高鹏讲

贝卡利亚

普通人都能懂

国家凭什么罚你?

高鹏　著

德麦国际出版社
DEMAI INTERNATIONAL PRESS

出 版 信 息

书名	《高鹏讲贝卡利亚：普通人都能懂——国家凭什么罚你？》
著者	高鹏
执笔	复合主体：高鹏确定承重母题、判词与全书方向，AI（Claude）协作成稿、统稿与修订，经作者审定
出版	德麦国际出版社（Demai International Press）· 新加坡
编号	德麦国际专著第 210 号
ISBN	979-8-90690-235-1
定价	US$20
开本	170 mm × 240 mm
字数	约 21 万字
版次	2026 年 9 月第 1 版

AI 协作声明：全书的承重母题、判词、SDE 分析框架与全书方向由作者确定；文字在作者指令下由 AI（Claude）协作成稿、统稿与修订，并经作者审定。选择什么判词、采用什么证据、怎样对待争议、错了怎样改，责任都在作者。书中关于《论犯罪与刑罚》的转述，以剑桥版英译本（Bellamy 编、Davies 译）为底本核对章号，中文页码留待引证终审；凡标〔待核〕处尚未逐字核对原文。第 44、56、64 章所用的盲编码，由不知本书结论的大模型编码者完成，人类编码者复核仍待做；附录二所列史实出版前应逐条核对原始法令。书中家庭、学校、公司的场景均为说明用的日常例子，不构成法律意见。

全书结构：出版信息、作者介绍、前言、导读、目录；导论与十五编八十二章；结语、参考书目、附录二件（判词演变表、死刑废止—恢复事件编码表）。

作 者 介 绍

高鹏，SDE 学员，法学硕士，高校教师，兼职律师。

高鹏早年修习金融专业，毕业后供职于国有商业银行市级分行，历经八年职场历练，获评中级会计师。缘于对法学的热忱与追求，他再度求学，考取全日制法学硕士。硕士毕业后，他任教于公立本科院校，主讲法学课程，并以兼职律师身份参与法律实务，多年来在制度规范、司法判断与现实经验之间持续往返。

2026 年上半年，王德生先生到他所在的学校讲学。高鹏当面受教，深受 SDE 思想体系启发，由此追随研习，成为 SDE 终身修习学员。

他的研究以法理学与法社会学为根基，把法律看作一种在感知、制度与时间中真实发生的过程。他从法律与文学母题起步，提出“法触觉”“道义语法”“活着对待”等命名；随后向司法认知、自由秩序与制度工程扩展，提出慢反馈回路、司法缄默、秩序自噬、纸权利通胀、命名权、惩罚边界的归因性基础设施等概念，始终在制度的缝隙处追问法律赖以生效的发生条件。贯穿其写作的价值，是让法律重新成为一种活着对待人、而非仅仅正确处理案件的实践。

他已在 SDE　Universes 学员专栏发表理论文章百余篇，著有德麦国际专著第 99 号《没有那一栏》、第 180 号《高鹏讲杜威：普通人都能懂》、第 184 号《高鹏讲加缪：普通人都能懂》。本书是他的第四部专著。

前 言

国家手里那把刀，要不要出生证明

◆ 一、一个二十六岁的年轻人，问了一句国家不爱听的话

一七六四年，米兰。一个二十六岁的米兰贵族青年，匿名出版了一本薄薄的小书，叫《论犯罪与刑罚》。书不厚，放在今天，一个下午就能读完。可它一出版，就被译成法文、英文、德文、俄文，伏尔泰为它写评注，叶卡捷琳娜二世把它的段落抄进给立法委员会的《圣谕》，托斯卡纳大公二十多年后干脆废除了死刑和酷刑。

这个年轻人叫切萨雷·贝卡利亚。他问的问题，今天听来平常，当时却很刺耳：国家凭什么罚你？

在他之前，这个问题几乎不用问。国王有剑，法官有权，酷刑是寻找真相的正常手段，死刑是维持秩序的正常手段。刑罚就像国家身上长的一个器官，天生就在那里。贝卡利亚做的事，是把这个器官拿到桌面上，要它说明：你从哪里来？你为了什么？你最多能做到哪一步？

这本书就从这三个问题出发，陪你把他那本小书，连同它身后两百六十年的回声，一章一章读下去。

这本小书不是一个人关在书房里想出来的。当时的米兰，属于哈布斯堡治下的伦巴第，行政上已经开始讲统计、讲税制、讲效率，刑法却还停在旧世界：秘密审讯、酷刑取供、按身份定罪、死刑与示众。几个年轻人常在韦里兄弟家里聚会，自称“拳头社”，办一份叫《咖啡馆》的小刊物，争论经济、法律和习俗。彼得罗·韦里那时正在收集一百多年前一桩冤案的卷宗，贝卡利亚就在这样的空气里写下了他的书。新旧两套治理同时存在，一个人忽然看见其中不协调的接缝，问题就出生了。

◆ 二、先还他公道

读一个两百多年前的人，最容易犯的错，是拿今天的尺子去量他，量出一堆“不够现代”。这本书先不这样做。前三十来章，我们先还他公道。

他最大的革命，不是把刑罚变轻，而是逼刑罚权拿出出生证明。国家处罚人，从此不再是天然的起点，而是一个必须不断证明的结论。他说，刑罚只有在“必要”的时候才正当，超过必要的部分就是暴政；刑罚要与罪行相称，偷一块面包和杀一个人不能是一回事；刑罚必须由事先公开的法律规定，法官不能在看见一个具体的人之后，才临时决定他犯了什么罪。

他还看穿了一件今天仍然有人不信的事：罚得狠，不等于管得住。真正让人不敢犯罪的，不是刑罚有多痛，而是“跑不掉”——规则、发现、证明、判决、执行这一整条路是不是可信。酷刑之所以荒唐，不只是因为残忍，更因为它会制造假象：强者咬牙不招，弱者先招，疼痛找到的不是真相，而是最怕疼的人。

死刑那一章，是他全书最有名的一章，也是最能看出他本事的一章。他没有只说“死刑太残忍”，而是先问了一个更冷的问题：组成国家的那一份份自由里，有没有哪一份，是一个人把自己的命交了出去？谁会把处死自己的权力交给别人？问完授权，他才去问效果：死刑真的比长期的劳役更能吓住人吗？一次处决的恐怖，很快会被遗忘；一个人在众人眼前长年劳作、偿还社会，才是日复一日的提醒。

他还把目光从刑场往前挪了一大步：最好的刑法，是不用出现的刑法。与其等犯罪发生了再罚，不如让法律清楚简明，让知识传开，让官员自己守法，让教育把人培养成能理解规则、自己作选择的人。一本讲刑罚的书，最后几章在讲教育和预防——这在当时，是很少见的眼光。

这些，是他站得住的地方。本书不打折扣地写出来。

◆ 三、一根柱子和一位作家

一六三〇年，米兰闹瘟疫。两个人被指控往墙上涂油传播瘟疫，在反复刑讯下互相指认，最后被处死。其中一人的房子被拆掉，原地竖起一根“耻辱柱”，把“这里曾有罪恶”写进城市的街道。这根柱子立了将近一百五十年，直到一七七八年才被拆除。

一七六二年，法国图卢兹，新教商人让·卡拉斯被控杀害想改信天主教的儿子，被处以车轮刑。伏尔泰为他奔走呼号，写下《论宽容》；三年后，卡拉斯被宣告无罪。

两桩冤案，一个纠正得极慢，一个纠正得很快。本书第 23 章把它们并排放着问：一个错误判决一旦被写进环境——一根柱子、一块碑、一份官方口径——会不会拉长它被纠正的时间？两个案子给出的方向一致，但也诚实地承认：有柱子的那一桩恰好没有伏尔泰，没柱子的那一桩恰好有，两件事分不开。这就是这本书读贝卡利亚的方式：看见一个问题，就给它找一个会输的办法。

◆ 四、再拆他的根

还了公道，才轮到拆。这本书用的拆法，叫 SDE：S 是显露，一样东西以什么样子出现在我们面前；D 是差异序列，它沿着一条什么样的路一步步变过来；E 是特征纠缠，它扎在一片什么样的土里。三者互相决定，没有哪个是第一位的。

用这把刀去切贝卡利亚，会看见他许多次亲眼看见了“发生”：工具会改变证据，规则会改变行为，秘密告发会改变人和人之间的关系，禁止得越多反而可能制造出更多的违法，教育会改变一个人以后怎样选择。把他写成一个冷冰冰的静态理性主义者，是误读。

可这本书也不打算替他鼓掌。它的方法有一个固定的节拍：先说“看见”，再说“收回”。每当我们以为抓住了他的一个缺口，就必须回到原典，看他自己是不是已经堵上了。结果，这本书一路收回了好几个自己最初很得意的判断：说他迷信

精确量刑，撤回；说他把人当成固定不变的原子，收窄；说他是“自由中心主义者”，改成“自由限权型公共幸福中心”；说他“搭了看台，少了楼梯”，从对他本人的指控降成了对一七六四年文本的结构诊断。这些撤回都留在书里，书末附录一把它们一条条列出来。

◆ 五、锁和钥匙

拆到最后，剩下的是一道闸。

先说一件小区里常见的事。业主公约写着：物业不得擅自进入业主家中——这是一把锁。紧接着一行小字：紧急情况除外——这是一把钥匙。什么算紧急？如果由物业自己判断，那么这把锁锁的是物业，钥匙也在物业手里。

贝卡利亚反死刑的那一章，正是这样一道闸。锁由每个人交出的最小自由铸成，对普通犯罪几乎焊死：偷盗、伤害，乃至普通的杀人，国家都无权处死。可闸上留了一扇门：一个人即使身陷囹圄，仍然握有足以颠覆国家的关系与力量，这时处死他可能是必要的。谁来判断一个人是否足以颠覆国家？正是自认为受到威胁的国家自己。一七九二年，他在晚年的死刑意见书里，依然保留了这扇门。

这本书把这种结构叫作“锁钥同手”。它让我们能用一句话，给全书下判词：

他逼国家为手里那把刀出示出生证明，却把最硬一道闸的钥匙，留在了国家手里。

这不是对他的道德指责。在一七六四年的米兰，除了国家，他找不到第二个可以握钥匙的人。但结构上的空位就在那里，而且两百多年来，一次次被人用上。

◆ 六、这本书怎样对待自己

一本解构别人的书，最容易在最后获得一种危险的自信。这本书给自己立了十一条推翻条件，写在第 80 章，每一条都注明：已触发、部分触发，还是未检验。

有几条，是在这一版里真的跑过的。我们把判断思想家“中心”的那组问题——谁上锁、锁多硬、钥匙在谁手里、门朝哪边开——交给一位不知道本书结论的

编码者，去量边沁、康德、罗尔斯、密尔、诺齐克和贝卡利亚。结果三处一致，一处打了作者的脸：我们漏读了康德留下的一个例外。我们又请编码者把一七八六年到一九七〇年间所有“废除死刑、后又恢复”的事件逐一登记，共三十六件，去检验“锁钥同手”的历史预测。结果是部分失败：钥匙在行政手里时，死刑并不总是“先”从国家安全这扇门回来。失败照登，改判照写。

这样做，不是为了显得谦虚，而是因为本书的方法有一条死规矩：不能输的方法，不配拿来判断别人。

有一条最要紧的，是关于这本书自己的方法。SDE 这把刀用顺了，很容易把一切都说成“发生出来的”，然后宣布：既然有来历，就没有资格说自己是真理。第 81、82 章专门转过刀口对着 SDE：发生能说明来历，不自动说明正当；一项权利即使有历史，也不等于它只是偶然的习惯。一把只会拆、不会立的刀，最后会连贝卡利亚想保护的那些东西——罪刑法定、比例、无罪推定——一起拆掉。本书给 SDE 也写了五种死法，其中一种叫“永不认输”：每遇到反证就换一个更抽象的说法继续成立。附录一里那张判词演变表，就是本书没有走上这条死路的证据。

◆ 七、为什么是现在

贝卡利亚最危险的愿望，是让刑罚像几何一样确定、可计算。他那个时代做不到。今天，算法可以给每个人算出一个风险分数，精确到小数点后一位。

于是他当年问的问题换了一张脸回来。过去，是法官或官员说“为了安全，这是必要的”；今天，可能是一个模型说“风险 78.6%，因此必要”。数字让判断看起来客观，却可能悄悄把“谁有权宣布必要”这把钥匙，从人的手里交给机器。这本书的第十三编专门处理这件事：必要性不是一个风险分数；机器可以帮我们估计会发生什么，却不能替社会签署权利让渡书。

家庭、学校、公司也是一样。父母、老师、老板都握着某种“罚”的权力。孩子打碎了邻居家的玻璃，是罚站十分钟，还是陪他去道歉、商量怎么赔？员工一次失误，是扣掉全年奖金，还是一起查清流程哪里有漏洞？学生打了架，是“一律停学三天”，还是先弄清谁先动手、谁在自卫？这些日常决定，和国家要不要判一个

人死刑，规模天差地别，结构却惊人地相似。第十四编把贝卡利亚请进这些地方，问同样的问题：你凭什么？非这样不可吗？钥匙在谁手里？

◆ 八、这本书不做什么，写给谁

这本书不是法律意见，也不替任何具体案件出主意；书中关于家庭、学校、公司的场景，只是帮助理解的日常例子。它也不宣称发现了一套全新的刑罚理论——贝卡利亚研究两百多年，许多区分早有人说过，本书在导读和第 64 章都老实交代了先占。

它写给三种人：想弄明白“国家凭什么罚人”的普通读者；每天都要做“罚不罚、罚多重”决定的父母、老师和管理者；以及想看一看 SDE 这把刀，在一个真实的思想家身上怎样下刀、又怎样被顶回来的读者。

你不必先读过《论犯罪与刑罚》。导读会告诉你从哪里进、怎样读、怎样推翻这本书。

导 读

五条判决，一把锁，三个随身问题

◆ 一、五条判决

全书最后给贝卡利亚留下五条判决。它们在结语里逐条验收，这里先放在你手边：

判决	一句话	主场
一 · 出生证明	他最大的革命，不是把刑罚变轻，而是逼刑罚权拿出出生证明	第一编、第六编
二 · 制度行为	他把刑罚从情绪化报复，改写成受一般法律、必要性、比例和公共理由约束的制度行为	第二至第五编
三 · 看见发生	工具改变证据，规则改变行为，秘密改变关系，过度禁止制造违法，教育改变选择	第四编、第七编、第八编
四 · 锁钥同手	公共幸福定方向，自由设硬闸；硬闸对普通犯罪几乎封死，在“足以颠覆国家”处留门，钥匙在国家手里	第九编、第十编
五 · 权力姿势	处罚者必须解释，强制者必须举证，制度必须允许纠错	第十三至第十五编

◆ 二、一句判词

五条判决压成一句，就是全书的判词。它在前言第一次出现，在这里第二次出现，在第 56 章总解构处和结语里还会一字不改地出现：

他逼国家为手里那把刀出示出生证明，却把最硬一道闸的钥匙，留在了国家手里。

◆ 三、三条读法

快读路线（一个周末）：导论四章，第 9 章“谁有权宣布必要”，第 30 章“先问授权”，第 44 章“自由中心还是幸福中心”，第 56 章“锁钥同手”，第 72 章“必要性不是一个风险分数”，结语。

理论路线：第 33—35 章康德与死刑，第 44 章与第 56 章的诊断工具及其校准，第 62—64 章康德、边沁与哈特，第 78—82 章死法与推翻条件。

生活路线：导论，第 25—29 章“罚得重才害怕吗”，第 36—39 章“法律也制造犯罪”，第十四编家庭、学校、公司五章，再看各章末的“小账”。

◆ 四、怎样推翻这本书

本书的判断都是可以被推翻的工作诊断，不是对贝卡利亚的永久定性。推翻它，有三个地方可去：

第一，第 80 章的推翻条件总表（F1—F11）。每一条写明“什么证据出现，本书就必须改判”，以及现在的状态。第二，附录一“判词演变”，列出原典怎样一次次把本书顶回来。第三，附录二“死刑废止—恢复事件编码表”，第 56 章那次部分失败的检验，底账全在这里，你可以逐条复核、重新编码。

凡引贝卡利亚原典处，正文用“原典锚点：剑桥版第某章”标出章号，你可以按下面第九节的对照表，到任何一个中译本里去找。

◆ 五、三个随身问题

读完这本书，最希望你带走的不是某个结论，而是三个问题。下一次，当有人——国家、学校、公司、平台，甚至你自己——要让别人承受一个不利后果时，问一问：

你凭什么？——这项权力从哪里来，谁授予的，授予了多少。

非这样不可吗？——有没有更轻、同样有效的办法。

钥匙在谁手里？——如果有例外，谁来判断例外成立；判断的人，是不是正好是被限制的那一方。

◆ 六、每章怎么写

每一章大体是同一个节拍：从一个日常场景或一个历史现场开头；再说贝卡利亚在这里说了什么、对在哪里；然后用 SDE 看这一处，“看见”了什么，又必须“收回”什么；最后放回今天的日子里。

每章末尾有两样东西。一是“带走的话”，一句话，记住它就记住了这一章。二是“小账”，两行：他看见了什么，接缝在哪里。小账是留给你继续算的。

◆ 七、凡例

一、底本与章序。本书关于《论犯罪与刑罚》的原典核验，以 Richard Bellamy 编、Richard Davies 译的 Cambridge Texts in the History of Political Thought 版（Cambridge University Press, 1995；2012 年上线）为主要英译底本，章序统一依其四十七章编排。中文表述以忠实转述为主；中文译本的版次与精确页码留待引证终审锁定，本书不编造页码。正文核心论证以“（原典锚点：剑桥版第 X 章）”标明章号；中文阅读以黄风译北京大学出版社 2008 年四十七章版为参照。

二、SDE 术语。S＝Show，统一译为“显露”；D＝差异序列；E＝特征纠缠。“形状、路、土”仅作为面向普通读者的生活比喻，不作为 S、D、E 的定义。三者不是线性三步，而是互相决定：S＝F(D,E)，D＝G(S,E)，E＝H(S,D)。本书另用两个工作词：“123 原理”指一次张力（1）逼出一次新显露（2），新显露再回写路径与环境（3）；“回写”指某个结果回到它所从出的结构、改变下一轮选择，使用时必须能说出什么结果、回到什么结构、经过什么通道、凭什么证据（第 82 章），说不出的就不用这个词。

三、判词的性质。书中“看见”“收回”“判词”“死法”都是可被原典、历史材料和竞争解释推翻的工作诊断，不是对贝卡利亚的永久定性。凡属二十一世纪

的制度延伸，均明确标作“现代延伸诊断”，不得倒装为贝卡利亚本人已经面对的问题。

四、增量的边界。本书不再把“契约与功利并存”“一般正当化与责任分配相区分”或“惩罚权来源”本身当作原创发现；Bellamy、H. L. A. Hart、Ronald Dworkin、Robert Nozick、Michael S. Moore、A. John Simmons、R. A. Duff 等已有重要论述。本书尝试保留的增量收窄为两类：第一，把“谁上锁、锁多硬、钥匙挂在哪、谁握钥匙、门朝哪边开”这一组锁钥问题转用为思想史诊断工具，先在边沁、康德、罗尔斯三个答案已知的对象上校准，再交不知结论的编码者盲编码（第 44 章）；由此得到“锁钥同手”的判断，并与施米特的例外论、罗西特的紧急权力准则和“任何人不得做自己案件的法官”划界，只以“位置×方向”的复合读数保留增量（第 56 章）。第二，检验贝卡利亚“先问授权、再问目的”的问题次序：作者六章编码与第二编码者十八章扩样表明，授权追问集中在“权力归谁”与“定罪之前的人身”两类问题上（第 64 章）。原先设想的第三项“失效图谱”，与预测警务的反馈回路研究对照后已被先占，只作教学框架保留（第 78 章）。

◆ 八、本书与既有研究的关系

贝卡利亚研究早已注意到他的契约语言与公共幸福语言并存。Richard Bellamy 的剑桥版导论把刑罚放入更广泛的社会组织、实践与制度中理解；Harcourt 也从现代刑法基础、形式主义与福柯式问题意识反读贝卡利亚。因此，本书所谓“自由限权型公共幸福中心”首先是解释性压缩，不以“第一次发现两条线并存”自居。

Hart 在《刑罚原则导论》中区分刑罚的一般正当化目的与处罚分配问题；他还专门讨论过边沁与贝卡利亚。本书据此撤回原先对“两阶段正当化”的原创主张。“第零闸”也不再作为理论发明：洛克的自然法执行权、Simmons 对洛克惩罚权的重构，以及 Duff 关于政治共同体何以有资格要求行为人作答的研究，都说明“谁有权罚”早已是独立课题。

大部分具体区分可以由证据法、行政法、标签理论和刑罚哲学重述，SDE 在本书中首先是一种迫使改判、暴露路径与保留反身性的研究纪律。第 23 章的“环境持续性”以米兰案与卡拉斯案对照后，仍与外部申诉力量混杂，只能记为待分离的候选；“锁钥同手”的历史预测在全样本编码中部分失败，留下待检的新假设（F11′）。

◆ 九、原典章号对照

《论犯罪与刑罚》在作者生前多次增订，章序几经变动；1766 年莫雷莱的法译本又按译者的理解重排了全书。本书一律按剑桥版（依意大利第五版，引言加四十七章）标注“原典锚点”。下表列出正文实际使用的核心章节，读者对照黄风译本或其他版本时，请按章名查找。

