：“看，早就预测到了。”

没有。那曲线不是预测。那曲线是推手。它推着他，从被标记到被掏空，从被掏空到被捕获，再从被捕获走到那个被社会终审为“邪恶”的临界点上。它在完成这一切之后，依旧静静地躺在服务器里，被当作科学治理的典范，写进下一轮系统升级的白皮书。

六、反驳、边界与自我限定

一份负责任的判断，必须事先回答那些最可能杀死它的质疑。本章依次处理三个最有力的反驳，并在结尾划定本文判断的解释力边界，给出明确的证伪条件。

6.1 反驳一：操刀者的正义

这个反驳不一定来自法庭上的法官，更可能来自那些在深夜写评估报告的社工、既成事实论生身上有伤而不得不报警的老师、反复权衡是否该把分数调高一档的评估员。它不是愤怒的，它是疲惫而无奈的：“我做这些，不是为了制造罪犯。我是在保护可能被他伤害的人。如果不打分、不干预、不把他标出来，万一他出了大事，谁对这个后果负责？”

这个反驳首先必须被承认。安全治理的每一个具体操作者，他们面对的困境是真实的：他们被社会赋予“防范风险”的职责，被制度要求做出“可被追溯、可被审查”的决策，而在每一个具体案例中，他们眼睛能够看到的信息都远远不足以做出关于一个人未来的确切判断。在这种情况下，依赖一个分数、启动一套标准干预流程、把你的决定可以追查到“我是按照风险评估结果做的”——这不是最差的决定，甚至可能是他们在现有制度空间里唯一能合理做出的决定。本文从未、也不会把矛头指向个体操刀者的道德意愿。

但承认这个困境，恰恰是本文判断更根本的地方。问题不在个体的意愿，而在他们共同被框住的那个系统逻辑本身。这套逻辑要求他们看到一个生命时，首先看到的是他的分数、标签、他未来可能造成的损失。它给他们留下的制度空间，小到几乎没有可能去问这样的问题：“我现在为他做的这一切保护性操作，在把他当作风险来管理的同时，是不是也在系统性地拆掉他那套能够自己管住自己、自己启动自己、在诱惑面前自己叫停的能力？”如果他们连这个问题的合法位置都没有——因为问了也无法改变任何一个流程的一步——那恰恰证明了本文的判断：那台机器已经越过了所有个体善意，自己在那里转着了。

本文必须在此正面回应反驳者手里最有力的那句话——“那你说怎么办？我不打分，万一他真砍了人，你来担责？”这正是本文的“外部锚定”判断第一次从理论上揭示的那个悲剧性困局：你当前采取的每一项保护他人安全的“预防性干预”，在短期内确实可能阻止一次可见的伤害；但同时，它们已经在为未来制造一个更不可见、更不可控、因而更危险的施害者——不是因为他在变坏，而是因为他自己内部那套能停止伤害的器官，被你的预防性操作一步步地拆掉了。你没有选择不保护。但你面对的不是“保护 vs. 不保护”的选择题。你是被锁在一台悲剧性的机器里：这台机器的结构，让你在尽职尽责地完成每一次“保护”的同时，合力制造着它自己所要防范的对象。认领这个悲剧性，不是要求个体背锅，而是要求我们把问责的方向，从“这个人为什么没拦住”扭转为——“这套必须制造出它自己的‘敌人’才能自我再生产的逻辑，能不能被顶掉？”

6.2 反驳二：被锚定者的能动性何在

这是任何试图用社会因素解释犯罪的论述都不可回避的质问。本文的整个分析链条，听起来确实像是在说：一个人犯罪，全都是社会替他安排好的，他自己没有选择。那我们对他的惩罚，还有什么道德基础？

本文对此的回应，不是“他其实有选择的自由”。那个回应太廉价了，只是在逃避本文自己已经拉开的那个核心追问。本文的回应是：在一个被外部锚定的生命那里，选择的“能力”与“空间”这两件事，必须被分开来讲。在“能力”这个层面上，本文的整个判断已经论证：当一个人的内部生成能力被系统性地剥夺到临界点以下，他那套能够在对错之间痛苦摇摆、能够在外部指令面前做出内心停顿、能够将指令放在自己的完整人生经验中加以审视的器官，已经不存在了。他当然可以在形式上说“不”——他的声带没有被切断，他的四肢没有被绑住。但是那个能让这个“不”字从他里面自己长出来的心理结构，在他被要求做出选择的那一刻，并没有在那个选择现场。