剑桥版章号	章名	本书主要使用处
引言	引言（“最大多数人分享的最大幸福”）	第44章
第2章	惩罚权	第5章
第4章	法律解释	第16—17章
第6章	罪刑比例	第12—13章
第7章	衡量刑罚的错误	第10—11章
第12章	刑罚的目的	第44章
第15章	秘密告发	第37章
第16章	酷刑	第20—23章
第19章	刑罚的及时性	第27章
第23章	耻辱	第38章
第25章	放逐与没收	第38章
第27章	刑罚的温和	第26章
第28章	死刑	第30—32章、第56章
第41章	如何预防犯罪	第36、41章

剑桥版章号	章名	本书主要使用处
第45章	教育	第42章
第46章	赦免	第26章、第32章后专题
第47章	结论	结语

版本提醒：意大利 1764 年初版的章序与此不同，最容易出错的是死刑章——它在初版中是第十六章，而剑桥版第 16 章是“酷刑”。意大利文献（如 Treccani 词条）引用“第十六章”谈死刑时，指的是初版章序。初版与莫雷莱法译本的完整章号对照，〔待核：据意大利国家版核定后补入〕。

目 录

作者介绍 · · · · · iii
前言　国家手里那把刀，要不要出生证明 · · · · · iv
导读　五条判决，一把锁，三个随身问题 · · · · · x

导论　先别问他主张什么，先看十八世纪的人是怎么被罚的 · · · · · 1
第 1 章　一次审判可以有多可怕 · · · · · 2
第 2 章　为什么所有人都觉得“国家当然可以罚人” · · · · · 7
第 3 章　米兰已经开始现代化，刑法为什么还留在旧世界 · · · · · 11
第 4 章　二十六岁的年轻人为什么突然插手刑罚 · · · · · 16

第一编　国家手里为什么有一把刀 · · · · · 20
第 5 章　处罚权不是从天上掉下来的 · · · · · 21
第 6 章　每个人究竟交出了多少自由 · · · · · 25
第 7 章　“必要”：贝卡利亚给国家上的第一道锁 · · · · · 30
第 8 章　如果没有必要，处罚就是另一种暴力 · · · · · 34
第 9 章　第一处接缝：谁有权宣布“必要” · · · · · 40

第二编　犯罪不是国王生气 · · · · · 43
第 10 章　到底什么叫犯罪 · · · · · 44
第 11 章　为什么不能按照一个人的“坏心眼”量刑 · · · · · 47
第 12 章　偷一块面包和杀一个人为什么不能是一回事 · · · · · 50
第 13 章　比例不是数学题 · · · · · 53
第 14 章　一把不能太精确的尺 · · · · · 56

第三编　谁写规则，谁判人 · · · · · 59
第 15 章　只有法律可以制造刑罚 · · · · · 60
第 16 章　法官为什么不能临时发明一个罪 · · · · · 63
第 17 章　“解释法律”为什么会让贝卡利亚害怕 · · · · · 73
第 18 章　裁判员真的只能照着记分牌念吗 · · · · · 77
第 19 章　理念自由：自由第一次进入九格表 · · · · · 81

第四编　酷刑：一个寻找真相的机器，为什么会制造假象 · · · · · 85
第 20 章　如果还没有判有罪，他是谁 · · · · · 86
第 21 章　疼痛真的能发现真相吗 · · · · · 91

第 22 章　强者不招，弱者先招 · · · · · 96
第 23 章　司法机器从什么时候开始参与制造“事实” · · · · · 100
第 24 章　贝卡利亚在这里离发生学有多近 · · · · · 107

第五编　最狠的刑罚为什么未必最好 · · · · · 109
第 25 章　人人都相信“罚得重才害怕” · · · · · 110
第 26 章　重不重要，为什么“跑不掉”更重要 · · · · · 113
第 27 章　三年以后再罚，还有多少意义 · · · · · 117
第 28 章　威慑不是一个点，而是一条路 · · · · · 120
第 29 章　现代刑罚第一次变成行为机制研究 · · · · · 123

第六编　国家能不能杀死一个公民 · · · · · 126
第 30 章　先别问死刑残不残忍，先问国家从哪里得到这项权力 · · · · · 127
第 31 章　一个人能不能把自己的命交给国家 · · · · · 130
第 32 章　即使国家有权杀，杀人真的有用吗 · · · · · 133
第 33 章　康德为什么不能接受贝卡利亚 · · · · · 146
第 34 章　两个人到底各自踩着哪块地板 · · · · · 149
第 35 章　死刑争论真正争的不是“善良” · · · · · 152

第七编　法律不仅抓犯罪，也制造犯罪 · · · · · 155
第 36 章　禁止得越多，犯罪为什么可能越多 · · · · · 156
第 37 章　告密制度怎样制造一个人人互疑的社会 · · · · · 159
第 38 章　一个人被叫成“罪犯”以后发生了什么 · · · · · 162
第 39 章　刑罚如何反过来改变社会 · · · · · 165
第 40 章　这是贝卡利亚最接近SDE的一刻 · · · · · 168

第八编　刑法最好的时候，是它不用出现 · · · · · 180
第 41 章　为什么预防比惩罚更重要 · · · · · 181
第 42 章　教育为什么会进入一本刑法书 · · · · · 184
第 43 章　公共幸福究竟是什么 · · · · · 187
第 44 章　贝卡利亚到底是自由中心，还是幸福中心 · · · · · 190
第 45 章　如果幸福才是根，前面的判词会不会倒 · · · · · 197

第九编　中心主义审判：把三大核心重新放上桌 · · · · · 200
第 46 章　把贝卡利亚三大核心重新放上桌 · · · · · 201
第 47 章　它们真的都从公共幸福长出来吗 · · · · · 204
第 48 章　九格总对账：他到底把哪一格放在了中央 · · · · · 207

第十编　看台与楼梯：谁替共同体判断公共幸福 · · · · · 217

第 49 章　哪里是中心，哪里只是结果 · · · · · 218
第 50 章　第一场正式判决 · · · · · 221
第 51 章　社会契约以前站着谁 · · · · · 224
第 52 章　那个自由的人有没有童年 · · · · · 227
第 53 章　语言、家庭、贫困、身份有没有参加犯罪选择 · · · · · 230
第 54 章　如果一个人的自由也是长出来的 · · · · · 233
第 55 章　“先放人，再限权”还能不能成立 · · · · · 236
第 56 章　权力被发生化，主体尚未被完全发生化? · · · · · 239

第十一编　给贝卡利亚找最强的辩护人 · · · · · 255
第 57 章　他并没有说人可以被精确计算 · · · · · 256
第 58 章　他已经承认激情、概率和社会复杂性 · · · · · 259
第 59 章　“自由人”只是规范起点，不是经验的人吗 · · · · · 262
第 60 章　为了限制国家，是不是必须先假定一个不可被国家制造的人 · · · · · 265
第 61 章　如果没有一个先在自由，还有没有人权 · · · · · 268

第十二编　康德、边沁以后：贝卡利亚到底留下了什么 · · · · · 271
第 62 章　康德：不是有用，而是应得 · · · · · 272
第 63 章　边沁：把效用推得更远 · · · · · 275
第 64 章　现代刑法为什么同时继承了两个人 · · · · · 278
第 65 章　罪刑法定为什么比贝卡利亚活得更久 · · · · · 293
第 66 章　死刑为什么没有被一本书杀死 · · · · · 297

第十三编　AI时代，贝卡利亚最危险的愿望可能真的实现了 · · · · · 300
第 67 章　如果机器真的可以预测犯罪风险 · · · · · 301
第 68 章　如果量刑真的越来越可计算 · · · · · 304
第 69 章　贝卡利亚会喜欢算法法官吗 · · · · · 307
第 70 章　“确定性”走到极端，会变成什么 · · · · · 310
第 71 章　预测犯罪是在预防犯罪，还是在制造“罪犯” · · · · · 313
第 72 章　算法第一次把“必要性”变成一个数字，怎么办 · · · · · 316

第十四编　处罚走出法庭：家庭、学校、公司 · · · · · 327
第 73 章　父母为什么不能只靠处罚教育孩子 · · · · · 328
第 74 章　公司管理为什么“罚得狠”不等于有效 · · · · · 331
第 75 章　学校为什么会通过规则制造“坏学生” · · · · · 334
第 76 章　一个制度怎样让人只怕处罚、不相信规则 · · · · · 337
第 77 章　自由不是“没人管我” · · · · · 340

第十五编　给贝卡利亚写死法，也给SDE写死法 · · · · · 343

第 78 章　贝卡利亚体系的主死法候选：把“最少必要”变成永久答案 · · · · · · · · · 344
第 79 章　次死法：三个没有彻底走完的地方 · 349
第 80 章　什么证据会推翻本书 · 352
第 81 章　SDE是不是也先放了一个“发生” · 364
第 82 章　给SDE自己写死法 · 366

结语　国家凭什么罚你? · **368**

参考书目 · **371**

附录 · **375**
附录一　判词演变：原典如何一次次把本书顶回来 · · · · · · · · · · · · · · · · · · · 375
附录二　死刑废止—恢复事件编码表（1786—1970） · · · · · · · · · · · · · · · · · · 378

导　论

先别问他主张什么，先看十八世纪的人是怎么被罚的

编　序

序编四章，是全书导论。第1章看十八世纪的审判怎样把人的身体当作工具；第2章追问国家的处罚权从哪里来；第3章讲米兰改革与旧刑法的错位；第4章讲二十六岁的贝卡利亚怎样在朋友圈的碰撞中写出这本小书。

序编是第一条判决的开场：贝卡利亚最大的革命，是逼刑罚权拿出出生证明。这四章先不急着讲他主张什么，而是把“国家当然可以罚人”这个被习惯盖住的问题重新翻出来，定下全书的起手式。

读这一编时，可以带着一个日常问题进去：物业、学校或单位对你说“这里一直这样”的时候，你有没有问过，这份权力当初是谁给的？

第 1 章

一次审判可以有多可怕

如果一个现代人走进十八世纪欧洲的一间刑事法庭，他最先感到陌生的，未必是法官穿什么衣服，也未必是判决书写得多么古老。他真正会感到不适的，是那种弥漫在整个制度里的确信：为了找出真相、维护秩序、震慑坏人，国家可以把人的身体当成工具。

今天我们已经习惯把“审判”和“处罚”分成两件事：先证明有罪，再决定处罚。至少在原则上，一个人还没有被证明有罪以前，不能先把他当罪犯处理。但在贝卡利亚生活的时代，刑讯并不是历史课本里的野蛮传说，而是一种仍然存在于大陆刑事司法中的制度技术。疼痛可以被当作获取口供的手段，秘密程序可以被当作维持秩序的必要办法，严酷刑罚可以被当作让别人不敢犯罪的公开语言。

最值得注意的不是“那时候的人比较残忍”。这样解释太容易，也太浅。真正的问题是：当一个制度长期存在以后，人们会慢慢忘记它曾经需要理由。酷刑最初也许需要解释，后来只需要程序；死刑最初也许需要辩护，后来只需要宣判；国家为什么有权把人关起来，为什么有权没收财产，为什么有权让身体受苦，甚至为什么有权剥夺生命，这些问题一旦被重复使用的制度盖住，就会从“问题”变成“常识”。

这就是本书第一把钥匙：最牢固的权力，不是天天大声宣布自己的权力，而是让人不再追问它为什么存在。

我们可以用一个日常例子理解。假设一栋楼的物业最初为了防火，请住户把公共消防门的钥匙交给它保管。很多年以后，物业不仅拿着消防门钥匙，还能开你家的门、翻你的柜子、没收你的东西。新住户搬进来时，大家只告诉他：“这里一直这样。”如果有人问：“谁同意它开我家的门？”别人甚至会觉得这个问题很奇怪。制度最厉害的时刻，就是“谁给的权力”不再被当作一个需要回答的问题。

贝卡利亚做的第一件事，正是把这个已经被遗忘的问题重新提了出来。

但是，要理解他的震动，我们还要看见刑罚为什么会那么依赖身体。今天国家处罚一个人，可以有罚金、社区矫正、限制资格、监禁等多种形式；而在旧刑罚世界里，公开示众、肉体痛苦、耻辱和死亡更直接地承担着权力表达。处罚不是悄悄完成的一项行政工作，而是一场让所有人“看见国家”的演出。国家通过罪犯的身体说话：你们看，违抗的结果就是这样。

这种制度并不愚蠢。恰恰相反，它有自己的逻辑。识字率不高、行政能力有限、警察和监狱体系远不如现代国家发达时，公开而强烈的刑罚确实是一种廉价而醒目的传播方式。今天一条规则可以通过学校、媒体、手机通知、数据库、摄像头和日常行政不断重复；那时，权力更需要通过广场、刑台和身体留下记忆。残酷在这里不是偶然装饰，而是传播技术的一部分。

这也是为什么我们不能站在今天简单嘲笑过去。SDE要找的，不是“古人怎么这么坏”，而是“这种做法为什么能在那片土里长出来”。只有把它的土找出来，才知道后来的人究竟拆掉了什么。

贝卡利亚当然看见了残酷，但他比“残酷不好”多走了一步。他发现，如果处罚的目标是减少犯罪，那么国家让人更痛，并不自动等于社会更安全。强度和效果之间没有一条天然直线。一个刑罚可以非常可怕，却因为执行极不稳定而失去威慑；也可以表面温和，却因为几乎无法逃脱而深深影响人的选择。到这里，刑罚第一次不再只是“国家能给多少痛苦”，而开始变成“什么样的制度路径真正会改变行为”。

更重要的是，贝卡利亚把刀伸向审判本身。刑讯最危险的地方，不只是痛，而是它假装自己在“发现真相”。可一旦疼痛进入取证过程，工具就不再中立。它测出的可能不是“谁有罪”，而是“谁更能忍”；不是“事实是什么”，而是“人在多大的疼痛下会说什么”。一个身体虚弱的无辜者，可能比一个意志顽强的罪犯更早认罪。于是，国家以为自己拿到的是事实，实际上得到的可能只是疼痛制造的结果。

请记住这一点，因为它会贯穿全书：贝卡利亚最厉害的时候，总是在把一个被当作透明工具的东西重新变成问题。酷刑不是透明工具，法官的解释不是透明工具，死刑不是透明工具，甚至“严厉”本身也不是透明工具。每一种工具都会参加它所要处理的事情。

这已经非常接近SDE所谓“发生”的眼光：结果不是孤零零地等在那里，被一个中立方法捞出来；方法、环境、人的身体、恐惧和制度一起参与了结果的形成。

所以，本书不会把贝卡利亚只写成“人道主义者”。人道是他的温度，但不是他的全部结构。他更深的动作，是把刑罚从一种习惯性权力，变成一项必须说明来源、目的和效果的公共技术。

第一章只需要留下一个问题：如果一种刑罚已经存在了几百年，甚至人人都习惯了，它还需要证明自己吗？贝卡利亚给出的答案是：越是能够伤害人的权力，越不能靠“我们一直这样”证明自己。

◆ 再看一层：不要把十八世纪简单写成“野蛮时代”

如果我们把这一章只理解成“过去很残酷，后来变文明”，就会把贝卡利亚写浅。因为真正困难的不是发现疼痛很可怕，而是说明为什么一种人人都知道会让人痛苦的制度，仍然能够长期被认为合理。答案在于：制度从来不是只靠道德感觉运行，它还靠一整套关于真相、权威、秩序和人的假定。只要这些假定仍然稳固，残酷就可能被解释成必要，秘密就可能被解释成效率，酷刑就可能被解释成取证。

因此，贝卡利亚真正拆的不是一个“坏习惯”，而是一套把强制合理化的连接方式。他把过去连在一起的几件事拆开：有权审判，不等于有权折磨；追求真相，不等于任何取证手段都可以；维护秩序，不等于处罚越重越有效；法律存在，不等于法律本身无需被评价。每拆开一处，旧制度就少一条“当然如此”的通道。

从这个角度看，现代刑法的形成也不是人类突然变善，而是越来越多原本捆在一起的东西被迫分开接受证明。调查与处罚分开，嫌疑人与罪犯分开，立法与司法

分开，公共安全与无限权力分开。所谓文明，很多时候不是人的心忽然柔软，而是权力再也不能把这些不同环节一句“都是为了正义”糊在一起。

这也是为什么今天仍然需要读贝卡利亚。我们当然不会把十八世纪的刑具原样搬回来，但“把手段藏进目的”这种动作从未消失。只要有人说“为了安全，所以程序可以少一点”“为了效率，所以解释可以宽一点”“为了惩恶，所以证据标准可以松一点”，旧结构就可能换一件现代衣服回来。贝卡利亚真正留给我们的，不是对某一种刑具的历史谴责，而是一种持续拆分的习惯：目的、权限、手段、效果，各自拿证据。

◆ 一个思想实验：如果酷刑真的很有效呢？

为了看清贝卡利亚的锋利，我们故意提出一个让人不舒服的问题：假如有一天有人证明，酷刑在某些案件里确实能更快得到信息，是否就可以恢复？这个问题逼我们区分两条理由。第一条是效果：它能不能得到想要的结果。第二条是权利：国家有没有资格把尚未被证明有罪的人当成获取信息的工具。贝卡利亚反对刑讯时，两条线其实同时存在。他既怀疑疼痛作为真相工具的可靠性，也强调在有罪未定时对一个受法律保护的人施加处罚本身存在权利问题。

这意味着即使某种手段“有效”，也不自动正当。现代人最容易被“效果证明一切”迷惑：一个办法只要能降低风险、提高破案率、提高秩序，就好像获得了通行证。但贝卡利亚已经逼我们看到，制度评价必须至少有两张账——结果账和权限账。结果账问有没有用，权限账问凭什么这样用。

后面讨论死刑时也会遇到同样结构。即使有人证明死刑在某些条件下具有威慑效果，仍然不能直接跳过“国家是否拥有这种权利”；反过来，即使有人从权利上反对死刑，也还要面对现实社会对安全的要求。贝卡利亚真正有生命力的地方，就在这两本账从此很难再被合并成一句“为了大家好”。

带走的话

真正需要解释的，不只是“为什么有人犯罪”，还有“为什么国家因此就可以让一个人受苦”。贝卡利亚的革命，从把第二个问题重新变成问题开始。

> 带走的话
>
> 越是能够伤害人的权力，越不能靠“我们一直这样”证明自己。

小账

他看见了什么？——酷刑测出的可能不是谁有罪，而是谁更能忍。

接缝在哪里？——效果账与权限账要分开，酷刑有效也不自动正当。

第 2 章

为什么所有人都觉得“国家当然可以罚人”

有些问题一旦问出来，我们会奇怪自己以前为什么没问。比如：国家为什么可以处罚人？第一反应往往是——因为有人犯罪了。可这个回答其实把问题绕开了。犯罪只能说明“一个人违反了规则”，并不能自动推出“另一个人或者一个组织可以合法地让他失去自由、财产乃至生命”。中间还缺一座桥：处罚权从哪里来？

在日常生活里，这座桥很好理解。一个陌生人不能因为你说错一句话就把你关进地下室；公司领导不能因为员工迟到就随意打他；学校老师不能因为学生考试作弊就没收学生家里所有财产。我们之所以立刻觉得这些行为不对，是因为“对方做错了”与“我因此可以对他做任何事”之间从来不是同一件事。

可一旦主语换成国家，这个跳跃常常被忽略。国家说：你犯罪了。然后我们自然接上：所以国家可以罚你。贝卡利亚就是在这里停了一下。他问：等一等，为什么？

这个“为什么”把刑法从技术问题推成了政治哲学问题。它不再只是法官如何判、监狱如何管、刑期如何算，而是在追问国家与个人之间最基本的关系。国家不是抽象的天，它是由人组成、通过制度行动的共同体。如果一个普通人不能随便剥夺另一个人的自由，为什么很多普通人集合起来以后，就突然获得了这种能力？如果大家本来都没有权任意杀人，那么把很多“没有这种权”的人加在一起，会不会凭空长出一个“有权杀人”的国家？

贝卡利亚借用了社会契约传统来回答。他的基本想法可以用非常朴素的方式说：孤立的人各自保有全部自由，但也因此生活在极不稳定的危险里。为了换取共同安全，人们同意拿出自己的一小部分自由，交给共同体，使共同体获得维持秩序的力量。国家的处罚权，不是神秘地从皇冠里长出来，也不是统治者出生时自带，而是从这些为了共同安全而交出的自由份额中形成。

这一步的力量非常大，因为它把“国家有权处罚”改写成了一个有来源的命题。既然有来源，就可以问范围。既然是交出来的，就可以问交了多少。既然目的在于共同安全，就可以问某一种处罚是不是超出了这个目的。

请注意这个变化：过去的问题是“国王愿不愿意仁慈”，现在的问题变成“国王有没有这个权”。一个是劝权力少用一点，另一个是要求权力证明它从哪里来。前者仍然默认刀属于国王，只希望他温柔一点；后者则先检查刀的产权证。

这就是贝卡利亚最大的开门动作之一。

从SDE的角度看，他把一块“发现”变成了“发生”。旧制度很容易把刑罚权呈现成一种先在事实：有国家，就有处罚；有统治，就有惩罚。贝卡利亚却说，不，处罚权有生成过程。它来自个人为安全而作出的有限让渡。于是刑罚权不再是“本来在那里”，而是“在一定关系中长出来”。

一个有出生过程的权力，和一个天生存在的权力，最大的区别是什么？不是它更温柔，而是它必须服从自己的出生条件。

假设我们成立一个小区业委会，只授权它处理公共维修。后来它突然决定检查每个住户的手机。即使它说“我是为了小区安全”，住户仍然可以问：当初我们授权你的时候，包括检查手机吗？这就是来源约束。权力不能只拿目的说话，还得拿授权说话。

贝卡利亚用同一把刀去碰死刑时，问题立刻尖锐起来。如果每个人只是为了保住更多自由，才交出最少的一部分自由，那么他是否真的把自己的生命也完整交给国家了？如果没有，国家从哪里获得杀死他的权利？即使后来我们不同意贝卡利亚关于死刑的全部论证，他提出问题的方式已经改变了讨论。国家不能只说“这个人罪大恶极”，还得回答“罪大恶极为什么足以产生我杀他的权利”。

当然，社会契约不是一份历史上真的签过的纸。贝卡利亚不是在档案馆里找到了一张“全体人民让渡自由协议”。它是一种规范性的解释：如果我们要把国家强制理解为正当，就必须把它想成自由的人为了共同安全可以同意的东西。也就是说，契约的作用不是描述某年某月发生过一次大会，而是给权力画边界。

这里也埋下了本书后面最大的疑问。贝卡利亚把国家权力从“先在”拉回了关系发生，可他似乎需要先放进去一个已经拥有自由、能够计算利害、能够交出自由份额的个人。国家不是天生的，那么这个“自由个人”是不是天生的？这一刀我们暂时不下。因为按照的纪律，先把他看见的地方写足，再去抓收回。

第二章真正要留下的，不是“社会契约论”这个名词，而是一种看权力的方法：任何能够合法伤害人的制度，都必须回答三个问题——你从哪里来？你为了什么？你最多能做到哪一步？

◆ 再看一层：社会契约不是一张真的合同

读社会契约最容易犯的错误，是追问“这份合同什么时候签的”。如果把它当作历史文件，当然找不到。贝卡利亚借契约做的事情更接近一种正当性测试：假如每个人仍然被当作一个有自己目的的人，那么国家的强制必须能够解释为这些人为了共同安全可以接受的有限安排。它是一种反向审查，而不是一份档案。

这一区分非常重要。因为一旦把契约理解成规范测试，我们就不会因为“现实中没人签字”而轻易否定它，也不会因为“多数人同意”就把任何强制都叫契约。真正的问题是：一个自由的人为什么有理由接受这一权力？如果答案只能是“因为你出生在这里”，那仍然只是事实，不是正当性。

同时，这也暴露出契约思维的困难。所谓“自由的人”究竟是什么人？他有没有家庭、贫困、教育、身份和情感？他是一个会恐惧、会冲动、会依赖别人的具体人，还是一个已经能独立计算利害的抽象人？贝卡利亚为了限制国家，需要把个人放到国家之前；可一旦把个人放得太干净，就可能把塑造个人的社会关系一并擦掉。

这一难题现在先记账，不急着结算。因为好的解构不是看到一个前提就宣布“你错了”，而是先问：这个前提为什么在当时如此有用？答案很清楚——如果不先给个人一个不能被国家随意吞没的位置，十八世纪限制刑罚权的工程很难启动。自由人的“先在”，也许既是贝卡利亚的边界，又是他的历史武器。

犯罪只能证明一个人违反了规则，不能自动证明国家因此拥有无限的处罚权。贝卡利亚做的，是把“所以可以罚”中间那座被省略的桥重新建出来。

> **带走的话**
>
> 犯罪只能证明有人违反了规则，不能自动证明国家因此拥有无限的处罚权。

> **小账**
>
> 他看见了什么？——“有人犯罪”与“国家可以罚他”之间缺一座桥。
>
> 接缝在哪里？——契约先放进一个自由个人，这个人是不是天生的？

第 3 章

米兰已经开始现代化，刑法为什么还留在旧世界

贝卡利亚不是在一个完全静止的社会里写作。恰恰相反，十八世纪中叶的伦巴第正在经历一种非常尴尬的并置：一边是财政、商业、农业和行政改革，一边是刑事司法中仍然保留的旧制度技术。社会的一只脚已经迈向“用理性改善治理”，另一只脚却还踩在“用痛苦证明权力”的地板上。

这种错位很重要。思想往往不是在一切都最糟糕的时候发生，而是在两个时代相互摩擦的时候发生。一个人只有先看见“这里已经可以不这样”，才会把“为什么还这样”变成问题。

Franco Venturi 对十八世纪改革运动的研究提醒我们：米兰启蒙不能只被理解成几个青年读书人的思想冒险，它同时嵌在哈布斯堡治下伦巴第的自上而下改革、行政理性与公共福利议程中。这个背景既解释贝卡利亚为什么把希望大量寄托于一般立法和开明改革者，也解释后文“看台与楼梯”为什么首先是时代政治形态，而不是简单的人格缺陷。

如果整个社会都把酷刑、秘密审判、严刑峻法当作不可想象的自然法则，人们甚至很难获得批判它们的语言。可是当商业、政治经济学、统计、行政效率、公共幸福这些新词进入公共讨论，刑事司法就突然显得笨重而古老：既然税收可以算成本，贸易可以谈激励，农业可以谈制度，为什么刑罚只能靠传统和恐惧？

这就是贝卡利亚思想的一片重要“土”：治理理性正在扩张。

在米兰的青年知识圈里，刑法不是孤立话题。彼得罗·韦里等人关心政治经济、商业、行政改革，也关心司法和酷刑。所谓“拳头学院”不是一所正式学校，而是一群年轻人聚在一起争得面红耳赤的讨论圈。这个名字本身就很形象：他们不是在抄老师答案，而是在互相撞。贝卡利亚正是在这样的碰撞中被推向犯罪与刑罚问题。

这件事对于“伟大思想家”的写法很有启发。我们喜欢把思想写成一个人的天才闪电：贝卡利亚坐在书桌前，忽然想到了现代刑法。但真正的思想生产往往更像一间吵闹厨房。有人递来材料，有人提出问题，有人讲监狱里的见闻，有人逼你把论证写清楚，有人替你检查一句话是不是太软。最后端出来的菜署一个人的名字，不等于厨房里只有一个人。

彼得罗·韦里对贝卡利亚的影响尤其重要。Peter Garnsey 2026年的剑桥研究指出，韦里既是赞助人又像导师，甚至很可能是他把贝卡利亚从数学经济问题转向刑法批判。贝卡利亚并不是多年沉潜刑法后写出巨著，而是在一个高强度知识网络中迅速进入问题、迅速阅读、迅速成文。这反而更能说明“土”的作用：思想并不只靠一个人的知识存量，还靠周围的人不断提供碰撞。

为什么这本不厚的书会迅速变成欧洲事件？因为它说中了一个正在成熟、但还缺少清楚语言的时代感觉：国家不能只用传统证明自己。

这种感觉不仅发生在刑法。启蒙时代越来越多的人开始相信，制度应该能被理性说明，权力应该服务于公共目的，法律应该可理解，政治不应只是祖传安排。贝卡利亚把这种新要求集中压到刑罚上：既然刑罚是国家对人的身体和自由最猛烈的干预，它尤其不能躲在传统背后。

这就像一家公司已经用数据管理库存、用成本核算采购、用流程优化服务，却在处理员工时仍然采用“老板看你不顺眼就罚”的办法。越是其他部门开始现代化，那个仍靠任性的角落就越刺眼。米兰的改革环境让刑罚制度的旧显得更加明显。

但这里也有一个容易误解的地方。贝卡利亚并不是从“不要国家”出发。他并不认为一切强制都是坏事。相反，他承认共同生活需要规则，需要处罚，需要社会保护。他的工作不是拆掉国家，而是给国家重新上锁。

这使他的思想既激进又克制。激进的是，他敢问国家最核心的强制从哪里来；克制的是，他的答案不是“不要处罚”，而是“处罚只能在必要范围内存在”。

这也是为什么“必要”会成为全书最值得盯住的一个词。它既是贝卡利亚限制权力的刀，也可能成为他自己的接缝。因为任何时候只要有人说“这是必要的”，我们都要继续问：谁判断？为了什么？有没有别的办法？如果“必要”本身停止接受追问，它就可能从限制权力的工具，变成新的免检章。

不过在这一阶段，先别急着批。我们先承认他的历史动作：他把刑罚拖进了改革时代的同一张桌子。财政要讲理由，商业要讲机制，行政要讲效果，那么刑罚也必须讲。国家不能在别处做现代人，在刑罚上继续做祖先。

第三章的关键不是“米兰有什么改革”，而是理解思想为什么会在错位中发生：当新旧两套治理方式同时存在，一个人忽然看见其中不协调的接缝，问题就出生了。

◆ 再看一层：制度为什么总比时代慢半拍

一个社会可以在经济上拥抱新办法，在刑罚上却长期保留旧办法，这并不奇怪。制度有自己的惯性。财政制度可以因为收入压力迅速调整，商业制度可以因为利益推动不断试验，而刑事司法往往被“安全”“秩序”“传统”包围，任何改变都容易被描述成对危险分子的纵容。于是，同一个国家会出现非常奇怪的并置：它可以在市场上相信激励，在刑罚上却只相信恐惧；在行政上追求效率，在审判上却容忍秘密和拖延。

这就是所谓路径依赖最朴素的样子：并不是没有更好的路，而是旧路上已经站满了机构、职业、习惯和正当化语言。每一个沿旧路生活的人，都可能真诚相信改变会带来混乱。

所以改革思想不能只说“新办法更好”，还必须重新安排恐惧。贝卡利亚做得高明的一点，是他没有要求社会在“秩序”和“人道”之间二选一，而是反过来证明：严酷未必更有秩序，温和也未必更软弱。只要确定性更高、程序更清楚、处罚更及时，较轻刑罚可能更能预防犯罪。这样，他就不是拿道德去和安全硬碰，而是试图证明二者可以换一种连接。

这种论证策略直到今天仍然重要。真正成功的制度改革，常常不是告诉旧制度“你太坏”，而是让它发现：你用来证明自己的目标，其实可能被另一种方式做得更好。贝卡利亚不只是道德控诉者，他还是在重新设计“秩序如何实现”的工程图。

◆ 为什么一本薄书能比一部大法典跑得更快

《论犯罪与刑罚》的传播还提醒我们：改变制度的，不一定是最厚、最完整的著作。有时恰恰是一套足够简洁的新问法先跑出去。法典擅长把答案固定，短论擅长让问题移动。贝卡利亚的文字不断使用普通人可以感受到的尺度——痛苦、恐惧、自由、安全、确定性——这让它不仅能进入法学院，也能进入更广泛的公共讨论。

制度变革往往需要这种“翻译层”。专业法律语言可以精确，却容易只在职业共同体内部循环；纯粹道德控诉可以激烈，却未必能变成规则。贝卡利亚恰好把两端接在一起：一方面提出权力来源、比例、程序等结构问题，另一方面用任何人都能理解的痛苦与安全说明为什么这些结构重要。

这也是本系列为什么要坚持“普通人都能懂”。通俗不是降低标准，而是检验一个判断是否真的抓住结构。如果一个道理只能在几十个术语包围中成立，很可能我们还没有找到它的承重梁。贝卡利亚自己的历史影响，恰好证明了清楚的语言也可以是制度力量的一部分。

思想往往不是在旧世界最完整的时候发生，而是在旧世界已经裂了一条缝、新世界刚好能从缝里照进来的时候发生。贝卡利亚就是站在这样一条缝上。

> 带走的话
>
> 思想常在新旧世界的裂缝里发生：国家不能在别处做现代人，在刑罚上继续做祖先。

> **小账**
>
> 这一章看见了什么？——米兰改革与旧刑法错位，让“为什么还这样”成了问题。
>
> 还要追问什么？——“必要”一旦停止接受追问，会不会变成新的免检章？

第 4 章

二十六岁的年轻人为什么突然插手刑罚

贝卡利亚写《论犯罪与刑罚》时非常年轻。这一点常常被当作传记趣闻：一个二十多岁的米兰贵族青年，写了一本改变刑法史的小书。可如果我们只把它写成“青年天才”，就会错过真正有意思的地方——为什么偏偏是这样一个人，在这样一个朋友圈里，在这个时间点，开始追问刑罚？

他出生于米兰贵族家庭，接受耶稣会教育，又在帕维亚学习法律。按传统路线，他完全可以成为一个熟悉制度语言的人，甚至在制度里过得不错。正因为如此，他不是从权力的最外边看刑法，而是从一个已经学会秩序语言的人内部发生怀疑。

这种怀疑往往更难。外部反抗很容易说“这套东西都是坏的”，内部怀疑却必须面对一个更棘手的问题：我明明知道制度为什么需要秩序，也知道社会不能没有规则，可为什么现有规则让我越来越难以辩护？

贝卡利亚的答案不是情绪爆炸，而是重新计算。这里的“计算”不是数学公式，而是把过去混在一起的东西分开：惩罚的目的是什么？什么程度才够？什么手段真的有效？谁有资格决定？为什么法官不能临时扩大处罚？为什么酷刑可能反而让真相更远？

这是一种非常现代的思维：不把“动机很好”当作制度正确的充分理由。国家可以说自己为了安全，法官可以说自己为了真相，立法者可以说自己为了震慑，但贝卡利亚追问的是：你的方法真的产生你说的结果吗？如果没有，目的越高尚，反而越不能替失败的方法遮丑。

在他的思想形成中，孟德斯鸠和爱尔维修等法国启蒙思想家提供了重要材料。孟德斯鸠让权力限制、法律与政治自由之间的关系进入视野；爱尔维修式的感受与

利益分析，则帮助贝卡利亚把刑罚效果放进人的实际心理反应里。再加上社会契约传统对权力来源的追问，几股力量在他这里碰到一起。

但“受谁影响”不是最重要的。最重要的是，他没有把这些思想原样搬进刑法，而是让它们在一个具体问题上重新组合。权力限制到了刑法里，就变成“法官不能任意解释和加重”；社会契约到了刑法里，就变成“处罚权只能来自最少自由的让渡”；人的感受机制到了刑法里，就变成“确定的小刑罚可能比不确定的重刑更有威慑”；公共幸福到了刑法里，就变成“刑罚不是为了让人受苦，而是为了减少将来的伤害”。

这才叫思想发生：不是把几本书装进一个脑袋，而是几个原本分开的东西，在一个新问题上重新连成结构。

我们也应该看到贝卡利亚的性格和写作处境。《论犯罪与刑罚》最初匿名出版并非偶然。对既有法律和司法制度提出系统批判，在当时不是纯粹的学术游戏。一个年轻人如果直接说许多被法律认可的做法“不必要”“残酷”“暴政”，他碰到的不只是概念，而是现实权力。匿名给了思想一层保护，也说明这个文本从一开始就在真实风险中发生。

书出版后迅速传播，进入法国启蒙圈，引发讨论并被翻译。一本小书从米兰青年圈走向欧洲，这个过程本身又反过来改变了作者和文本。原本在朋友讨论中形成的判断，突然变成各国改革者、法学家和思想家要回应的公共命题。

这是一种典型的回写：思想从环境里长出来，长成以后又改环境。

如果我们用“天才发现真理”的传统写法，会说贝卡利亚发现了罪刑相称、反酷刑、反死刑等先进原则。但用发生的眼光看，事情更有意思：旧刑罚制度提供压力，米兰改革提供对照，朋友网络提供碰撞，启蒙思想提供工具，年轻人的写作把这些力量压缩成一套新的刑罚语言；这套语言传播以后，又参与改造欧洲对刑罚的想象。

这样看，贝卡利亚不是一个孤立的“源头”，而是一个非常强的结点。很多路在他这里相遇，又从他这里分出去。

这也提醒我们，后面不能把所有现代刑法原则都简单归功于一个人。思想史不是发明专利表。贝卡利亚的价值不在于“第一个说过所有正确话”，而在于他把若干正在出现的改革直觉压成一个可以传递、可以争论、可以进入制度的结构。

第四章结束时，我们可以把他的起手式说得更准确：他不是先从“罪犯有多坏”出发，而是从“国家应该怎样证明自己的强制是必要的”出发。这个起手式，决定了后面很多章节的方向。

◆ 再看一层：不是集体写作，也不是孤独天才

强调朋友网络，并不是要把贝卡利亚个人抹掉。发生学如果走到“反正都是环境造成的”，一样会把人写死。韦里可以提出题目，可以提供材料，可以推动写作，但为什么最后是贝卡利亚把这些材料压成那种简洁、锋利、可以迅速传播的文本？这里仍然有个人选择、文字能力、胆量和判断。

所以“土”不是决定论。土提供压力与资源，却不能替人写出每一句话。同一群朋友面对同一制度，并没有写出完全相同的书；同一个时代生活着无数人，也没有人人成为贝卡利亚。发生真正有意思的地方，就在环境和个人不是二选一：环境进入个人，个人重新组织环境提供的材料，写成以后又回到环境里改变别人。

这也是本书不愿意用“时代必然产生贝卡利亚”这种说法的原因。必然会把创造抹掉。更准确的说法是：那个时代积累了让这种思想变得可能的张力，而贝卡利亚在一次具体的知识网络和个人选择中，把可能变成了文本。

把这一点写清，我们才既不会制造“天才神话”，也不会制造“环境神话”。伟大的思想不是无中生有，也不是时代自动打印出来。它是一场有人签名负责的发生。

伟大的思想很少是一个人从零开始的发现。它更像许多力量在一个人身上相撞以后，突然长出一种此前没有被这样说清楚的结构。

带走的话

贝卡利亚的起手式，不是罪犯有多坏，而是国家怎样证明自己的强制是必要的。

小账

他看见了什么？——几股启蒙思想在刑罚这一具体问题上重新连成结构。

还要追问什么？——不写天才神话，也不写环境神话，个人选择放在哪里？

第 一 编

国家手里为什么有一把刀

编 序

第一编五章。第5章讲处罚权来自为共同安全交出的有限自由；第6章追问究竟交出了多少；第7章把“必要”立为第一道锁；第8章说明超过必要的处罚就是越界的暴力；第9章追问谁有权宣布“必要”。

本编是第一条判决的主场。贝卡利亚挪动了举证责任：不再是个人求国家少罚一点，而是国家必须说明刀从哪里来、为什么非罚这么多不可。第9章同时露出第一处接缝：锁装上了，钥匙在谁手里。

读这一编时，可以带着这个问题进去：有人以“为了安全”“为了你好”限制你时，他能不能说清楚，谁给的权力，较轻的办法为什么不够？

第 5 章

处罚权不是从天上掉下来的

现在，我们正式进入贝卡利亚最重要的问题。（原典锚点：剑桥版第2章“惩罚权”。）

设想一个最简单的场景：甲偷了乙的钱。乙当然受到了损害，但这并不意味着任何路人都可以把甲关十年。即使所有邻居都很愤怒，也不意味着他们可以临时搭一个刑台处决甲。我们会说：必须有法律，必须有程序，必须由有权机关决定。可“有权机关”四个字只是把问题往后推了一步——它为什么有权？

贝卡利亚的回答从人的不安全开始。人在完全没有共同规则的状态下，每个人拥有很大的自由，却也暴露在别人的力量之下。你今天可以做很多事，别人明天也可以对你做很多事。自由很多，安全很少。为了避免这种持续危险，人们愿意牺牲一部分可以任意行动的自由，换取一个共同规则保护剩余自由。

如果用家常话说，就是一群人决定共同装一扇门。每家愿意交一点钱、接受一点门禁限制，因为大家换来了更大的夜间安全。关键在于：大家不是为了让物业变成主人，而是为了让所有住户更安全。

这就是贝卡利亚刑罚权理论里的方向：个人不是为了国家而存在，国家的强制是为了保护共同生活而出现。

这句话在今天听起来不惊人，可放到旧制度背景里，它改变了问法。以前可以从统治者、神圣秩序或传统出发说明处罚；贝卡利亚则从被处罚的人也参与构成共同体这一点出发。处罚权因此不是站在个人外部的陌生力量，而必须能够说明自己如何从共同生活中获得正当性。

更关键的是，贝卡利亚不是说人们把全部自由交出去了。恰恰相反，他强调让渡必须是最少的。为什么？因为人只有在迫不得已时才愿意放弃自由。每个人加入共同体，是为了更好保存自己，而不是为了制造一个可以随时毁掉自己的巨人。

这里已经出现现代法治的一种核心直觉：权力不是越多越好。权力只能多到完成任务所必需的程度。

把这一点放到刑罚上，逻辑非常清楚。社会为了防止成员互相伤害，需要有能力制止和处罚某些行为；但超过维护共同安全所需要的那部分痛苦，就失去了来源上的正当性。贝卡利亚因此说，凡不是出于“绝对必要”的刑罚，都带有暴政性质。我们不必把“绝对必要”当成一句永远正确的公式，但必须看到它在当时的锋利：国家第一次被要求证明，为什么这种痛苦非加不可。

这和“仁慈”完全不同。仁慈是有权多罚的人选择少罚；必要性原则则说：超过必要的那部分，你本来就没有权。

这就像医生给病人做手术。传统思路可能是“医生有资格切，所以切多少由医生决定”；现代思路则要求每一步侵入都必须有医学必要性。不是医生善良才少切，而是没有必要就不能切。贝卡利亚把类似逻辑放到国家强制上：不是国家仁慈才轻罚，而是多余的痛苦没有权利根据。

从这里，很多看似分散的主张都可以重新串起来。为什么反对过重刑罚？因为多余。为什么反对酷刑？因为在证明有罪以前把痛苦施加于尚不能称为罪犯的人，而且它并非可靠取证方式。为什么限制法官解释？因为法官如果临时扩大处罚，就超出了共同规则授予的范围。为什么质疑死刑？因为它尤其需要证明个人真的把这么大的权利交给了国家。