但这绝不意味着他在法律上不能被惩罚。法律惩罚的门槛，历来不以“他是否拥有一套完好的内部生成能力”为条件。它只问客观行为是否能被证明、行为当时的认知能力是否能辨认为性质。本文反对的不是惩罚；本文反对的是：我们在法律惩罚完成之后，把惩罚暗中升级为一种道德上的永恒放逐——把他叫作“邪恶的人”，并用这个标签堵死一切对他被做成的那个过程进行追问的可能。惩罚可以是社会必须对他施加的后果；但“邪恶”，是社会拒绝看到自己在其中占着的那个位置的防御机制。我们可以惩罚他的行为，但我们必须拒绝把“邪恶”这个词当成一个解释来用。那不是解释，那是一道让所有人都不必回头看机器的幕布。

6.3 反驳三：是否存在不可还原的极端恶

这是全文最后一个、也是最尖锐的理论挑战。有些罪行——酷刑折磨、对毫无反抗能力者的仪式化虐待、跨越数年的系统性强暴——似乎在任何意义上都无法用“外部锚定”来解释。施害者不是在执行指令时“没想那么多”；他是在设计痛苦、享受痛苦、在受害者的求

饶中获得满足。这种东西，是否需要一个完全不同于“外部锚定”的、关于“极端恶”的概念来面对？

本文的回应分三层。第一层是承认边界的诚实：是的，这类罪行超出了“外部锚定”的解释范围。“外部锚定”解释的是一种特定类型的参与——那些在系统性、制度化的暴力中，将自己作为他人指令的执行终端、在自身内部空无的状态下做出行动的人。它不解释出于内在邪恶快感的施虐。后者需要另一套概念工具，且很可能涉及精神病态或极端人格障碍的临床与心理学解释。

第二层是边界划定而非理论退缩：承认这个边界，不是本文的弱点，而是本文的纪律。“外部锚定”从一开始就没有声称自己是“所有犯罪”的统一理论。它切入的，是那一长期被既有理论混为一谈的因果路径：当一个人的内部生成器官被结构性拆除之后，他如何被外部指令系统捕获并锚定。如果一起案件的行为人，其内部生成能力从未被系统性地剥夺过——比如一个成长环境相对正常、却在青春后期逐渐表现出以他人痛苦为乐的施虐倾向的人——那么本文的判断不适用于他。那条因果路径需要另一套分析框架。两者可以在同一个刑法典里被判同一条罪名，但在生成机制上，它们不该被同一个概念涵盖。区分这两条路径本身，就是在完成一场政策与理论上都极其必要的分辨：前者指向治理机制的激进改革，后者指向个体层面的干预与控制。

第三层是把“极端恶”本身也推入追问。我们称之为“极端恶”的那些行为，往往被社会迅速收纳进“邪恶”“魔性”“人性之恶”这类不言自明的容器里——这些容器的功能，恰恰是阻止追问。本文不假装自己能回答“极端恶”的起源问题。但它坚持一件事：当我们说一个人“天生就是邪恶的”或者“心里本来就住着恶魔”时，我们不是在解释，我们是在用一个修辞动作，把那个我们还不理解的东西封进一个不透明盒子里，然后在盒子上贴一个标签。本文的整个方法论立场拒绝参与这个贴标签的动作。因此，在本文力所不及之处，本文的诚实态度是：承认边界，悬置判断，拒绝把“极端恶”当成一个起点来用。它也应该是一个终点之前的追问——不是“人为什么这么恶”，而是“我们为什么这么快就把这么恶这件事叫作‘恶’，然后停止往下想了”。

6.4 解释力边界与证伪条件

综合以上三组反驳，本文的判断可以将其适用边界概括如下：本文提出的“外部锚定”这一辨别维度，适用于解释一个在成长史中长期被安全治理密集干预的人，其内部生成能力被系统性地拆除之后，被某种强大而清晰的外部指令系统所捕获，并在此后将该系统的指令作为自身行动依据来执行的一类案件。它不宣称能解释：偶发性激情犯罪、以内在施虐快感为驱动力的极端暴力、精神病态者出于功能性目的的操控性伤害、白领犯罪中高度理性计算的违法行为，以及那些从未经历过安全治理密集干预的个人所犯下的罪行。

本文的判断同时给出以下证伪条件。一个有力的证伪，不需要否定本文所有的经验描述；它只需要满足以下两项条件之一。条件一：找到一组在成长史中经历了同等密度的、以风险预防为名的安全治理干预的个体，他们在内部生成能力被同等程度地系统性拆除之后，在面临“必须行动却无据可依”的悬空状态时，有相当比例的人没有走向任何外部锚定系统，而是在长时间的悬空之后，从自己内部重新发生出了一套稳定的行动依据和自我约束能力——且这一“重新发生”的比例，等于或高于他们被外部锚定系统捕获的比例。这样的证据将直接重创本文的因果论证。条件二：找到一组被外部锚定的人在脱离锚定系统之后，在未接受任何针对性的、长期的特殊干预（即重新提供那些被剥夺了十几年的生长条件）的情况下，其内部生成能力在一段相对较短的时间内（例如几年内）自行恢复到了能够独立进行道德判断与自我约束的水平。如果这种情况是常见的而非常态之外的个案，那么本文判断中关于“拆除的结构性”和“恢复的非对称性”的核心论证将被削弱——尽管这不会完全推翻本文对“剥夺如何发生”的论述，但它会要求本文重新评估“锚定”的持久性与深度。