你会发现，贝卡利亚不是先有十条改革意见，再找一个共同理由；更像是抓住一个“权力必须有来源”的根，然后很多枝条从这个根上长出来。

这也是我们在SDE前期研究里把“国家刑罚权的正当性”确定为全书主题中心的原因。反酷刑、反死刑、罪刑相称都很重要，但它们是主场，不是总根。总根是：国家手里的刀从哪里来。

当然，这套理论立即遇到两个难题。第一个是：所谓“最少自由”怎么确定？第二个是：共同安全到底能要求个人牺牲到什么程度？这两个问题会一直跟着贝卡

利亚。因为只要“必要”没有自动刻度，就需要人来判断；只要有人判断，就可能重新出现权力。

但这并不取消他的突破。一个制度理论未必需要一次解决所有问题，才能完成历史动作。贝卡利亚完成的动作是把举证责任挪了位置。以前是个人求国家少罚一点，现在是国家需要说明为什么罚这么多。

这次挪动，看似只是一句话，实际上改变了整张桌子的方向。

◆ SDE小结：从“发现”到“发生”

在旧结构里，处罚权像一件早已存在的物品：国家有它，法官使用它。贝卡利亚把它改写成关系的结果：为了共同安全，个人让出一部分自由，才形成有限的处罚权。于是“有权处罚”不再是起点，而变成需要解释的结果。这就是本卷第一处最明确的“发生化”。

◆ 再看一层：有处罚资格，不等于有任何处罚方式

权力来源理论还有一个常被忽略的后果：即使承认国家拥有处罚资格，也不等于承认任何处罚方式都可以。资格和方式必须分开。父母有教育孩子的责任，不等于可以采用任何教育手段；医生有治疗资格，不等于可以做任何侵入；国家有维护秩序的责任，也不等于可以把一切强制都塞进“刑罚”二字。

这一区分让贝卡利亚的论证具备了层次。第一层问“能不能罚”，第二层问“谁来罚”，第三层问“罚什么”，第四层问“罚到什么程度”。旧权力最喜欢把这些问题压成一个：既然你承认社会需要处罚，那就别再挑方式。贝卡利亚恰恰把它们重新拆开。

这种拆分在现代法律里已经变成习惯，以至于我们容易忘记它当初的革命性。今天说“即使有罪也享有程序权利”“即使要处罚也不得酷刑”，听起来像常识。但这两个“即使”背后都有同一个结构：国家目标的正当性不能自动洗白实现目标的所有手段。

◆ 从“权力属于谁”到“每一步权力各凭什么”

传统政治哲学常把注意力集中在一个大问题上：主权属于谁？国王、贵族、人民，还是某种混合结构？贝卡利亚的刑罚论把这个大问题往制度细部推了一步。即使我们已经回答“主权来自人民”，仍然不能直接推出“每一种具体强制都获得了人民授权”。来源正确只是第一关，具体使用还要逐项过关。

这就像一家公司董事会由股东合法选出，并不意味着董事会可以因此查看每个员工的私人聊天。组织来源合法，与具体行为合法，是两个不同层次。现代法治的重要成熟，恰恰在于不满足于问“这个机关有没有权”，还要问“这项措施是否属于它的权限、程序是否正确、手段是否相称”。

贝卡利亚当然没有写出今天完整的行政法和宪法审查体系，但他的刑罚逻辑已经把权力从一个整体名词拆成可核对的动作。每一次关押、每一次取证、每一种刑罚，都应该有自己的理由。这种“逐项说明”比抽象说国家权力来自人民，更接近普通人真正遭遇权力的地方。

因此，本卷接下来每讲一种具体制度，都要回到这一层：不是问贝卡利亚有没有同情罪犯，而是问他如何把国家的“资格”继续拆成一项项有限权限。这正是刑法从权力说明书变成权力限制书的关键过程。

贝卡利亚不是请求国家把刀磨钝一点，而是先问：这把刀到底是谁给你的？

> 带走的话
>
> 超过必要的那部分处罚，国家本来就没有权；这不是仁慈问题，而是权限问题。

> **小账**
>
> 他看见了什么？——处罚权是个人为安全让渡有限自由后长出来的。
>
> 接缝在哪里？——“最少自由”怎么确定？必要没有自动刻度。

第 6 章

每个人究竟交出了多少自由

如果说第五章解决的是“权力从哪里来”，第六章就要面对一个更难的问题：到底交出了多少？

◆ 四种社会契约语言：霍布斯、洛克、卢梭与贝卡利亚

把贝卡利亚的“最小让渡”放进社会契约史，锋芒才真正显出来。霍布斯首先关心的是结束相互恐惧的战争状态，因此主权必须足够集中；洛克更强调人进入政治社会是为了更可靠地保护生命、自由与财产，政府仍受信托目的约束；卢梭《社会契约论》第一卷第六章则用“每个结合者连同全部权利向整个共同体作完全转让”的语言来建立共同体。三种说法都不是贝卡利亚的简单前身，它们把“为什么结合”“交出多少”“交给谁”放在不同位置。

贝卡利亚最值得注意的变化，是把这场宏大的主权论证压进刑罚权的具体边界。他并不满足于说“既然已经组成国家，国家就能处罚”，而是反过来追问：为了获得共同安全，人究竟不得不牺牲多少自由？答案不是越多越好，而是尽可能少。主权因此不是一张空白支票；刑罚权也不是从“国家存在”自动推出的全部权限。

这与卢梭形成了尤其尖锐的对照。卢梭一方面主张完全转让，另一方面在第二卷第五章为共同体在必要时要求公民付出生命留下位置；贝卡利亚恰恰从“人并没有把生命交出去”出发质疑死刑。两人都用契约语言限制私人强力，却在国家能从契约中取得什么权限上走向不同方向。后文死刑编将回到这一分歧。

因此，本书不把贝卡利亚叫作“社会契约论的应用者”就算结束。真正要抓的是契约语言内部的参数变化：霍布斯把安全压力推高，卢梭把共同体结合推到完全，洛克保留信托边界，贝卡利亚则把“只交最少必要的一份”变成刑罚权的减法原则。这个减法原则，才是他限制国家暴力时最有辨识度的动作。

社会契约最容易被说成一个漂亮故事：人人拿出一点自由，组成国家，于是国家获得处罚权。可真正决定制度边界的，恰恰是那个含糊的“一点”。一点是多少？是允许国家拘留你，还是允许国家打你？是允许国家没收犯罪所得，还是允许国家没收全部家产？是允许国家终身监禁，还是允许国家处死？

贝卡利亚的回答方向非常明确：既然人们放弃自由只是因为共同生活的需要，那么让渡应当尽可能少。换句话说，国家的正当权力不是“所有人能交多少”，而是“维持共同安全最低需要多少”。

这是一种非常重要的最小化逻辑。

我们可以拿手机权限来理解。一个手电筒应用需要相机权限，因为它要调用闪光灯；但它如果同时要求通讯录、麦克风、定位、短信和银行账户权限，你会觉得不对。为什么？不是因为这些权限绝对不能给任何应用，而是因为它们与这项功能无关。权限必须和目的相称。

贝卡利亚对国家的要求，很像给最强大的“应用”做权限管理。国家需要保护社会，所以它可以获得某些强制能力；但不能因为“安全”是一个重要目标，就自动拥有所有权限。

这时，罪刑相称不再只是“轻罪轻罚、重罪重罚”的常识，而是权力来源理论的自然结果。如果处罚超过保护共同体所需，它就不只是“不人道”，而是“没有权限”。

这一点在死刑问题上尤其尖锐。贝卡利亚问：一个人加入社会，是为了保存自己的生命和安全；那么他为什么会在同一动作中，把国家杀死自己的权利也交出去？如果说他真的同意了，那就出现一种奇怪结构：为了更好活下去，我同意别人有条件杀死我。贝卡利亚正是抓住这种张力来质疑死刑权的来源。

后来很多哲学家并不同意他。康德尤其不接受把刑罚正当性建立在效用或个人让渡上。但争论一旦进入“国家的权利根据是什么”，就已经接受了贝卡利亚制造的新战场。人们可以反驳他的答案，却很难再轻松回到“国家当然可以”。

然而，“最少让渡”也有一个难以回避的技术难题：现实社会没有一台机器，能够显示今天的共同安全需要每个人交出12.7%的自由。必要性不是温度计。

这意味着最小化原则本身不能自动完成判断。它需要立法者比较风险、手段、效果和替代方案。到了这里，贝卡利亚会越来越依赖一种理性立法者形象：冷静、一般、面向公共幸福，不被激情和私人利益控制。这个形象很重要，也很危险。因为当我们说“让理性立法者决定必要范围”时，问题又会回来：理性由谁定义？

这就是本书后面要抓的“必要性接缝”候选。贝卡利亚用“必要”把旧权力拉回边界，可“必要”一旦被某个新权威垄断，它也可能再次变成封顶词。

但是我们暂时仍然站在他这一边。为什么？因为在一个严刑习以为常的时代，“最少”本身就是方向革命。旧刑罚逻辑容易把严重性当作威慑资源：越害怕，越不敢犯。贝卡利亚则要求反过来——不是问国家还能加多少，而是问达到目的后还有哪些痛苦必须减掉。

这种思维今天仍然非常重要。刑法里经常出现一个诱惑：发生严重事件以后，舆论要求“加重”“从严”“顶格”。强烈情绪会让我们觉得严厉本身就是答案。但贝卡利亚提醒我们：刑罚不是表达愤怒的音量键。它必须接受功能与边界双重检查。

一项处罚即使能满足报复情绪，也未必能降低犯罪；即使能降低某类犯罪，也未必意味着没有更温和的替代手段；即使社会整体受益，也仍要问个体是否被迫承担了超出共同生活所需的牺牲。

这三层检查——有效吗、必要吗、相称吗——正是现代比例原则式思维的远亲。我们不能把今天所有精密法理都倒算成贝卡利亚已经完整拥有，但可以看到他的方向：国家越强，越需要说明为什么不能少做一点。

第六章到这里，社会契约终于不再只是哲学名词。它变成了一张权限表。每一项刑罚都必须问：这一项权限，个人真的为了共同安全交出去了吗？如果没有，国家不能靠“大局”两个字补签。

◆ 再看一层：“最小让渡”不等于“国家越小越好”

“只让渡最少自由”很容易被读成一个简单政治口号：国家越小越好。其实贝卡利亚的逻辑没有这么机械。关键不是国家规模，而是强制与共同目的之间的关系。有些社会如果没有稳定警察、公开审判和有效执行，表面上国家干预很少，实际个人反而更容易被私人暴力、报复和任意权力支配。为了减少这种支配，社会可能需要更强的公共能力。

所以最小让渡并不是“什么都别管”，而是“每一种强制都必须有对应理由”。有时为了让规则确定、公平执行，国家需要建立更完善的机构；有时为了保护个人，国家需要制止私人复仇。国家能力可以增强，同时任意权力可以缩小。两者并不矛盾。

这也解释了贝卡利亚为什么既反对残酷刑罚，又强调处罚的确定性。他并不希望国家软弱到法律无人执行。恰恰相反，他要的是一种“更少但更准”的国家：少一些任意痛苦，多一些稳定规则；少一些恐怖表演，多一些可预期执行。

这种区分对今天也有价值。讨论刑罚时，我们经常把“严厉”与“有力”混为一谈，把“温和”与“软弱”混为一谈。贝卡利亚逼我们拆开：一个国家可以处罚较轻，却执行得稳定；也可以法条极重，却大量选择性执法。真正的强，并不一定写在刑期数字上。

自由不是一次性交给国家的整块财产，而更像一项项有限授权。贝卡利亚真正想做的，是把国家的每一项处罚权限重新拆开核对。

> 带走的话
>
> 自由不是一次交给国家的整块财产，而是一项项有限授权，每一项都要核对。

> **小账**
>
> 他看见了什么？——契约变成权限表，超出必要的刑罚就是没有权限。
>
> 接缝在哪里？——必要性不是温度计，理性立法者由谁来定义？

第 7 章

“必要”：贝卡利亚给国家上的第一道锁

如果要从《论犯罪与刑罚》里挑一个最值得贴在整本书墙上的词，我会选“必要”。

不是“残酷”，不是“死刑”，甚至不是“自由”。因为贝卡利亚真正让国家不舒服的地方，是他不断要求：你做这件事，到底有没有必要？

“必要”听起来很普通。可把它放进权力关系里，它会立刻变成一把刀。一个父母说“我是为你好”，一个公司说“我是为了效率”，一个学校说“我是为了纪律”，一个政府说“我是为了安全”，这些目的都可能是真的，但目的正确并不能证明手段当然正确。必要性要求在目的与手段之间再加一道检查：不用这么重的办法，目标是不是也能实现？

这道检查让权力第一次必须面对替代方案。

比如为了防止学生考试作弊，学校可以加强监考，也可以没收所有学生手机，还可以要求所有人接受全天候摄像。三个办法可能都能减少作弊，但它们对自由和尊严的侵入完全不同。如果较轻手段已经足够，最重手段就很难说“必要”。

贝卡利亚把这种直觉放进刑罚：处罚不是越重越稳，而应当只到防止新的损害所需要的程度。超过这个程度的痛苦，不增加正当性，只增加权力的剩余。

这也是为什么他那么重视刑罚的确定性。旧逻辑容易把“威慑”理解成痛苦强度：砍得越狠，大家越怕。贝卡利亚却发现，人实际作决定时，会同时考虑另一件事——会不会真的被抓、真的被罚。一个极重但几乎逃得掉的处罚，可能不如一个温和但高概率发生的处罚有影响。

这意味着刑罚效果不是一个点，而是一条路。行为人面对的不是刑法典里孤零零的一行刑期，而是一整串经验：规则是否清楚、执法是否稳定、案件是否及时处理、处罚是否真的执行、身边人是否经常逃脱。

这就是贝卡利亚厉害的地方：他没有只在“刑罚有多痛”上做加减法，而是开始看整个差异序列。用SDE的大白话说，他从“显露”走向了“差异序列”。

一旦看到路，很多严刑逻辑就会松动。国家如果执法松散，却不断加重法定刑，可能只是把纸上的恐怖越写越大，而现实中的确定性并没有提高。一个社会如果大量行为被规定为严重违法，但大家长期看到“有关系的人没事”，那么再严厉的条文也会被日常经验抵消。

贝卡利亚甚至进一步指出，刑罚过于残酷可能产生反效果。人对极端痛苦会变得麻木，执法者也可能因为处罚太重而不愿严格适用，公众可能同情被处罚者，最终反而增加逃避与不执行。也就是说，重刑不仅有成本，还可能破坏自己的目标。

这是一种非常现代的制度分析：规则会回写执行者，执行方式会回写公众，公众反应又会回写规则的实际效果。刑罚不是立法者写完以后单向落到社会上的石头，它进入社会以后会引起新的行为。

然而，正是在“必要”这个词上，我们必须开始提高警惕。

因为“必要”非常容易变成权力最喜欢的自我证明。历史上许多扩大监控、延长拘禁、限制程序保障的措施，都可能以“特殊时期的必要”为理由。必要性如果没有比较、证据、期限和复查，就只是换了一个更聪明的“我认为需要”。

贝卡利亚本人试图用公共幸福、最小痛苦、犯罪预防等标准约束必要性，但这些标准仍然需要解释。什么叫公共幸福？少数人的权利能否为了多数人的安全被无限压缩？长期安全与即时恐惧如何比较？有些处罚虽然能提高威慑，却会不会损害社会信任？

所以，“必要”既是贝卡利亚的第一道锁，也可能是全书最值得检查的锁眼。

我们可以暂时把它画成一扇门：旧制度的门没有锁，国家想推多远就推多远；贝卡利亚装上一把“必要”锁。这个动作无疑是进步。但谁拿钥匙？如果钥匙还是只在国家手里，锁会不会变成装饰？

现代法治后来发展出更复杂的比例原则、程序审查、证据要求、司法救济，某种意义上都在回答这个问题：不能让权力自己宣布“必要”以后就结束讨论。

◆ 必要性为什么必须带着“失败条件”

一个真正可检验的“必要”判断，必须允许自己失败。比如我们说“只有延长拘禁才能防止证据灭失”，那就应该说明：在哪些案件、什么证据条件、延长多久、由谁复查；如果较短期限已经足够，原来的“必要”就应退出。反过来，如果一句“情况特殊”可以无限续期，必要性就不再是判断，而成了护身符。

这和科学假说很像：一个理论如果无论出现什么结果都说自己对，它就无法被真正检验。权力判断也是一样。必要性必须写出“在什么条件下我不再必要”。这就是后面本书会反复使用的“推翻条件”。

把这条要求反过来用到贝卡利亚自己身上也很重要。我们不能只说“贝卡利亚认为必要才可处罚”，还要问他的必要性标准能否处理复杂现实，是否会把某些价值偷偷排除在计算之外。SDE的纪律是：不能让用来批别人免检的词，到了自己这里变成免检词。

因此，“必要”不是终点，而是一道要求持续说明的程序。它最现代的意义，或许不在于给出答案，而在于让答案永远带着期限和条件。

因此，我们对贝卡利亚最好的尊重，不是把“必要性”背成名言，而是继续按照他的精神追问必要性本身。

◆ SDE小结：一把锁和一个接缝

“必要”把国家从无限强制拉回目的约束，这是贝卡利亚最有力的开门；但如果必要性被当成无需继续说明的最终词，它又可能成为新的封顶。后续全书会专门检查：贝卡利亚有没有给“必要”留下足够的发生空间。

◆ 再看一层：为什么“确定性”比“严厉性”更接近现代治理

贝卡利亚强调确定性，还有一个比威慑更深的意义：它把治理从偶尔爆发的权力，变成持续可预期的规则。严厉往往依赖戏剧性——一次重判、一次公开震慑、

一次轰动性行动；确定性则依赖日常制度——清楚规则、稳定执法、及时审判、相对一致的适用。前者让人看见国家有多凶，后者让人知道规则会怎样运行。

从社会心理看，这两种秩序完全不同。如果大家认为法律平时可以绕，只有碰到运动式治理才突然严惩，人们学习到的不是“遵守规则”，而是“判断风向”。如果规则稳定而处罚适度，人们才更可能把法律当作可预测环境的一部分。

因此，确定性不仅是犯罪人的成本计算，也与普通人的信任有关。一个处罚不重但程序稳定的制度，会让守法者知道别人也受同样规则约束；一个处罚极重却执行随意的制度，则可能让人害怕权力而不相信法律。贝卡利亚没有用今天“制度信任”的术语完整展开这一点，但他的方向已经触到这里。

这也帮助我们理解为何他对赦免、任意裁量等问题保持警惕。偶尔的仁慈表面上很温暖，却可能说明法律本身过于严酷或执行不稳定。一个制度如果必须不断靠个人恩典修正自己，就说明规则设计可能有问题。现代法治更希望善意进入一般规则，而不是等某个好人临时救场。

“为了安全”不是答案，只是问题的开始。真正的法治，要继续问：为什么非得用这个办法，为什么不能少一点？

> 带走的话
>
> “为了安全”不是答案，只是问题的开始：为什么非得用这个办法，为什么不能少一点？

> **小账**
>
> 他看见了什么？——威慑是一条路，确定性比严厉更能影响行为。
>
> 接缝在哪里？——“必要”是第一道锁，但钥匙若只在国家手里呢？

第 8 章

如果没有必要，处罚就是另一种暴力

到这里，我们终于可以理解贝卡利亚那句极具攻击性的判断：不是出于必要的刑罚，就是暴政性的。

今天听这句话，很容易把它当作十八世纪启蒙作家的漂亮修辞。但如果把前面几章连起来看，它其实是一个非常严格的逻辑结果。

第一步，个人并不是为了让国家变得强大才组成共同体，而是为了在共同生活中保存更多安全与自由。第二步，因此个人只会让出达到共同目的所必需的那部分自由。第三步，国家处罚权只能从这部分让渡中产生。第四步，凡超过必要范围的处罚，就没有来源。

所以，“多余的刑罚”不是简单的政策失误，而是权力越界。

这和我们评价一个人的暴力非常相似。假设有人为了阻止另一个人正在伤害孩子，必要时可以使用一定强制。但危险已经解除以后，他继续殴打对方十分钟，这十分钟的性质就变了。前面的力量可能是制止，后面的力量变成报复。区别不在拳头本身，而在必要性已经结束。

贝卡利亚把这种边界意识放到国家身上。国家不能因为最初有一个正当目的，就让所有后续强制自动变得正当。

这使刑罚的讨论第一次具有一种“剩余审查”：每增加一分痛苦，都要问它有没有额外作用。没有作用的痛苦，就是纯粹剩余。

这也是为什么他会反对把折磨当作惩罚美学。旧刑罚有时追求一种象征对称：犯罪越严重，处罚越要让人震撼；冒犯越大，国家越要显示力量。贝卡利亚却把问题拉回功能：这种痛苦究竟防止了什么？如果只是满足观看者的愤怒，它不能自动进入正当刑罚。

这里有一个非常重要的现代转折：刑法不再首先是一门“表达愤怒”的学问，而逐渐变成一门“控制强制”的学问。

很多人会说，这样是不是把刑罚说得太功利了？犯罪人明明做了恶，难道不该受到应有报应吗？这正是后来康德等人与贝卡利亚拉开距离的地方。贝卡利亚更关心刑罚是否能够防止未来损害，康德则强调不能把人只当作实现社会目的的手段，刑罚必须与其应得相关。

这场争论后面会专门展开。此刻我们只需要看到：贝卡利亚把刑罚从“过去发生了什么，所以必须报复”转向“未来怎样少发生伤害”。时间方向变了。

一旦时间方向变了，处罚的轻重就不能只看过去罪行的道德愤怒，还要看未来效果。酷刑如果不能可靠发现事实，就没有必要；死刑如果并不比长期剥夺自由更有效，就很难仅靠恐惧说明必要；极端重刑如果降低执法确定性，就可能适得其反。

这是一整套围绕“效果—必要—限度”展开的链条。

但我们也必须指出其中的第一处深层风险：如果所有处罚都只用未来效果证明，那么个人会不会变成社会治理的工具？假设公开重罚一个人可以让一万人更守法，是否就一定正当？如果答案只是“只要总体收益够大”，那就可能牺牲个体。贝卡利亚本人的自由让渡理论正是在努力防止这种滑坡：国家不是因为总体幸福大就可以任意对待个人，因为国家的权利来源本来就受个人让渡范围限制。

这说明贝卡利亚并不是一个简单的“结果最大化”思想家。他的结构里至少有两股力量同时存在：一股是公共幸福和预防效果，一股是个人自由与权力来源。

这两股力量以后会不会冲突？一定会。

例如，假设某种高强度监控可以大幅降低犯罪，却严重侵入每个人的私人生活。公共安全提高了，自由却减少了。贝卡利亚的理论应该怎么选？如果他坚持最少让渡，就不能只看安全收益；如果他坚持最大公共幸福，又可能要求更多让渡。

这正是我们前期研究时没有急着把他判成“自由中心主义”的原因。真正的母题必须经过敌意检验：那些看起来支持幸福中心的原文，能不能推翻“自由是根”的判断？如果能，我们就改判。

一本好的解构书不应该把作者变成被告以后，所有证据只往一个方向收。我们必须让贝卡利亚有机会把自己的判词推翻。

不过，到第八章为止，前文正文已经可以给出一个阶段性结论：贝卡利亚的革命至少完成了三次方向转换。

第一，从“国家有权罚”转向“国家必须说明权从哪里来”。第二，从“刑罚越重越有力”转向“刑罚必须接受效果与必要性检验”。第三，从“犯罪以后怎样报复”转向“怎样让未来少发生犯罪”。

这三次转换加在一起，才构成《论犯罪与刑罚》的真正现代性。

他当然没有完成现代刑法。无罪推定、程序正义、比例原则、司法审查、权利保障后来都经历了漫长发展。把所有成果倒贴到贝卡利亚身上，会把历史写成英雄神话。但如果没有人把“国家当然有权”重新变成问题，后面的制度精细化也缺少一个重要起点。

所以本章最后，我们再回到那个最简单的问题：一个人犯罪以后，国家为什么不能想怎么罚就怎么罚？

贝卡利亚的回答不是“因为国家应该善良”，而是：因为国家手里的权力本来就是有限生成的。它不是自然现象，不是神秘属性，也不是胜利者的永久财产。它只在共同生活的必要范围内成立。

这句话的含义，比“刑罚温和一点”大得多。它把刑法从国家用来管人的工具，变成一套同时用来管国家怎样管人的规则。

◆ 再看一层：暴政不是“很残酷”的同义词

贝卡利亚把多余刑罚称为暴政，最值得注意的并不是修辞强度，而是他悄悄改了“暴政”的定义。暴政不只发生在统治者特别凶残时，也可以发生在权力越过自

己正当来源的时候。这样一来，一项处罚即使看起来不血腥，只要缺乏必要根据，也可能带有暴政性质；相反，一项给人造成真实痛苦的强制，如果确实是防止更大侵害的最小手段，也不能仅凭“让人难受”就判定为不正当。

这使评价标准从感官转向结构。我们不再只看“痛不痛”，而要看“为什么、由谁、通过什么程序、有没有替代、到什么限度”。这正是现代权利审查比单纯人道主义更复杂的地方。

同时，这种结构也防止我们把贝卡利亚浪漫化。他并不排斥刑罚本身。他接受国家在必要范围内制造痛苦、剥夺自由。这意味着他的思想从一开始就活在一个张力里：怎样既承认公共强制的必要，又不让必要变成无限授权？

后面的全书，真正要研究的就是这条张力如何运行。只赞美他的温和，会丢掉国家强制这一半；只批判他的功利，又会丢掉自由来源这一半。贝卡利亚之所以值得一整本书，正因为这两股力量没有轻易合并，它们一直在彼此牵制。

◆ 阶段性判词

贝卡利亚最重要的革命，不是要求国家少罚，而是改变了谁来证明：从“个人为什么不该受罚”，变成“国家为什么必须这样罚”。

◆ 阶段小结　八章之后，我们已经走到了哪里

写到这里，我们还没有进入罪刑比例、法官解释、酷刑、死刑这些最著名的贝卡利亚主场，甚至还没有真正对他下总判词。可前八章并不是“铺垫”。它们完成了一件更重要的事：把后面所有具体主张的共同地基挖出来。

这块地基不是“人道”。如果只说贝卡利亚希望刑罚更人道，就会把他和许多同情受刑者的人放在同一层。人道当然重要，却解释不了为什么他会同时关心法官解释、刑罚确定性、处罚及时性、死刑权来源、秘密控告和犯罪预防。真正把这些问题连起来的，是对强制权来源和限度的追问。

前八章已经看见三条线。第一条是来源线：处罚权不是国家天然拥有，而必须从共同生活的需要中说明。第二条是限度线：即使有权处罚，也只能使用为共同安全所必要的那部分强制。第三条是效果线：一种刑罚不能因为足够严厉就自动正当，还要看它在真实行为路径中是否有效。三条线一交叉，贝卡利亚就不再只是“温和刑罚的倡导者”，而成为一个重新设计国家强制的人。

这三条线之间也已经出现张力。来源线把个人自由放得很高，效果线却不断诉诸公共幸福和犯罪预防；限度线用“必要”限制国家，却又必须让某个立法者判断什么叫必要。换句话说，贝卡利亚每打开一扇门，也会留下新的门框。SDE下一步要做的，不是急着说门框就是错误，而是检查：哪些只是工作所需的暂时支架，哪些后来被悄悄当成了地基。

因此，后文正文将真正进入“接缝”。第9章会追问：谁有权宣布某项刑罚是必要的？第10章以后则转向犯罪的尺度和罪刑比例。我们会看到贝卡利亚试图把“国王受了多大冒犯”换成“社会受了多大损害”，这又是一次重要发生化；同时也会追问“社会损害”是不是能够被独立、稳定地计算。

再往后，法官解释和酷刑将把问题推到更细。法官是只负责把事实装进法律格子，还是会参与法律意义的形成？酷刑是在发现真相，还是在制造口供？到这些地方，贝卡利亚会从政治哲学走进知识论：国家不仅在决定谁该受苦，也在决定什么可以被当作事实。

前文写作还有一个方法上的收获：我们已经收回了一个过早的判断。前期研究阶段曾考虑把贝卡利亚的“收回动作”命名为“量刑尺”，仿佛他把复杂社会重新压成可以精确计算的尺度。但原典显示，他自己已经承认人的激情和利益过于复杂，政治判断不能达到几何学式精确。既然证据反对我们，就必须改判。这个收回不是失误，而是本系列必须保留的诚实机制：不能为了保住一个漂亮母题，牺牲哲学家自己的原文。

目前更值得继续验证的，是“先放人，再限权”。贝卡利亚把国家权力从天上拉回关系发生，却需要一个能够先于国家拥有自由、再把其中一部分交出的个人。

如果后面的原典证明，他其实已经充分看见人的社会形成过程，那么这个判词就必须再次收回；如果没有，它才有资格进入总解构。

所以，前文不是把贝卡利亚判完了，而是完成了开门：刑罚权有出生证明；国家强制有权限表；“为了安全”必须继续举证；严厉不等于有效；目标正确不能洗白所有手段。到这里，我们才真正有资格进入《论犯罪与刑罚》的核心房间。

如果把这八章压成普通人能立刻使用的一种判断习惯，它其实很简单：凡是有人准备以“为了你好”“为了安全”“为了秩序”为理由限制别人，先不要急着争论动机好坏，而要把问题拆成四步——谁给了这项权力，目标是什么，较轻办法是否足够，怎样证明结果确实有效。贝卡利亚两百多年前处理刑罚，我们今天可以把同样的问法带回学校、单位、家庭和公共生活。

这也是全书接下来不会放掉的一条纪律：每当我们准备赞美贝卡利亚时，都去找反对他的原文；每当我们准备批判他时，都先把他的最好辩护写出来。只有这样，后面的总判词才不是提前写好的答案。

八章之后，本书只确定一件事：国家可以处罚人，不应当是刑法的起点，而应当是刑法必须不断证明的结论。

> 带走的话
>
> 国家可以处罚人，不是刑法的起点，而是刑法必须不断证明的结论。

> **小账**
>
> 他看见了什么？——多余的刑罚不是政策失误，而是权力越界。
>
> 接缝在哪里？——公共幸福与最少让渡两股力量冲突时，谁让步？

第 9 章

第一处接缝：谁有权宣布“必要”

“只有必要的刑罚才是正当的。”这句话听起来几乎无可反对。一个国家如果能用较轻的手段维护秩序，却偏偏选择更重的手段，强制就开始越界。前文已经看到，贝卡利亚正是用“必要”给国家的处罚权加上第一道锁。问题是，锁一旦装上，新的问题马上出现：钥匙在谁手里？谁有资格宣布某种刑罚已经到了“非用不可”的程度？

这是本书第一次真正进入“接缝”。所谓接缝，不是抓一句话证明作者前后矛盾，而是找一个思想从开放状态变成稳定结构的地方。贝卡利亚打开了“国家权力必须说明理由”的门，但要让这套制度真的运转，他又必须让某个判断暂时稳定下来。否则每一次判刑都要从头讨论国家是否有权处罚，刑法根本无法成为公共规则。问题不在于稳定本身，而在于稳定之后是否还允许继续追问。

先用一个最普通的例子。小区为了防止电动车进电梯，物业可以采取很多办法：张贴提醒、安装识别设备、设置集中充电点、罚款、停用门禁，甚至把违规者的车直接锁起来。物业说：“这是为了安全。”居民完全可能承认安全目标正当，却继续追问：为什么一定要锁车？有没有更轻办法？锁多久？谁认定存在危险？有没有申诉？“为了安全”只回答了目标，没有自动回答手段。

贝卡利亚的“必要”也是这样。必要性并不是一句神奇咒语，而是一个比较关系。只有在至少两个方案之间比较，才谈得上哪一个“必要”。如果根本不展示替代方案，所谓必要常常只是“我决定这样做”的另一种说法。因此，必要性原则真正的锋利之处，不是把“必要”写进法律，而是逼权力说明：目的是什么、备选手段有哪些、为什么轻的手段不够、重的手段增加了什么效果。

但贝卡利亚所处的十八世纪，并没有我们今天已经熟悉的比例原则审查、违宪审查或成熟的行政法技术。他面对的是更原始也更危险的局面：法官、行政者、君主以及各种旧习惯都可能以秩序之名动用强制。因此他必须先解决一个更基础的问

题——不能让判断必要性的权力散落在每一个具体裁判者手里。于是他把越来越多的决定提前交给法律和立法者。

这一步有充分理由。假如“必要”完全由法官在个案中判断，同样的行为就可能因为法官性格、身份、情绪、社会地位关系而得到不同结果。一个法官觉得严罚才能震慑，另一个法官觉得宽缓更有效；一个同情贵族，一个同情平民；一个刚遭遇盗窃，另一个正好心情平和。此时，“必要”不是限制任意，而可能成为任意的入口。贝卡利亚如此厌恶司法解释，与这一恐惧有直接关系。

所以，他的制度选择是把必要性尽量前置：由一般法律先规定什么行为是犯罪、对应什么刑罚，法官只判断事实是否落入法律已经划出的格子。这样做，至少把“今天这个法官觉得有必要”改成“社会事先以法律形式认为有必要”。这就是他心中的自由保障：不是靠法官的善良，而是靠规则在案件发生之前就存在。

可接缝也在这里出现。贝卡利亚一方面告诉我们，法律和刑罚都来自社会成员为了共同安全而形成的安排；另一方面，当这种安排写成法律以后，他希望它尽可能稳定、明确，不再在法官手中继续生长。用SDE的大白话说，前半段他很会讲“这套制度是怎么长出来的”，到了裁判现场，却更希望已经稳定的显露保持不动。

这不是简单的“自相矛盾”。一个制度如果永远处在正在成形状态，人们无法预期后果；但一个制度如果一旦定型就不许现实重新进入，新的不公平也会被一起固定。真正困难的是：规则必须足够稳定，才能限制权力；规则又必须留出某种更新机制，才能处理新事实。现代法治仍然没有彻底摆脱这个两难。

因此，第九章不能匆忙下判说“贝卡利亚迷信立法者”。更公平的说法是：他把自己时代最危险的任意主要识别在司法端，于是选择把不确定性往立法端搬。今天我们更容易看到，立法者同样会判断错误、信息不足、受利益影响，法律文字也可能产生空白。但是在贝卡利亚面对的历史环境里，先把法官从“半个立法者”位置上拉下来，本身就是巨大的自由工程。

问题也因此更准确了：贝卡利亚没有消灭“谁来判断必要”这个难题，他只是重新分配了判断的位置。过去的答案可能是“法官、君主和传统随时决定”；他的答案更接近“让社会通过一般法律预先决定”。这显著减少了个案中的随意，却把新的压力集中到一个地方：立法者凭什么知道什么最必要？

这就是我们要留到后面继续攻击的第一处接缝。“必要性盖章”暂时不作为总收回动作，但它成为一个重要候选：当贝卡利亚用必要性拆掉旧的残酷刑罚时，必要是批判工具；当他开始设计新刑法时，必要又可能变成给新制度盖章的理由。两种用法之间有没有足够透明的转换程序，决定了“必要”到底是一把锁，还是一枚印章。

普通读者可以把这一章带回生活。任何人说“这是必须的”，先别急着争论他是不是坏人。只问四句：必须为了什么？有没有别的办法？为什么别的办法不够？谁来复查“必须”这个判断？这四句就是把“必要”从口号变成程序。贝卡利亚给国家装了第一道锁；我们今天能做的，是继续检查这把锁有没有被同一只手同时握住钥匙。

> **带走的话**
>
> 贝卡利亚用“必要”限制权力，却没有让“必要”自动获得真理资格。真正的难题不是要不要必要性，而是谁来证明“非这样不可”。

> **小账**
>
> 他看见了什么？——由法官个案判断必要，必要会成为任意的入口。
>
> 接缝在哪里？——判断搬到了立法端，立法者凭什么知道什么最必要？

第 二 编

犯罪不是国王生气

编 序

第二编五章。第10章把犯罪的尺度从“冒犯谁”换成“伤害什么”；第11章拒绝按“坏心眼”量刑；第12章讲轻罪重罪为何不能同罚；第13章说明比例不是数学题；第14章把量刑尺改造成一张有主干刻度的公共地图。

本编是第二条判决的主场。刑罚从情绪化报复，改写成受公共理由和比例约束的制度行为：犯罪要说明损害了什么，刑罚要守住轻重顺序。本编也正式撤销了“贝卡利亚以为刑罚可以精确计算”的指控。

读这一编时，可以想想家里、学校或单位：是不是每次犯错都被说成“最严重”？一套没有刻度的规则，会教人学到什么？

第 10 章

到底什么叫犯罪

人们很容易以为“犯罪是什么”比“怎么处罚犯罪”简单。杀人、抢劫、盗窃，这些不是一眼就能看出来吗？但只要把时间拉长，就会发现答案并不稳定。某个时代禁止的行为，另一个时代可能不再禁止；某些地方被称为冒犯神圣秩序的行为，在另一些地方根本不进入刑法。犯罪当然包含真实行为和真实损害，可“什么行为值得国家动用刑罚”从来不是自然界自动贴好的标签。

贝卡利亚在这里做了一次极重要的换尺。他反对用被冒犯者的身份高低衡量犯罪，也反对把宗教意义上的罪孽直接等同于国家可以处罚的犯罪。他把尺度转向一个更世俗、更公共的问题：这种行为对社会造成了什么损害。原典里那句著名判断很明确——衡量犯罪的真正尺度，是对社会造成的损害。（原典锚点：剑桥版第7章“衡量刑罚的错误”，第8章开头重申。）

为什么这一步重要？想象两个社会。第一个社会把“让国王难堪”看成比伤害普通人严重得多的罪；第二个社会要求法律先说明一个行为怎样破坏了共同安全、财产、生命或公共制度。两者都可能有刑法，但它们的刑法不是同一种东西。前一种刑法围绕身份秩序旋转，后一种至少尝试围绕共同生活旋转。

贝卡利亚把“国王被冒犯多少”换成“社会实际受损多少”，等于把刑法从垂直等级关系往横向公共关系拉了一步。一个行为的严重性，不应因为受害者头衔更高就自动膨胀。这一改变的意义，不只是保护平民，也是在改写国家与犯罪的关系：刑法不是替强者维护尊严的私人武器，而应当说明它维护的是所有人共享的公共条件。

但“社会损害”并不是一个透明玻璃杯。谁是“社会”？什么算损害？损害只包括已经发生的身体和财产后果吗？对制度信任的破坏算不算？制造恐惧算不算？长期污染环境算不算？操纵金融市场造成的系统风险算不算？如果我们把今天的问

题倒推回去，就会发现“社会损害”比“君主尊严”进步，却并不因此变成无需解释的自然量。

贝卡利亚自己其实知道社会关系复杂。他在谈罪刑比例时明确承认，人类激情、利益和行为组合不能用几何学精确安排。这一点非常重要，因为它提醒我们不要把“社会损害”误读成一台仪器能直接测出来的数字。它更像一套公共排序：哪些损害会动摇共同生活的基础，哪些较轻，哪些虽然令人反感但不应动用最严厉的国家强制。

这使“犯罪”第一次显出发生结构。一个行为不是先天带着“犯罪”二字，而是在法律与公共秩序的关系中被定义。定义以后，它又会回头改变人的行为。例如税率越高、边境越难监管，走私的诱因可能随制度设计改变；法律如果禁止过多日常行为，也会扩大“违法者”的数量。贝卡利亚后来甚至直接提醒，某些犯罪是法律自己制造出来的。

这是一条非常现代的线索。我们常把刑法理解成警察拿着手电筒，在黑暗里“发现坏人”。贝卡利亚让我们至少看到另一面：法律不是只有探照灯，它还有刻刀。它一边发现已经存在的伤害，一边划定哪些行为从此属于国家可以强制介入的领域。刻刀刻得越宽，犯罪的边界就越宽；刻得太细，严重伤害又可能漏掉。

所以，问“什么叫犯罪”不能只答“法律规定的就是犯罪”，否则所有残酷刑法都天然正确；也不能只答“道德上坏的就是犯罪”，否则国家会把道德争议无限刑罚化。贝卡利亚的贡献，是在两者之间插入“公共损害”这块地板：刑罚必须面对共同生活受到的真实伤害，而不能只是替宗教、身份或个人愤怒服务。

但这块地板仍然会移动。一个农业社会、商业社会、数字社会对“公共损害”的理解不会完全相同。盗窃一袋粮食、窃取数百万人的数据、传播恶意程序、操纵算法造成大规模歧视，它们并不共享同一种物理伤害，却都可能破坏社会合作的条件。贝卡利亚给了我们方向，却没有也不可能给出两百多年以后所有损害的清单。

这里需要特别防止一种误读：如果“社会损害”是核心，就可以为了社会利益牺牲任何个人吗？不可以。贝卡利亚的整个体系同时受另一条线约束——个人只交出维持共同安全所需的最少自由。也就是说，社会损害不是一张无限授权书。它只能说明国家为何可能介入，还必须再通过必要性、比例、法律预先规定等闸门。

这也是为什么“社会损害”比“公共利益”四个大字更难滥用。公共利益如果不拆开，很容易成为统治者替自己说话的宏大名词；社会损害则要求回到具体关系：谁受到什么损害，损害如何发生，为什么需要刑罚而不是民事、行政或社会治理手段，刑罚会不会制造更大的二次损害。贝卡利亚没有用今天的术语提出这一整套问题，但他的换尺已经把路打开。

到这里，我们可以给犯罪一个更适合本书的通俗定义：犯罪不是“一个坏人做了坏事”这么简单，而是某种行为被共同体判断为严重破坏共同生活条件，并因此决定动用最强强制之一去回应。这个定义里有行为、有损害、有规则，也有制度选择。它比道德谴责冷一点，却更适合法治，因为刑罚不能靠愤怒决定边界。

贝卡利亚在这一章真正完成的，是把犯罪从“冒犯谁”改成“伤害什么”。前者把目光盯在地位，后者把目光转向关系。只要这一步成立，后面的罪刑比例就有了可能：不是因为贵族比平民重要，所以罚得更重，而是因为不同犯罪对共同生活造成的损害不同，所以必须建立梯度。下一章我们就要碰上新的难题——如果不看一个人的“坏心眼”，刑法还能不能把故意、过失和偶然区别开？

> 带走的话
>
> **贝卡利亚把犯罪的尺度从“谁被冒犯”转向“共同生活受到什么损害”。犯罪因此不再只是道德标签，而成为国家必须说明公共理由的一种法律判断。**