这两个证伪条件，是可以被经验研究来检验的。它们不是不可证伪的自封命题。本文欢迎这种检验。

七、结论：既不回归，也不放弃

本文的全部论证，可以被凝缩成这样一个判断：一个被社会事后锁定为“罪犯”的人，他在一条被治理逻辑从头铺到尾的轨道上，被系统性地拆除了一套对自己行动进行内部发问、停留与校准的能力；当这套能力被拆除殆尽，他被迫在自身的空无中，将第一个足够强大的外部指令系统接纳为自己唯一的骨骼——此后他做出的那些被社会编码为“邪恶”的事，不过是这台机器在他身上完成的最后一道工序。

这个判断不把“邪恶”当成解释项。它把“邪恶”本身，还原为一个有待追问的社会命名过程。

本文在展开中，与鲍迈斯特、阿伦特、津巴多逐一正面碰撞，论证了他们各自握住的解释线在本文所聚焦的那类案件上同时失效——因为他们全都预设了一个曾被允许存在、只是后来被耗尽、被放弃、或被暂时覆盖的内部器官。而本文所追踪的那些生命，那个器官的生长土壤，在它有机会长出任何东西之前，就已经被铲平了。这就是“外部锚定”从这些先行者的共同失效之处，被逼出来的那条新的辨别线。

在结论处，本文必须诚实面对一个在论证全程被悬置、却不可再回避的问题：出路是什么？

这里有一个极易滑入的陷阱：把“内部生成能力”默认为“一个人原本该有的样子”，然后暗示出路是“回到那种未被治理污染的、能自我生成的状态”。这是既成事实论最深的口音——预设一个被夺走的、更本真的人性本质，等着被回归。本文拒绝这个预设。“内部

生成能力”本身也不是什么先在的、每个人都有权拥有却被社会夺走了的“善良种子”。它也是被发生出来的——在特定的、在社会治理密集干预之前可能偶然存在过一阵子的土壤条件里，被发生出来的。它并不“本该”被任何人拥有。

因此，出路不是“回归”。出路是一个更难、也更诚实的问题：什么样的社会条件，可以让这套能力在当下被重新发生出来？这种重新发生，与安全治理的逻辑之间，是否必须是一种非此即彼的、不可调和的矛盾？

如果答案是“必须非此即彼”——如果任何以预防风险为名的外部干预，都必然挤占内部生成能力所需要的那些不被填满的缝隙——那么本文所揭示的那个悲剧性机器将没有出口。它只会继续转下去，继续在保护人的同时，合力制造着它所防范的对象。承认这一点，不是悲观，是诚实。如果答案是“有可能调和”——那本文没有能力在此给出一个现成的方案，但它至少可以指出那个方案必须解决的核心张力所在：如何在治理为预防风险而必须介入一个人生命的那些时刻，为他的内部生长保留或主动开辟出不可被外部指令占领的缝隙——那些他能自己发愣、自己困惑、自己挣扎、并且在挣扎失败之后不被当作故障来惩罚的空间？这不是在要求治理变得更“人性化”一些；这是在要求治理在结构上，为它自己的反面留出位置。本文不能替代政策制定者来回答“怎么做”。但本文至少把问题摆到了它该在的地方：不是在“这个人怎么没拦住”，而是在“这台机器的逻辑本身，怎么被顶掉”。

最后，有一笔必须收在明处。本文在引言中承诺过，要将那个把被锚定者的行为命名为“邪恶”的社会命名过程本身，作为待解释项来开启——而不把它当作不言自明的自然属性。这笔追问，本文在论证全程中始终将它作为一个持续的暗线在运行：每一次当“邪恶”这个词可以被当作解释项来使用的时候，本文都逼自己绕开它，转而追问“这个行为是从什么样的生成路径里被逼出来的”。但这笔追问本身，还没有成为文本的正面主题。它应该成为本文之后、下一篇论文的起点：那个把某种行为叫作“邪恶”的社会装置，它的发生条件、它的制度功能、它与法律惩罚体系之间的分工与共谋是什么？当我们说一个人“邪恶”的时候，我们究竟是在描述他，还是在用这个命名来防御一个我们不愿面对的、关于我们自身治理逻辑的真相？

这个问题不在本文的结论之内。但本文的结论，必须为这个问题留出它的入口。

（完）

注释

[1] 以下典型化生命史基于多起公开法律文书的综合分析。由于相关案件多涉及未成年人犯罪记录封存等法定保密要求，本文在此不逐一提供可公开检索的具体案号。这不是举证上的回避，而是对现行法律框架中隐私保护边界的尊重。对封存档案中所反复出现的结构性模式的分析，不受个案具体身份信息的影响。