> **小账**
>
> 他看见了什么？——法律不只是探照灯，也是刻刀，某些犯罪由法律制造。
>
> 接缝在哪里？——“社会损害”不是透明玻璃杯，谁来界定损害？

第 11 章

为什么不能按照一个人的“坏心眼”量刑

说犯罪要看社会损害，马上会遇到常识反击：难道一个人是故意杀人还是完全意外，都不重要吗？难道抱着恶意但没有造成严重后果的人，就一定比造成重大后果却没有恶意的人更轻？今天受过基本法律教育的人都知道，主观心态在刑法里非常重要。因此如果把贝卡利亚的“不能以意图作为犯罪真正尺度”理解成“故意过失都不分”，会把他写成一个比实际更粗糙的人。（原典锚点：剑桥版第7章。）

他真正反对的是另一件事：把一个人心里到底有多邪恶，当成可以直接测量并据此安排刑罚的最终标尺。原典解释得很清楚，意图会随每个人当时的印象、心理状态、激情和处境而变化；如果要按照这种内在状态精确量刑，就仿佛必须为每个人、每一次犯罪重新制定一部特别法律。

这句话背后的恐惧仍然是任意。人的内心并不像桌上的尺子。法官如果可以说“我感觉他恶性特别深”，量刑就会迅速滑向人格判断。一个衣着体面、表达顺畅的人可能被看成“可教化”；另一个沉默、粗鲁、贫穷的人可能被看成“危险”。表面上是在判断内心，实际上很容易把阶层偏见、性格偏好和社会成见带进刑罚。

贝卡利亚因此宁愿把刑法重心拉回可公开讨论的行为及其社会后果。因为行为和损害至少能进入共同证据，而“灵魂有多坏”无法由国家可靠读取。他还特别区分了人与人关系和人与上帝关系：宗教意义上的罪孽可以涉及内心，但国家刑法没有资格假装自己能像上帝一样看穿一个人的全部道德价值。

这一点非常锋利。刑法的任务不是给一个人做完整道德鉴定。一个人在生活中可能自私、虚荣、冷漠，这些都不自动成为犯罪；另一个人可能出于看似良好目的做出造成巨大公共损害的行为，也不能仅凭“初心是好的”免除法律评价。贝卡利亚说，有时最好的意图会造成最大的社会恶果，有时极坏的意图反而没有造成相应损害。法律若只追心，就会失去公共尺度。

但我们今天当然还要追问：主观责任怎么办？答案不是回到“坏心眼量刑”，而是把主观心态放在更有限的位置。故意、明知、轻率、疏忽，之所以重要，不是因为法官终于看见了一个人的灵魂颜色，而是因为这些心理状态与行为控制、风险认知、可责性和预防方式有关。也就是说，主观要素可以进入法律结构，却不等于国家获得了全面审判人格的权力。

用大白话说，刑法可以问：“你知不知道会发生什么？你能不能避免？你是不是有意让它发生？”但要非常谨慎地问：“你到底是不是一个坏人？”前一组问题和具体行为相关，后一问则容易把人整个吞进刑罚。贝卡利亚对“意图尺度”的警惕，正是在替这条边界寻找早期表达。

一个学校例子能看得更清楚。两个学生都打碎一块玻璃。甲故意拿石头砸，乙在体育活动中不慎撞碎。如果学校只看损失金额，显然不公平；但如果老师说“甲这个孩子骨子里坏，所以不仅赔玻璃，还要停学半年”，又把一次行为扩展成对整个人格的审判。合理的处理既要区分行为控制和主观状态，又不能把“坏学生”身份当成无限加码的依据。

这恰好帮助我们理解贝卡利亚的限度。他为了对抗旧刑法的道德化、身份化与法官任意，倾向于把评价压向客观社会损害。这一方向在历史上有解放作用，但如果推到极端，就可能低估主观责任结构。于是不能把他的观点当作今天刑法的完整答案，而应把它看成一次矫正：当一个制度太热衷于审判灵魂时，先把它拉回行为与公共后果。

SDE在这里看到的是一次“中心移动”。旧制度很容易把自我界里的身份、品格、忠诚、信仰放到刑法中心；贝卡利亚把重心移向现实界中可观察的行为后果与理念界中的一般规则。这个移动让刑法更可公开、更可预测，却也会产生新的缺口：人的行为并不是纯粹外部机械运动，意图、认知和处境仍会参与发生。

因此，本书不能简单赞美“客观主义”，也不能简单用现代主观责任原则反驳贝卡利亚。更有价值的做法，是把他的真正敌人抓准：他反对的不是一切主观要素，而是把不可测量的内心恶性直接升级为刑罚尺度。今天刑法继续使用故意、过

失，也必须把这些概念写成可证成、可争辩的法律标准，而不能偷偷恢复“我觉得这个人坏”的人格直觉。

如果把这一章压成一句普通人能使用的话：判断一个行为，可以问他做了什么、知道什么、能避免什么；但一旦开始用“他就是这种人”代替证据，刑罚就从行为滑向身份。贝卡利亚让国家从审判灵魂后退一步，这一步还没有解决全部主观责任问题，却为现代法治留出了一条重要底线。

> 带走的话
>
> 贝卡利亚不是要把人的意图从刑法里全部赶走，而是拒绝让国家把“内心有多坏”当成一把无限伸缩的刑罚尺。

> **小账**
>
> 他看见了什么？——按“坏心眼”量刑，会把阶层偏见带进刑罚。
>
> 接缝在哪里？——重心压向客观损害，可能低估故意、过失等主观责任。

第 12 章

偷一块面包和杀一个人为什么不能是一回事

比例原则最容易被说成一句废话：重罪重罚，轻罪轻罚。谁会反对？可历史真正的问题不是人们从来不知道“轻重有别”，而是他们经常在恐惧、愤怒、身份政治和象征秩序面前忘掉这句话。贝卡利亚举过一个刺眼例子：如果杀人和较轻违法都面临同样极刑，那么一个已经犯下较轻罪行的人，反而缺少避免更严重犯罪的额外理由。

这说明比例并不仅仅是“公平感”，还与制度激励有关。刑罚本身会进入人的选择结构。假如盗窃一只动物和杀人都判死刑，违法者一旦跨过第一条线，后面的线就失去意义。法律原本想把所有人吓住，却可能因为处罚失去梯度，反而把不同严重程度的行为推平。

这里有一个非常重要的思想转变：刑罚不是贴在犯罪后面的价格标签，而是整个行为环境的一部分。法律怎样分级，会反过来改变人怎样区分风险、怎样理解严重性。贝卡利亚甚至担心，法律如果对性质不同的行为施加同样处罚，会损害社会长期形成的道德区别。也就是说，刑法不仅反映社会价值，还会回写社会价值。

用一个日常比喻。公司规定：迟到五分钟和泄露商业秘密都直接开除。管理者可能觉得“严才能管住人”，但员工真正学到的不是纪律，而是规则没有层次。一旦有人已经迟到，他对后续更严重违规的额外顾忌反而可能下降。更糟的是，当所有错误都被同一种最高处罚覆盖，“严重”这个词本身会失去区分能力。

贝卡利亚因此需要一条犯罪梯度。从最危及社会存在的行为，到对个体造成较小不公的行为，应当形成由重到轻的序列；刑罚也相应形成梯度。这个设想并不是说每一个行为都能得到唯一精确分值，而是至少不能把顺序颠倒。最严重的犯罪不应得到比轻罪更轻的反应，轻微违法也不应承受为最严重侵害准备的刑罚。

注意，这里已经出现“路”的思想。重罪与轻罪不是两个孤立盒子，而是处在一个差异序列上。序列最重要的不是每个点之间恰好相差多少，而是方向不能乱。如果一部刑法把轻罪和重罪混成一锅粥，它失去的不只是量刑技术，还失去对社会成员传递意义的能力。

但为什么“偷一块面包”通常比“故意杀人”轻？如果回答“因为大家都知道”，那只是把习惯当理由。贝卡利亚的答案仍然回到社会损害：对生命、共同安全和社会结构造成的破坏不同。我们今天还会加入行为人的责任、被害后果、风险程度等更多维度，但最基本的公共说明不能省。比例不是情绪强弱，而是理由的等级。

这一点对现代刑法依然有用。每当一个社会遭遇高度关注案件，最自然的反应是“加重处罚”。但比例原则会逼我们问：这类行为在整个刑法体系中处于什么位置？如果为了回应舆论，把某一罪名的刑罚突然抬高，它与相邻犯罪的顺序会不会被打乱？一个局部的正义感，可能让整个体系出现新的不协调。

贝卡利亚的厉害之处，是他早早看见刑法是一张相互关联的表，而不是一串互不相干的罪名。调整一个刑罚，会改变其他刑罚的相对位置；过度提高某一轻罪处罚，不仅伤害该案被告，也会使严重犯罪和轻微犯罪之间的价值差异变模糊。刑罚体系因此像一架琴，某根弦拧得过紧，整架琴都可能走调。

但比例也会遇到一个难题：不同类型损害真的能放在同一条线上吗？身体伤害、财产损失、公共信用、国家安全、隐私侵害，它们不是同一种单位。贝卡利亚用“社会损害”提供总方向，却没有一把能把所有价值换算成单一数字的万能尺。正因如此，下一章必须专门说明：比例不是数学题。

这并不削弱比例原则，反而让它更现实。我们日常也不断做不能精确量化的比例判断。医生比较治疗收益与副作用，学校比较纪律与教育机会，法院比较侵害严重性与制裁强度。没有共同单位，不等于完全不能比较；只是比较必须公开理由，并承认边界。

从SDE角度看，贝卡利亚在这里把刑罚从一个静态“东西”变成一条差异序列D。每一个犯罪的严重程度都要靠与其他行为的关系显出来，刑罚也通过相邻差异获得意义。没有差异，就没有等级；没有等级，刑法只能剩下“罚”与“不罚”两档。一个只有总开关、没有刻度的制度，很容易走向残酷。

所以这一章的真正结论不是“轻罪轻罚”四个字，而是：法律必须保留差异。一个制度越是害怕复杂，越容易用同一重罚压平差异；而一旦差异被压平，人的行为选择、社会价值判断以及法治的可解释性都会同时受损。贝卡利亚要求刑法有刻度，不是为了把人变成数字，而是为了阻止权力只剩下一个音量——最大声。

> 带走的话
>
> 比例的第一意义不是“算得准”，而是“不能把差异压平”。一部好刑法首先要守住轻重的顺序，才谈得上精细量刑。

小账

他看见了什么？——同罚不同罪，会让已犯轻罪的人失去不犯重罪的理由。

接缝在哪里？——生命、财产、公共信用单位不同，能放在同一条线上吗？

第 13 章

比例不是数学题

讲到罪刑比例，最容易犯的错误是把十八世纪理性主义想象成一台巨大的计算器：仿佛贝卡利亚只要找到“社会损害”这个变量，就想把每个犯罪和刑罚精确换算成数字。前期研究阶段我们一度也接近这种理解，所以提出过“量刑尺”的候选收回动作。但原典自己否定了这个过强判断。（原典锚点：剑桥版第6章“罪刑比例”。）

贝卡利亚明确写到，人类激情彼此冲突，人口增长、私人利益交错，政治生活不可能按照几何学的精确度被导向公共福利。在政治算术里，概率计算必须替代数学精确。他又说，如果人类行动无限而隐晦的组合真的能够数学化，我们才可能拥有一条完全对应的罪刑刻度；现实中，明智立法者只能标出主要分界，避免把顺序颠倒。

这段话非常关键，因为它直接告诉我们：贝卡利亚想要的是“有秩序的近似”，不是“绝对精确的算法”。他相信犯罪之间有轻重，刑罚必须形成相应差异；但他并不声称可以用一个公式把所有案件算到小数点后两位。把他写成量刑数学家，不仅不公平，也会让我们错过他真正的难题。

真正的难题是：没有精确数学，我们仍然如何反对任意？如果任何比例都只是大致的，那法官是不是又可以说“我觉得这样差不多”？贝卡利亚的回答，是把主要分界和处罚区间尽可能写进一般法律。也就是说，他放弃精确数值，却不放弃预先规则；承认复杂性，却仍然要求公共排序。

这就像城市道路限速。我们无法证明某一条路每一秒的“唯一正确速度”恰好是60公里而不是58或62，但这不意味着限速完全任意。道路宽度、视距、行人密度、事故风险可以提供理由，规则需要给多数人一个可预期边界。精确到最后一公里并非重点，重点是不能今天因为司机穿什么衣服就说40，明天因为法官心情好就说100。

贝卡利亚的比例观因此有两层。第一层是结构比例：不同严重程度必须有不同反应，顺序不能倒。第二层是制度确定性：主要界线要由法律事先规定，不能完全留给个案情绪。两层合在一起，他追求的是一种“可预期的不精确”。这听起来矛盾，其实正是现代制度经常采用的方式。

很多法律概念都不是数学量，却要求稳定适用。什么叫“重大损失”“明显不当”“合理注意”“严重情节”？它们有弹性，但不是没有边界。真正的法治不是把所有不确定性消灭，而是把不确定性放进公开、可争辩、可复查的程序。贝卡利亚在十八世纪还没有完整发展出这套现代技术，他更倾向把弹性压缩在立法端，但问题已经被他逼出来。

这也帮助我们重新理解理性。理性不是“凡事都能算”，而是“知道哪些能算、哪些只能排序、哪些必须承认不知道”。一个理论如果面对复杂对象还假装绝对精确，反而是不理性。贝卡利亚承认概率与近似，说明他并没有把人的社会生活误认成简单机械装置。

不过，承认不精确之后，另一个风险出现了：谁来决定“主要分界”？如果答案仍然只是“明智立法者”，我们又回到第九章。贝卡利亚极其相信一般法律，却较少展开立法形成过程中的利益冲突、代表性和信息问题。这是他的历史边界之一。对他来说，最大的敌人是司法任意；对我们今天来说，立法任意同样需要制度化控制。

因此，SDE不能在这里说“贝卡利亚缺D，因为他想固定一切”。恰恰相反，他已经看见差异序列：犯罪从最严重到最轻微呈连续变化，社会行动复杂且不能几何精确。更准确的诊断是：他看见了D的连续性，却希望在制度操作层面把主要节点尽可能定型，以免司法过程重新把不确定性放大。

这是一种很典型的“看见之后收窄”。他知道世界复杂，但担心权力借复杂之名任意；于是宁愿牺牲一些个案适配，也要先换取规则确定性。今天我们容易批评这种形式主义太硬，可如果忘掉当时司法权的不透明和等级任意，就无法理解他为什么宁愿把钟摆推向另一边。

本章因此正式撤销“贝卡利亚以为刑罚可以精确数学化”的指控。以后总解构如果仍需要批评，只能批评他怎样安排不确定性，而不能捏造他否认不确定性。这个改判本身也符合本系列的纪律：当哲学家的原文比我们的母题更复杂，先改母题，不许删掉原文。

普通人可以从这里学到一种很实用的判断方法：没有精确数字，不等于没有道理；有数字，也不等于就有道理。真正重要的是比较依据是否公开、顺序是否一致、边界是否可复查。比例不是一道只需要计算器的题，而是一项必须给出理由的公共判断。

> **带走的话**
>
> 贝卡利亚并不相信罪刑关系可以达到几何学精确。他追求的不是“完美计算”，而是“主要顺序不能颠倒、规则不能任意漂移”。

> **小账**
>
> 他看见了什么？——政治算术只能用概率，立法只能标出主要分界。
>
> 接缝在哪里？——承认不精确以后，谁来决定主要分界？

第 14 章

一把不能太精确的尺

如果比例不能数学精确，我们为什么还把它比作“尺”？因为尺有两种理解。第一种是工厂卡尺：一毫米都不能差。第二种是地图比例尺：它不会把每一块石头画进去，却能告诉你大体距离和方向。贝卡利亚真正需要的，更像第二种。

他的刑罚体系至少要完成三件事：先把最严重和最轻微区分开；再让相邻行为保持合理顺序；最后让公民在行动之前大致知道越过哪条线会承担什么后果。这三件事都不要求绝对精密，却要求规则足够稳定。于是“一把不能太精确的尺”反而比“量刑计算器”更接近他的思想。

为什么不能太精确？第一，社会损害有异质性。生命、财产、自由、公共信用、国家安全并非同一单位。第二，同类损害也受情境影响。偷取同样金额，对不同被害人的实际后果可能不同。第三，行为人状态、未遂既遂、共同犯罪、风险程度都会加入。任何想把所有变量压成一个单数的模型，都可能把最难解释的部分偷偷藏起来。

但为什么又必须有尺？因为没有尺度，权力就容易退回身份、情绪和临时政治需要。今天同情就轻，明天愤怒就重；对普通人严格，对有地位者宽松；媒体关注时重判，无人关注时轻放。比例原则的价值，恰恰是让判断者必须把一个案件放回整个制度序列里，说明它为什么位于这个位置。

这可以用医院分诊来理解。急诊不会给每一个病人计算一个宇宙唯一的“严重度真值”，但必须区分立即抢救、紧急、普通和可等待。分级不完美，却比“谁喊得最响先救谁”更公平。刑法的比例结构也类似：它不是发现宇宙中本来存在的一串刑罚数字，而是为了让公共强制有秩序而建立的一套可争辩分类。

这样一来，贝卡利亚的比例思想就从发现学转向发生学。比例并不是等在那里让立法者找到，而是在社会对不同损害、风险和自由边界的长期判断中形成。形成

之后，它又回写人们的道德直觉：什么被视为严重，什么被视为轻微，法律的等级安排会参与塑造公众感受。贝卡利亚已经部分看见这种回写，所以他担心同罚不同罪会破坏社会的道德区分。

这里也出现一个值得后续专门处理的难题：如果法律会塑造道德，我们还能单纯用“社会损害”来证明法律吗？因为社会对损害的理解本身已经受到法律影响。比如某种行为长期被严厉处罚，人们可能因此觉得它天然严重；反过来，长期不处罚又可能使人低估其危害。法律与社会判断互相塑造，形成一个回路。

贝卡利亚的写法里，社会利益往往像立法者需要识别的对象，尚未充分展开“法律本身如何参与生产社会利益的定义”。这可能成为后面SDE总解构的一条支线：他已经看到刑罚会影响行为，却没有总是把这种回写继续用到“犯罪严重性如何被制度塑造”上。

但现在还不能下重判。因为《论犯罪与刑罚》中已经存在多处发生性洞察：秘密控告会改变人与人关系，过度法律会制造犯罪，刑罚严厉度和确定性会改变行为选择，教育会改变犯罪的发生条件。这说明贝卡利亚并非只把法律当被动镜子。他已经看到镜子有时会反射回去，只是没有把“回写”统一成体系。

所以，本章的任务是把“量刑尺”从错误指控改造成更准确的装置：不是一把绝对刻度尺，而是一张有主干刻度的公共地图。地图要稳定，否则没人知道路；地图也要更新，否则新道路、新风险和新生活方式永远进不来。贝卡利亚极力保护前者，现代法治必须继续补足后者。

从普通人的角度，比例原则最后可以变成一句朴素规则：别因为生气就把所有错误都判成“最严重”。家庭教育、学校管理、单位纪律都一样。如果每次犯错都上升到“你这个人完了”，规则就失去刻度。真正有效的制度，会让人知道哪一种行为错在哪里、严重到什么程度、还有没有修复和返回的路。

> **带走的话**
>
> 比例更像地图比例尺而不是卡尺：要有主干刻度，也要能够更新。

小账

他看见了什么？——法律的轻重安排会回写公众的道德直觉。

接缝在哪里？——社会损害本身受法律塑造，还能单独证明法律吗？

第 三 编

谁写规则，谁判人

编 序

第三编五章。第15章讲只有事先公布的法律才能制造刑罚；第16章讲法官不能临时发明一个罪；第17章讲解释法律为何让贝卡利亚害怕；第18章讲裁判不只是照着记分牌念；第19章把理念自由第一次放进九格表。

本编是第五条判决的主场：处罚者必须解释。贝卡利亚想把法官压成三段论机器，本书承认解释无法消除，于是要求解释留下理由、接受反驳和复查。第19章又为第四条判决里的自由闸搭好了骨架。

读这一编时，可以带着这个问题：公司手册里“造成不良影响”这类宽条款，领导“灵活解释”时，你有没有申辩和复核的入口？

第 15 章

只有法律可以制造刑罚

到这里，我们终于走到现代刑法最熟悉的一道门：罪刑法定。今天说“法无明文规定不为罪、法无明文规定不处罚”，很多人会觉得这只是法学院第一课。但在贝卡利亚那里，它不是一句技术规则，而是前面所有论证汇到一起后的政治结论。既然处罚权来自社会成员为了共同安全而作出的有限让渡，那么谁也不能在这份公共授权之外临时增加刑罚。

原典的逻辑非常清楚：只有法律能够规定犯罪的刑罚；任何作为社会成员之一的司法官，都不能正当地对另一个成员施加法律没有预先规定的处罚。法官如果把法定刑再加重一点，看上去只是“多罚一点”，实际上等于他自己增加了社会契约中从未授权给他的那部分强制。

这句话的革命性在于，它把刑罚从人的意志转成规则的后果。不是“法官觉得你该坐五年牢”，而是“社会通过事先公布的一般法律规定，这类行为在满足这些条件时承担这种刑罚”。法官当然仍然拥有巨大权力，但权力的正当形式改变了：他不是刑罚的创作者，而应当成为法律适用者。

为什么贝卡利亚如此在意“事先”？因为刑法最危险的一种形态，就是行为发生以后再根据结果、舆论或统治者需要决定它原来“其实是犯罪”。这种制度下，人无法在行动时安排自己的自由。今天不违法的事，明天可能因为得罪了谁被重新定义；同样行为，对不同身份的人可能出现不同规则。规则若不能事先知道，自由只能靠猜。

罪刑法定因此并不是单纯保护罪犯，而是保护所有还没有犯罪的人。它告诉每一个普通人：国家强制的边界必须先画出来。你当然可能不喜欢某些边界，但至少不应在走过之后才发现地面突然出现一条新线。法律的确定性，本质上是一种生活可规划性。

贝卡利亚还把这一点和主权结构连起来。谁有资格制定一般刑罚规则？在他的框架里，是代表社会整体意志的立法者，而不是处理具体案件的法官。这样，制定规则与适用规则被分开。这个分工后来成为现代刑事法治的重要基础：立法负责一般规范，司法负责具体判断，不能轻易相互吞并。

当然，我们今天知道现实没有这么整齐。立法不可能预见所有事实，语言也不可能完全没有模糊。所谓“法律制造犯罪”，并不是说法律凭空创造全部恶，而是说只有经过合法、明确、事先的规范过程，某种公共伤害才可以升级为国家刑罚意义上的犯罪。道德谴责、行政不便、舆论愤怒，都不能直接跳过这一步。

这条原则尤其适合用来抵抗“类推冲动”。当社会遇到一种令人讨厌的新行为，最容易出现一句话：“虽然法律没写，但这和那个犯罪差不多，也应该判。”从结果正义看，这句话常常很诱人；从罪刑法定看，却必须高度警惕。因为一旦允许法官按“差不多”扩张犯罪，法律边界会在案件发生之后不断移动。

贝卡利亚宁愿承受某些行为暂时不被刑罚覆盖的代价，也不愿让法官拥有无限补洞权。这是一种非常强的自由选择：宁可法律出现漏洞，也不要通过司法任意把漏洞随时补上。现代制度会用更复杂的立法更新、解释规则和违宪审查处理这些问题，但“刑罚不能靠事后创造”仍然是核心底线。

从SDE看，这里出现了一个漂亮的结构：犯罪不是自然物，但也不能由任何人随时发生。它必须通过一个被认可的制度路径发生。也就是说，贝卡利亚不是否认“犯罪概念是制度形成的”，相反，他非常在意形成路径D：谁提出、以什么形式、何时公布、怎样适用于所有人。只有走完这条路，国家强制才获得合法显露S。

问题是，他为了保护这条路径，把法律文本的稳定性放得很高。法律一旦形成，法官最好不要再参与创造。这就是下一章的主场。我们将看到贝卡利亚几乎把理想裁判写成一个“三段论机器”：大前提是法律，小前提是事实，结论只有自由或处罚。这个模型极其清楚，也极其值得今天继续追问。

先不要急着嘲笑“机械法官”。在一个法官可以根据身份、关系、气质和传统任意解释的时代，把法官降到规则之下，是一种解放。只有理解这一历史方向，我们才有资格进一步问：法官真的可能完全不解释吗？事实真的会自己走进法律格子吗？如果不可能，贝卡利亚到底压住了什么，又因此失去了什么？

所以第十五章的结论很简单，却非常硬：国家要处罚你，不能只因为有人认为你该受罚。必须有一条在你行为之前已经有效、由适格制度形成并向所有人公开的规则。贝卡利亚把刑罚的发生路径拉到阳光下，这一步直到今天仍然是法治最重要的防线之一。

> 带走的话
>
> **罪刑法定的真正意义，不是替“坏人”找技术漏洞，而是禁止国家在看见一个具体的人以后，才临时决定他原来犯了什么罪。**

> **小账**
>
> 他看见了什么？——刑罚必须走完一条事先公布的制度路径才能发生。
>
> 接缝在哪里？——为保护这条路，他把文本稳定放得很高，宁可留漏洞。

第 16 章

法官为什么不能临时发明一个罪

贝卡利亚对法官解释权的不信任，今天读起来甚至有些极端。他说，刑事案件中的法官没有解释刑法的权利，因为法官不是立法者。法官要做的，只是判断一个人是否实施了法律禁止的行为。他还用一个极有名的模型说明裁判：大前提是一般法律，小前提是行为是否符合法律，结论是自由或处罚。（原典锚点：剑桥版第4章“法律解释”。）

这就是“完美三段论”。像一道逻辑题：所有符合A条件的行为承担B刑罚；甲的行为符合A；所以甲承担B。整个过程最好不再加入第二套、第三套推理，因为一旦法官开始用所谓“法律精神”补充规则，不确定性就会进来。贝卡利亚甚至讽刺，法官的判断可能受情绪、关系乃至当天身体状态影响。

为什么他会如此警惕？因为他面对的不是今天理想化的职业司法，而是一种可能深受等级、秘密程序和个人裁量影响的审判。对他而言，“法官解释”首先不是精致法理学，而是权力绕过公开规则重新进入案件的通道。因此他宁愿把裁判想象得像机器，也不愿把人的任性重新放回去。

这个立场的第一层价值非常清楚：同样的法律，不应因为换一个法官就变成另一种刑罚。人的自由不能取决于审判者今天是否愤怒、是否认识被害人、是否喜欢被告的社会阶层。法官越能临时创造含义，刑法越接近“人治的文字版”。所以贝卡利亚把“解释”视为危险，不是因为他讨厌思考，而是因为他害怕思考变成无法控制的权力。

但第二层问题也同样清楚：没有解释，法律真的能工作吗？“禁止危险驾驶”里的危险怎样判断？“重大损失”多大算重大？“公共场所”包括哪些边界空间？即使法律写得再详细，语言也要在具体事实中获得意义。法官不是把一个已经贴好标签的事实机械塞进条文；事实本身如何描述，就已经包含选择。

举一个最普通的例子。规则写着“禁止车辆进入公园”。救护车算不算车辆？儿童玩具车呢？工作人员维修车呢？如果只说“按字面”，仍然要决定“车辆”在这个规则里是按物理特征、功能还是规范目的理解。解释并不是法官故意多事，而是一般语言进入具体生活时不可避免的中间过程。

所以从今天看，贝卡利亚的“三段论法官”并不是完整的司法理论。可如果我们因此说他“幼稚”，就错过了他真正想解决的问题。他在强调的并非解释从认识论上不存在，而是解释在刑罚领域必须受到极强约束。因为解释一旦扩大犯罪或加重刑罚，它造成的不是普通语言争议，而是国家对人的自由、财产甚至生命的强制。

换句话说，现代法治可以把贝卡利亚的命题改写成：法官无法不解释，但不能借解释成为第二立法者。解释必须有文本边界、体系理由、既有规则、公开论证和上诉复查，尤其不能用类推把本来不属于犯罪的行为硬塞进去。这不是背叛贝卡利亚，而是把他的自由担忧放进更成熟的制度工具。

SDE在这里看到一个很典型的“看见—收回”。贝卡利亚非常敏锐地看见裁判者的处境、情绪和关系会影响判断，这其实已经是在看E——法官并不是悬在真空中的纯理性机器，他有身体、有身份、有利益、有环境。可他的解决办法，却是尽量把这些土从裁判模型中消掉，让法官退回一个只做逻辑连接的角色。

这就是一个值得认真研究的接缝：他看见了“土会推”，却希望通过严格法律把土从司法判断中消音。为什么？因为在他的时代，那片土太容易变成特权和任意。于是“器具可消”的幻觉与法治追求纠缠在一起：如果规则足够明确，法官个人差异仿佛可以被消去，只剩下一部法律面对所有人。

这种愿望至今仍然诱人。我们总希望同案同判，希望不同法官得到相近结果，希望个人偏见尽量不影响裁判。这些愿望没有错。问题在于，消除任意不能靠假装解释不存在。真正成熟的制度，是承认解释不可避免，然后让解释留下理由、接受反驳、形成先例或受到上级法院审查。透明的解释，比不可见的解释更安全。

这也让AI时代出现一个特别有意思的回声。如果将来算法能够根据法条与事实自动给出判决，看上去贝卡利亚的“三段论法官”终于变成现实：大前提是法律数据库，小前提是事实输入，结论是刑罚输出。但算法训练数据、特征选择、分类方式同样包含解释。机器并没有消灭中间环节，只是把中间环节藏到模型里。

于是贝卡利亚两百多年前的问题突然重新变得尖锐：我们真正想消灭的是“人的解释”，还是“不可控的任意”？如果后者才是敌人，那么机器替代人并不自动解决问题。一个无法解释、无法质疑、无法纠错的算法，可能比有名字、有理由、能上诉的人类法官更接近他最害怕的那种权力。

这就是本卷以后进入AI编时要重新接上的线。贝卡利亚希望通过清楚法律让公民预先知道后果；AI也许能把预测推到极致。但可预测性如果以隐藏判断过程为代价，就会出现新的悖论：结果更稳定，理由却更不可见。我们不能只因为机器“算得一样”，就认为自由获得保障。

回到十八世纪，贝卡利亚的选择仍然值得尊重。他宁愿让法律承担主要错误风险，也不愿把每个被告人的命运交给法官即时创造。这不是司法理论的终点，却是一条现代刑法必须经过的窄门：先把处罚权从个人判断中抽出来，再讨论如何在规则内部承认解释。

前文写到这里，我们已经能更准确地描述贝卡利亚的制度性格。他不是简单崇拜严密逻辑，而是在一个任意权力浓度过高的时代，用一般法律、罪刑比例和法官受限去降低个体命运对权力者人格的依赖。他把自由放在“事先知道规则”的结构里。后面我们会看到，这套结构在酷刑、证据、秘密控告和死刑问题上继续发挥作用，也会暴露更深的边界。

◆ 理论深化　六个岔路口：如果继续往前追问

◆ 一、社会损害是不是后来“伤害原则”的同义词

读到这里，很容易把贝卡利亚的“社会损害”直接等同于后来更成熟的“伤害原则”，仿佛只要一个行为伤害别人，就可以犯罪化；不伤害别人，就绝对不能犯

罪化。这样写太快。贝卡利亚的工作首先是把刑罚从宗教罪孽、身份冒犯和统治者尊严拉回共同生活的公共后果，他没有在《论犯罪与刑罚》中建成一套后来意义上的自由主义犯罪化总理论。把后来的概念倒装回去，会让他显得比自己更系统。

更准确的读法是：社会损害在他那里承担“去神圣化”和“去身份化”的功能。它要求国家说出一种世俗的公共理由，而不是说“传统如此”“上帝不喜欢”“国王受辱”。这一变化本身已经足够巨大。至于哪些损害达到刑罚门槛、个人对自己的伤害是否可以犯罪化、公共道德能否单独构成刑罚理由，都还需要后来的理论继续争论。

这提醒我们做思想史时不要把一个人的“开门”写成后人已经装修完的房间。贝卡利亚打开的是公共理由之门，后人还要在里面继续区分伤害、冒犯、风险、道德和家长主义。SDE所谓发生，正是在这里拒绝倒装：后来的成熟概念不能被假装成早期思想一开始就完整拥有。

◆ 二、如果主观恶性不能直接量刑，刑法怎样保留责任

第二个岔路口是主观责任。贝卡利亚反对用“意图有多坏”作为犯罪的真正尺度，这很容易被误解成只看结果。实际上，一个可工作的刑法不可能完全不区分故意、轻率、疏忽和意外。问题不在于要不要主观要素，而在于主观要素怎样被法律化。只要国家说“我看你这个人就危险”，主观责任就重新变成人格标签；只有当它被拆成对风险的认识、对结果的希望或放任、可以避免却未避免等可举证命题，主观才从道德直觉进入法律程序。

因此，贝卡利亚对意图尺度的反对，可以被理解成一条现代提醒：刑法必须评价行为人的可责性，但不能假装能够读取整个人格。最危险的不是问“他是否故意”，而是从“他故意过一次”直接跳到“他本质上就是坏人”。前者仍然以行为为中心，后者会把一次案件变成对一个人全部未来的预判。

以后进入累犯、危险性预测和AI风险评分时，这条线会重新出现。技术越能给人打分，刑法越容易从“你做了什么”滑向“你是什么样的人、以后可能做什

么”。贝卡利亚为了反对旧时代的灵魂审判而退回行为，我们在算法时代可能再次需要这一步。

◆ 三、比例原则真正困难的不是轻重，而是“不可通约”

第三个岔路口是不同价值怎样比较。财产损失可以用金额表示一部分，生命损害却不能换算成金钱；公共信用、人格尊严、国家安全、环境风险更没有统一货币。比例原则因此不是把所有价值放进同一个计算器，而是承认它们常常不可通约，再建立可公开辩护的排序。

这正是“不能太精确的尺”最深的含义。一个制度如果为了显得科学，强行把所有损害压成单一分值，往往只是把价值判断藏进权重。谁决定生命值多少分、隐私值多少分、社会恐惧值多少分？数字不会消灭判断，只会让判断变得不容易看见。贝卡利亚没有面对算法评分，但他对“政治算术不能几何精确”的承认，给了我们很好的入口。

所以，比例原则最值得保存的不是数学外观，而是“不得无理由地颠倒顺序”。一个较轻侵害若被施加远重于严重侵害的处罚，制度必须说明为什么。说明不了，就暴露出比例失衡。法治的理性很多时候不是找出唯一答案，而是排除明显没有公共理由的答案。

◆ 四、法律确定性和个案正义为什么总在拉扯

第四个岔路口是本编最核心的结构矛盾。规则越细，越能减少法官随意；规则越细，也越可能遇到没有预见的新事实。规则越宽，越能适应复杂生活；规则越宽，又越给解释者留下空间。确定性和个案适配不是简单的敌我关系，而像跷跷板两端。

贝卡利亚把重量明显压向确定性，这是可以理解的历史选择。他所面对的首要危险是任意，所以宁愿让立法规则硬一点，也不愿让法官自由创造。但现代法治在继承这条底线后，需要再发展程序：解释要公开、裁判要说理、同案差异要说明、

上诉和再审能够纠错、立法能够及时回应新问题。这样才能把不可避免的解释从“个人权力”改造成“可复查的公共工作”。

SDE在这里不应该宣布“稳定就是死”。定型本来就是发生的一站。没有定型，任何规则都无法传递；真正的问题是定型以后有没有回写通道。法律最危险的不是稳定，而是把暂时稳定宣布为永远不许更新。贝卡利亚帮我们完成了第一次定型，后来的法治任务是给定型保留合法更新的路。

◆ 五、贝卡利亚的形式主义为什么既是药，也是副作用

第五个岔路口是形式主义。把法官约束在法律之下，是对专断的一剂药；但任何药进入不同身体，都可能产生副作用。形式主义的好处是让身份、关系和情绪不那么容易直接决定命运；副作用则是，当规则本身带有偏差或遗漏时，机械适用会把偏差稳定复制。

例如，一条表面中性的规则可能在现实中对不同群体产生极不相同的影响。法官若只能说“文字就是如此”，制度就可能把结构性不平等伪装成中立。贝卡利亚时代的第一任务是把人的任意从裁判中拿掉；今天的第二任务则是检查规则本身如何形成、谁参与、谁被遗漏、结果怎样回写社会。

因此不能在“规则统治”与“自由裁量”之间二选一。真正的问题是怎样把两者都装进可追责结构：规则限制裁量，裁量处理规则无法穷尽的事实；裁量必须说理，规则也必须接受宪法、权利与实践效果的持续检验。贝卡利亚给了我们一边，后人要补另一边。

◆ 六、这一部分对总母题“国家发生，主体先在”增加了什么

最后回到全书母题。第9—16章并没有直接证明“主体先在”，却给它增加了一个关键中间层：贝卡利亚不断把制度权力追溯到人的自由与共同利益，同时又努力把制度运行写成一般规则和可预期逻辑。这个结构说明，他最怕的是国家对个人的随机侵入，所以必须先给个人划出一块不应被任意穿透的自由空间。

这块自由空间究竟是一种规范假定，还是被他当成经验上已经完成的“人本来如此”，后面还需要继续分。若只是规范假定，它可以是一种限制权力的工作装置；若被扩展成关于人的完整事实判断，就可能成为新的先在。SDE不能因为看见“自由”二字就直接定罪，必须把规范层和发生层分开。

因此总母题现在更精确了：不是简单说“贝卡利亚认为个人先于社会”，而是问——为了给国家划边界，他是否必须假定一个在制度之前已经拥有可让渡自由、会权衡利害、能够理解规则的主体？这个主体是法律上的出发点，还是关于人的本体论断言？只有后面的威慑、教育、预防、激情与社会条件全部进入后，我们才能最终裁决。

普通人试验场　把贝卡利亚带回今天的三个场景

场景一：学校说“为了安全”，能不能什么都禁止

学校最容易出现贝卡利亚式难题。为了安全，禁止学生带某些物品、限制出入、安装监控、进行纪律处分，目标通常并不难理解。可“目标正当”不等于“手段自动正当”。如果一个学校以防止少数风险为由，把所有学生的日常活动都压到最低，必要性就可能从一道锁变成一枚通行章。

贝卡利亚式问法不是先说学校恶意，而是逐层拆：风险具体是什么？发生概率和后果如何？较轻手段是否足够？限制持续多久？有没有申诉和复查？同样的违规是否有轻重等级？如果迟到、打闹、考试作弊和暴力伤害最终都被“严重违纪”一把罩住，学校就犯了和旧刑法相似的错误——把差异压平。真正好的规则，不是没有处罚，而是让每一层强制都有公开理由和可预期边界。

场景二：公司制度写得很宽，领导能不能“灵活解释”

公司手册常有一些弹性很大的条款，比如“损害公司形象”“造成不良影响”“违反职业道德”。这些词并非完全不能使用，但如果它们成为领导临时处理

员工的万能口袋，贝卡利亚对法官解释的担忧就会在现代组织里重演：规则表面存在，真正决定后果的却是具体管理者当时怎样理解。

贝卡利亚会逼组织把关键处分条件尽量说清，让员工事先知道红线，并让同类行为得到相近处理。现代制度当然不能把管理者变成纯机器，但至少要要求重大处分有事实、有规范依据、有理由、有申辩和复核。所谓“灵活”如果只对管理者开放，对员工却没有反驳入口，那就不是适应复杂事实，而是单向裁量。

场景三：家庭教育为什么也需要“比例”

家庭不是法院，但权力结构同样可能失去刻度。孩子一次忘记作业、一次撒谎、一次危险行为，如果父母每次都上升到“你怎么这么不听话”“以后什么都不许做”，孩子得到的并不是清楚规则，而是一个不可预测的权力环境。今天挨批多少取决于父母情绪，而不是行为本身。

贝卡利亚式家庭规则并不是把亲子关系法律化，而是借一个简单原则降低任意：先区分轻重，再让后果和行为有关；处罚的目的应是帮助以后不再发生，而不是释放成年人的愤怒；规则最好事先知道，不要等事情发生以后临时加码；如果较轻办法已经有效，就没有必要继续加重。这样看，贝卡利亚写的是刑罚，却触到了所有使用强制的关系共同面对的问题——权力如何在有目标的同时，仍然给被约束者保留尊严和可预期性。

◆ 阶段小结　从“必要”走到“法官不能造罪”

前文八章完成以后，贝卡利亚的思想已经从刑罚权来源进入制度内部。前文问“国家凭什么有刀”，这一部分开始问“刀上的刻度谁来画、法官能不能临时改刻度”。问题看似越来越技术，实际上始终围绕同一件事：一个自由人面对国家强制时，怎样避免自己的命运取决于另一个人的即时意志。

我们也完成了一个重要改判。原先“量刑尺”作为收回动作已经正式退出。贝卡利亚并不主张把人的社会行为几何精确化，他明确承认政治判断只能使用概率和主要分界。以后对他的批评，不能再说“他天真地相信一切可计算”，而要更精确

地问：当他知道现实复杂时，为什么仍然把不确定性尽量压到立法端，把司法端想象成逻辑适用？

这八章还出现了一个新的SDE线索：贝卡利亚已经看见“土”。他知道法官的情绪、关系、处境会改变判断，也知道人的激情和利益让社会行为无法精确计算；可他解决这些问题的方式，经常不是让土进入制度模型，而是通过一般规则把土的影响压低。可以暂时把这一动作称为“把土静音”，但它还不够资格成为全书总收回，需要后续在酷刑、证据、死刑和教育章节继续核验。

后文将进入第17章以后：为什么“解释法律”会让贝卡利亚如此害怕，接着转入酷刑、证据、秘密控告和刑罚及时性。那里会出现本书另一条极强的发生线——司法工具并不只是发现事实，它会参与制造事实。酷刑尤其典型：表面上它在逼出真相，实际测量的却可能是谁更怕痛、谁更能忍、谁更愿意说出审讯者想听的话。

到那一步，我们对贝卡利亚的评价会进一步改变。他不仅是限制刑罚的人，也是早期看见“工具会改变所测对象”的思想家。国家为了找到真相使用的手段，可能反过来生产所谓真相；制度为了减少犯罪设计的规则，也可能反过来制造犯罪。只要这条线被原典继续支持，“贝卡利亚接近发生学”的判断就会比单纯“自由中心主义”更有厚度。

本编仍然保留总母题“国家发生，主体先在”，但暂不封顶。因为目前我们主要证明了国家权力、犯罪尺度和司法角色都被他放回制度关系中；至于“自由主体”是否真的被他当作发生之前的完成物，还需要后面处理人的激情、教育、预防和社会环境。只有那些材料仍然不能把主体发生补出来，总判词才有资格进入最后一编。