[2] 关于困境儿童群体中高度结构化干预与自主性衰减之间的关系，可参见例如 Lareau, Annette. *Unequal Childhoods: Class, Race, and Family Life* (University of California Press, 2003; 2nd ed. 2011), 其中对中产阶级“协作培养”与贫困/工人阶级“自然成长”两种模式下儿童自主时间安排差异的分析，为本文关于“时间被结构化”之长期后果的论证提供了相关的民族志背景。需要指出，Lareau 的分析框架与本文的规范性判断并不相同——她并不将中产阶级的“协作培养”视为剥夺——但她的田野记录（特别是对于工人阶级儿童无结构的、自发组织游戏的大量描述）为本文关于“不被填满的时间”的生长性功能提供了一个来自平行论域的经验参照。

[3] 关于学校纪律处分体系中“挣扎不被识别”的微观机制，可参见 Rios, Victor M. *Punished: Policing the Lives of Black and Latino Boys* (New York University Press, 2011), 尤其是其中关于被标记青少年在试图向上努力却被读作“尚不足以证明真实改变”的过程的质性记录。Rios的分析对象是美国青少年司法体系中的拉丁裔与非裔男孩，与本文关注的语境不同，但他所描述的“挣扎被污名化”机制——每一次达不到戏剧性转变标准的行为改善，都被视为“还不够好”而非“正在变好”——与本文案例中的动态具有高度的结构同源性。

深化增补：被放跑的最近邻、操作化与证伪

以下四节为编辑增补，与作者正文分开标注。

增补一：与平庸之恶、中和技术的硬边界

阿伦特的平庸之恶与赛克斯、马茨阿的中和技术是本文最近的两个对手。平庸之恶解释加害者为何不觉得自己在作恶（思之缺席），其主体仍在场只是不思；中和技术解释行为人如何用一套说辞解除自身的道德约束，其主体主动使用了这些说辞。本文主张的是更前一步：那套能够思、能够为自己找说辞的内部生成器官，在他做出所谓选择之前已被治理过程预先拆除，因此他不是不思，也不是自我开脱，而是无从生成。可判别面：不思者与中和者在被给予充分的道德对话空间后，应显现出可被唤起的内部判断；本文所述者在同等条件下显现的不是判断，而是对现成脚本的再次索取。若后者不存在，本文与前两家应视为同一现象的不同措辞。

增补二：内部生成能力的可编码形式

本文的核心变量若不可编码，撤除机制就无法被检验。三项：脚本外停留时长——个体在没有任何现成答案可用时，主动保持未决而不转向求助或服从的时长；理由自源率——其行动说明中来自自己界定的判准而非引用外部权威的比例；校准记录——一生中曾因自身反思而改变过某个立场的可追溯次数。三项都测环境是否给过生成机会与个体是否曾生成，均不给个体贴任何倾向性标签。这一取向是本文的实质主张之一：拆除发生在治理侧，指标就必须指向治理侧。

增补三：与同批《善的慢发生》《罪涌与恶壁》的分工

同批三篇处理同一原初问题的不同环节，须明确标注以免读成竞争解释。《善的慢发生》处理善的发生条件被效率逻辑清除；本篇处理内部生成能力被安全治理拆除后，个体被外部指令系统捕获的衔接机制；《罪涌与恶壁》处理压力溃决与社群凝固的共生。三者的共同底盘是同一个判断——那些看似无用的条件被清除之后，剩下的位置只能被现成指令填充；分野在清除者不同（效率逻辑／安全治理／关系失败）与后果形态不同（枯竭／捕获／溃决）。

增补四：不得用于筛查的硬限界

本文最危险的误用是被当作一套识别“将被外部指令捕获者”的前置评估。这一误用不仅在伦理上不可接受，而且在逻辑上会使本文自我否定：任何以风险防范为名的识别程序，恰恰就是本文所描述的那套拆除内部生成能力的治理动作，因此该程序的推行只会扩大它声称要预防的现象。本文的实践含义只有减法方向——在治理设计中为不确定、为犯错、为自我校准保留未被优化掉的位置。

编辑增补所依据的核验文献

Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil*, Viking Press, 1963.

Gresham M. Sykes & David Matza, “Techniques of Neutralization: A Theory of Delinquency,” *American Sociological Review*, 22(6), 1957.

Michel Foucault, *Sécurité, territoire, population: Cours au Collège de France (1977–1978)*, Gallimard/Seuil, 2004.

Ulrich Beck, *Risikogesellschaft: Auf dem Weg in eine andere Moderne*, Suhrkamp, 1986.

Jonathan Simon, *Governing Through Crime*, Oxford University Press, 2007.