前文总判词：贝卡利亚把任意从法官手里往一般法律中收，但没有真正消灭不确定性。他改变的是不确定性出现的位置，以及谁必须为它负责。

带走的话

贝卡利亚想消灭的不是思考，而是“无法预期的个人意志”。他的三段论法官过于干净，却把一个直到今天仍然正确的问题钉在法治门口：谁有权在案件发生以后改变犯罪的边界？

小账

他看见了什么？——法官的情绪、关系、身体会影响判断，土会推。

接缝在哪里？——他想用三段论把土静音，可解释无法消除。

第 17 章

“解释法律”为什么会让贝卡利亚害怕

如果只读到贝卡利亚反对法官解释法律，很容易觉得这个年轻人把审判想得太简单：法律文字怎么可能不需要解释？同一句话放进不同案件，事实不一样、语境不一样、词义也会变化，法官不解释，难道让法条自己开口吗？可如果把问题放回十八世纪，他真正害怕的并不是“思考”本身，而是解释权重新变成处罚权的秘密入口。

上一章已经看到，贝卡利亚喜欢把裁判写成一个三段论：法律是大前提，事实是小前提，结论是自由或者处罚。这个模型当然过于干净，但它有一个非常清楚的政治目标——同样一条规则，不能因为换了一个法官，就长成另一种刑罚。一个普通人的命运，不能取决于审判者今天心情如何、出身如何、与当事人关系如何，也不能取决于法官是否愿意替模糊文字补一段自己的道德判断。

因此，他对“法律精神”的警惕也不是反智，而是反任意。只要允许裁判者说“文字虽然没有写，但法律真正的精神就是这样”，处罚边界就可能在案件发生以后移动。移动得越晚，被告越无法事先安排自己的行为。贝卡利亚要保护的是一种非常朴素的自由：我在做一件事的时候，应当大体知道国家会不会因此处罚我，而不是等我做完以后，才由另一个人告诉我这件事原来属于犯罪。

今天的法治已经不可能回到“零解释”。语言必然要进入生活，事实也必须被描述和分类。但是贝卡利亚逼我们保留一条底线：解释不能成为重新立法的化妆间。尤其在刑法里，解释如果把本来没有被法律覆盖的行为扩张成犯罪，或者把轻的责任解释成重的责任，就等于把立法权悄悄搬进审判席。

所以，现代人读贝卡利亚，不必接受他的全部答案，却应当保留他提出的问题：法官不可避免要解释，可谁来限制解释？限制不是让法官闭嘴，而是要求他把理由放到桌面上，让文字、体系、先例、程序和上诉共同约束这一步。真正危险的不是解释存在，而是解释不留下可争论的痕迹。

用SDE的大白话说，这里已经出现一条“路”。法律从纸面走进案件，不是瞬间跳跃，中间要经过事实认定、概念归类、语义选择和理由说明。贝卡利亚为了防止这条路被任意权力占领，曾经想把它压得极短。今天更成熟的做法不是假装路不存在，而是把路照亮。

把问题再压深一层，贝卡利亚其实在处理一种“时间上的权力”。立法发生在行为之前，司法发生在行为之后。如果司法可以在行为之后通过解释扩大犯罪边界，它就获得了一种追溯性的规范创造力：人行动时面对的是A规则，被审判时却被B含义包围。罪刑法定最重要的自由效果，就在于阻止这种时间倒流。

但现代法治也不能因为害怕追溯扩张，就把所有解释冻结成字典释义。文本有核心也有边缘，事实会出现新组合。真正需要区分的是“使既有规则可适用的解释”和“借解释制造新的禁止”。前者是司法不可避免的工作，后者越接近新增犯罪和加重刑罚，越应把决定权退回立法。

因此，判断解释是否越界，可以用三个普通问题检验：一个普通人在行为时能否合理预见这种解释？这种解释是否仍然围绕条文已经公开的核心危险？如果存在两种同样合理的解释，国家是否把不确定性的成本全部推给被告？这三问比“法官能不能解释”更接近贝卡利亚真正想守住的边界。

从SDE看，这里适合从S→D→E逆拆：先看一个扩张后的犯罪边界已经显露成什么，再追查它通过哪些语义、类比和事实分类路径形成，最后检查这些路径受哪些制度激励和时代恐惧推动。若三步一拆，所谓“法律本来就是这个意思”往往会显出自己的发生史。

贝卡利亚害怕的不是法官会思考，而是法官在案件发生以后重新决定“什么算犯罪”。刑法的解释必须存在，但解释越接近增加犯罪、加重刑罚，就越要承担更重的公开说明义务。

◆ 再看一层：为什么“可预见性”比“字面不变”更重要

如果把贝卡利亚的担忧压缩成“法官不许解释”，我们会错过其中最能活到今天的部分。普通人真正需要的，不是每一个词永远只有一种含义，而是在行动之前能够大体知道什么会把自己带进刑罚。可预见性因此比字面僵硬更重要：法律可以在边缘处解释，但解释不能突然把昨天合理的人变成今天的罪犯。

这一区分还能说明为什么刑法中的类推尤其危险。一般生活里，我们常说“差不多就按一样处理”；可刑罚不是普通管理。只要“差不多”能够把新的行为纳入犯罪，国家就在案件发生以后扩大自己的权限。贝卡利亚宁可让某些新型危害暂时暴露立法漏洞，也不愿让法官凭相似性把漏洞悄悄补成刑罚。这是一种有代价的选择，但代价恰恰显示他把自由放在什么位置。

当然，另一面也不能遮住：规则写得越死，越可能出现新型行为明明造成重大损害，却暂时无法纳入旧条文的情况。因此现代制度真正继承贝卡利亚的方式，不是重复“法官绝不解释”，而是建立一种分工——司法可以解释既有规则的适用边界，但新增犯罪和显著加重责任，应尽量由公开立法完成。这样，法律既能进入变化的生活，又不至于让处罚权在个案中偷偷长大。

所以，第17章最终要留下的不是对解释的恐惧，而是一条权力纪律：越是会扩大国家强制的解释，越需要更清楚的文本依据、更强的公开理由和更严格的复查。贝卡利亚把“解释”问题政治化了——它从语言技术变成了自由边界问题。

> **带走的话**
>
> 越是会扩大国家强制的解释，越需要清楚的文本依据、公开的理由和严格的复查。

小账

他看见了什么？——事后扩张解释，是一种让时间倒流的处罚权。

还要追问什么？——适用既有规则的解释与制造新禁止的解释，怎样分界？

第 18 章

裁判员真的只能照着记分牌念吗

想象一场足球比赛。记分牌写着二比一，裁判不能因为觉得落后的一队“踢得更漂亮”，就宣布比分其实是二比二。这就是贝卡利亚心里的理想法官：规则先在比赛之前公布，裁判负责判断事实是否符合规则，而不是根据好恶重新发明规则。

可真实审判比记分复杂得多。足球里“球进没进门”还有摄像机和门线，刑事案件里许多关键事实却不会自己贴着标签走进法庭。什么叫威胁？什么叫明知？什么叫公共场所？什么叫重大损失？一句看似普通的法律语言，到了生活里都会出现边界。裁判员不仅要“念分”，还要决定什么算一个有效得分。

这正是贝卡利亚模型最值得继续推进的地方。他已经看见法官的个人处境会污染裁判，却把解决办法主要理解成消除裁量。今天我们更容易看见，完全消除裁量既不可能，也未必是好事。真正可做的是把裁量变成受约束的判断：同样事实需要相似理由，不同处理必须说明差异；被告能够反驳；上级法院能够复查；社会能够知道一条模糊规则在实践中究竟被解释成什么。

这也解释了为什么“法官机械适法”和“法官自由裁量”不应当被写成两个极端选项。机械适法假装中间过程不存在，自由裁量又可能让中间过程失去边界。法治真正困难的地方，恰恰是在两者之间搭一条看得见的路。

贝卡利亚的贡献，是先把终点钉住：不能由个案法官临时造罪。我们的继续，是把从法律到终点的中间步骤打开：事实如何进入法律，证据如何被选择，含义如何形成，理由如何被检验。这些步骤越透明，法官个人越难把自己的偏好伪装成法律本身。

如果把这一点带回普通生活，也很容易理解。学校规定“严重扰乱课堂秩序者给予处分”，老师当然要解释什么叫严重；公司规定“重大失职可以解除合同”，

管理者也必须判断什么算重大。真正公平的制度不是把所有判断都删掉，而是让判断有标准、有记录、有复核，而不是“我说严重就是严重”。

裁判员比喻还可以再推进一步。体育裁判面对的规则相对封闭，刑事法官面对的却是自然语言、证据不完整与价值边界。所以司法不可能只“读比分”；它必须说明自己为何把某组事实归到某个法律概念下。真正接近法治的，不是没有中间步骤，而是中间步骤可被复查。

这也解释了理由公开的重要性。只给结果，不给推理，法官的解释就仍然可能是私人的；把文本依据、事实认定、反方意见和排除理由写出来，解释才从个人脑中的动作变成公共程序的一部分。上诉与审级监督的价值，也正在于让另一个判断者能够沿原路检查哪里出了问题。

贝卡利亚的时代更需要先压低法官个人裁量，现代制度则多了一层任务：承认无法消除的判断，并为判断设置可见边界。这不是对贝卡利亚的否定，而是把他的反任意目标从“减少解释”推进成“约束解释”。

如果换成对手理论，结果会不同。一个把个案实质正义置于最前面的理论，可能愿意让法官用目的解释填补更多空白；贝卡利亚式立场则宁可承受部分规则不完备，也不愿用事后扩张换取眼前的处罚。两者的分歧不是谁更聪明，而是谁承担法律不完备的成本。

◆ SDE小结：不是取消中间，而是让中间可见

贝卡利亚用规则限制判断者，今天我们要再补一步：判断不可消失，因此必须被程序化。把不可避免的中间步骤藏起来，只会让权力更难被看见。

◆ 再看一层：裁判理由本身也是一种限权装置

如果法官不可能完全机械，那么限制任意就不能只靠减少裁量，还要靠“把裁量留下痕迹”。一份判决为什么认定这个事实、为什么采用这个解释、为什么不接

受被告的另一种说法，若都能被读者沿着文字复查，判断就从个人直觉转成了公共理由。理由写得越完整，后来的人越有机会指出某一步其实跳得太远。

这也意味着，同案同判不能被理解成“所有案件输出同一个数字”。真正的同案同判，是重要差异相同时给出相近处理，重要差异不同时说明为什么不同。换句话说，法治需要的不是把法官变成复印机，而是让差异的选择标准稳定、公开、可争论。贝卡利亚想消除的是身份和情绪造成的随机差异；现代司法还要进一步说明，哪些差异在法律上值得被看见。

从这个角度看，上诉制度的意义也变得更清楚。它不是简单地“再找一个更聪明的人看一遍”，而是承认第一次判断可能受事实遗漏、概念选择和路径依赖影响，因此给制度安装第二条观察路径。一个只允许作出判断、却不允许重走判断之路的体系，很容易把偶然错误固定成权威。

如果把这一章换成一句普通人的话：裁判员当然不能改比分，但他必须解释为什么那个球算进、为什么那次动作算犯规。贝卡利亚最重要的遗产不是让裁判沉默，而是要求裁判不能把自己的声音冒充成规则本身。

再往前一步，理由公开还会反过来塑造法官自己。一个判断如果必须写给当事人、上级法院和公众看，裁判者就不能只停在“我觉得如此”。他必须把直觉拆成可以被别人检验的步骤。程序因此不仅约束结果，也训练判断。贝卡利亚没有用今天的司法论证语言说出这一点，但他的反任意目标正可以在这里继续生长：不是让人的判断消失，而是让判断在公开理由中接受第二次审判。

这也是法治成熟的一种标志：权力不仅要作出决定，还必须说明自己为什么这样决定，并接受别人沿着同一条理由之路重新检查。

> **带走的话**
>
> 裁判员不能改比分，但必须说明那个球为什么算进；判断不会消失，只能被程序化。

小账

这一章看见了什么？——裁判不只念分，还要决定什么算有效得分。

还要追问什么？——理由公开与上诉，能否让判断接受第二次审判？

第 19 章

理念自由：自由第一次进入九格表

到这里，我们第一次可以把贝卡利亚的“自由”放进九格，而不是只当一句政治口号。所谓理念自由，首先不是“我想做什么就做什么”，而是一个人能够在事前知道规则、理解边界，并且不被国家用事后制造的标准突然处罚。

这是一种非常制度化的自由。它依赖公开的法律、稳定的分类、可预期的后果，也依赖处罚权不能随人的身份和情绪变化。换句话说，自由不是规则的反面。对贝卡利亚来说，恰恰是一般规则使自由成为可能：因为规则把权力提前锁住，人才能在剩余空间里安排自己的行动。

九格中的“理念界×自由”因此非常强。罪刑法定、限制解释、罪刑比例、一般法律，都在这一格集中显影。它们共同回答一个问题：国家要怎样写规则，才能让公民不是生活在随时可能改变的意志之下？

可一旦这一格太强，也会出现副作用。为了让规则稳定，我们可能把现实复杂性压得过平；为了消除法官任意，我们可能假装语言没有边界；为了让任何人事前可预期，我们可能低估新型伤害、新技术和新生活不断改变法律含义的速度。理念自由需要稳定，但稳定不能变成永久冻结。

所以，自由第一次进入九格以后，立即带出一个更深的关系：理念界的自由如果没有现实界和自我界，就会变成“纸面上每个人都一样”。同一条法律对拥有律师、知识、时间和资源的人，与对贫穷、恐惧、语言能力不足的人，实际可用程度可能完全不同。贝卡利亚已经要求法律清楚、一般、平等，却还没有完整展开这种“形式可知”与“现实可用”之间的差异。

这里不能急着拿今天的平等理论去判十八世纪文本。更公平的结论是：贝卡利亚已经把理念自由的骨架搭得极强，而这副骨架随后逼我们继续问，规则怎样进入身体、生活和人格。下一编的酷刑问题，正好把自由从纸面推到身体。

◆ 阶段判词

“理念自由”还可以理解为一种免于私人意志的地位。它不保证每条法律都温和，却要求强制通过一般、公开、事先的规则说话。在这个意义上，规则不是自由的敌人，而是把“谁碰上谁倒霉”改成“所有人都能预先知道边界”的技术。

但理念自由不能独自完成正义。一个人形式上知道规则，不等于现实中有能力理解、申辩和使用规则；同样的程序成本、律师费用、语言障碍，会让纸面平等在不同人的生活里显露成不同自由。正因为如此，下一编转向身体不是题外话，而是检验理念自由能否落地。

这也给九格方法加上一条自限：九格不是把贝卡利亚硬塞进九只抽屉，而是检查一格过强时会遮住什么。理念界×自由很强，便要主动去现实界看身体与刑讯，去自我界看恐惧与承受力。若其他两界一进入，原判断必须调整，九格才不是分类游戏。

到这里，第三编可以形成一个完整推进：第15章把刑罚创制交给法律，第16—18章把不可消除的司法中间步骤公开化，第19章再把这种结构识别为一种理念自由。它的价值不在于机械，而在于把任意压到必须说明理由的位置。

贝卡利亚的自由首先表现为“免于任意”：法律必须事先存在，边界必须公开，处罚不能由个人意志临时创造。理念自由是他的强项，也正因此暴露下一问——当制度真正碰到人的身体时，这种自由还能不能站住。

◆ 再看一层：从“知道规则”到“真正能够使用规则”

理念自由最容易让人满意，因为它看起来已经很公平：同一部法典公开给所有人，同一个犯罪构成适用于所有人，同样的程序写在纸上。可一旦把镜头从法典移向生活，就会发现“知道规则”与“能够使用规则”并不是一回事。有人能请律师、找到判例、保存证据，有人连自己正在失去什么权利都说不清。形式上的同一入口，可能通向非常不同的实际能力。

这并不意味着罪刑法定和法律明确性不重要，恰恰相反，它说明理念自由是一道必要但不充分的闸。没有它，权力可以随时改规则；只有它，自由又可能停留在纸面。因此九格分析的意义不是否定贝卡利亚，而是迫使我们把同一判断放到另外两界继续受审：现实界追问身体、资源和器具，自我界追问恐惧、身份、理解能力以及一个人怎样成为能够行动的主体。

这也是为什么第三编必须在这里停一下。此前我们主要讨论“国家如何把规则写得不任意”，下一编则要问“国家为了取得事实，能不能直接进入人的身体”。一旦酷刑登场，理念自由不再只是法条上的边界，而要接受最残酷的现实检验：一个制度可以嘴上承认人人受法律保护，同时用疼痛让某些人失去作为主体说话的能力。

于是，理念自由真正完成自己的方式，不是把其他两界排除出去，而是允许现实和自我不断回写它。规则如果发现自己在现实中只保护了能使用它的人，就必须重新检查程序；程序如果发现某种取证工具正在摧毁被取证者的主体地位，也必须重新检查所谓效率。自由不是一块写完就封存的条文，而是一套不断阻止强制滑回任意的制度工程。

这条线还提醒我们，普通人真正需要的“懂法律”，也不是背会多少条文，而是知道当国家要改变自己的自由状态时，有哪些问题可以追问、哪些理由必须由国家提供、哪些决定能够被复查。法律可知性若只剩下专家之间的技术语言，理念自由仍可能在形式上存在、在生活中失效。因此，清楚的规则、可理解的理由和能够实际进入的救济程序，应被看成同一自由结构的不同层面。

> **带走的话**
>
> 理念自由是免于任意的自由：规则事先存在、边界公开；它是必要的闸，却不充分。

小账

这一章看见了什么？——一般规则把权力提前锁住，自由才成为可能。

接缝在哪里？——知道规则不等于能使用规则，纸面平等未必落地。

第 四 编

酷刑：一个寻找真相的机器，为什么会制造假象

编　序

第四编五章。第20章讲未定罪的人仍是法律上的无罪者；第21章讲疼痛测出的是忍耐力而非真相；第22章讲强者不招、弱者先招；第23章借米兰1630年案讲程序怎样参与形成案件事实；第24章估量贝卡利亚离发生学有多近。

本编是第三条判决的主场：他看见工具改变证据。酷刑不是透明的取证工具，它会生产口供。本编同时撑起第五条判决：强制者必须举证，制度必须允许以“证据不足”收场，并给错误故事留出退出门。

读这一编时，可以想一想：老师逼学生“主动交代”，家长不认错就不停批评，这样得到的答案，是事实，还是压力的回声？

第 20 章

如果还没有判有罪，他是谁

刑事司法里最容易被遗忘的一个人，是“还没有判有罪的人”。我们习惯把案件名称、媒体报道和警方行动放在一起，很快就用“犯罪嫌疑人”替代一个人的全部身份。可贝卡利亚抓住了一条极简单又极难坚持的原则：在有罪被证明以前，法律不能把人当作已经有罪。（原典锚点：剑桥版第16章“酷刑”。）

这句话的重要性，在酷刑问题上最清楚。国家说：我们不知道你是不是罪犯，所以先让你受痛苦，看看你会不会承认。可如果不知道他是否有罪，凭什么先施加一种只有罪犯才应承担的痛苦？贝卡利亚的反问几乎把整个刑讯逻辑翻了过来：为了证明某人应受处罚，不能先用处罚去生产证明。

这也是无罪推定最朴素的道理。它不是说所有嫌疑人事实上都无辜，而是规定举证责任的方向：国家想把一个人从自由状态推进到受刑状态，必须自己完成证明。不能要求个人先证明“我不是坏人”，更不能用强制迫使他替国家补证据。

普通生活里也有同样结构。一个孩子丢了东西，家长觉得是哥哥拿的，于是先罚站，直到他说出“真相”；公司发现文件泄露，先停职最可疑的员工，再要求他证明不是自己；网络上有人被指控，所有人先把他当成事实上的加害者，然后说“如果冤枉，你自己出来解释”。这些做法的共同点，是把证明责任压给被怀疑的人。

贝卡利亚的革命，在这里不是温柔，而是顺序。先证明，再处罚；先把事实放在公共程序里检验，再决定国家能做什么。顺序一旦倒过来，程序本身就开始制造罪名。

无罪推定不是一句替坏人说话的口号，而是一道关于权力方向的规则：自由是起点，处罚必须由国家说明理由。怀疑可以启动调查，却不能自动兑换成有罪。

◆ 原典锚点：刑讯先是权限问题，再是证据问题

《论犯罪与刑罚》第16章处理酷刑时，贝卡利亚的第一步不是问“疼痛能不能提高破案率”，而是先把被告的法律位置摆正：判决以前，他还不能被称为有罪的人。于是一个极简单的二分出现了——如果犯罪已经确定，依法定刑即可，额外刑讯没有必要；如果犯罪尚未确定，就不能先把一个法律上仍属无罪的人当作罪犯去折磨。这个论证之所以重要，是因为它把结果账和权限账分开：即使某种手段偶尔有效，也不能跳过“你现在有没有权这样做”。（参见 Beccaria, ch.16; Bellamy/Davies, 1995, pp.39–44。）

旧制度中的刑讯并不只有“逼口供”一种用途。贝卡利亚逐项反对的，还包括在矛盾陈述之间“试出真话”、逼问同伙、为另一些尚未查明的犯罪寻找线索，甚至用受刑去洗刷所谓名誉污点。把这些用途放在一起看，会发现刑讯像一把万能钥匙：证据不够时拿它补证据，叙述矛盾时拿它选版本，不知道还有没有别的犯罪时拿它去“钓”。正因为它什么都能被用来证明，反而最需要先问其权限从哪里来。

这也是“无罪推定”在本书里的真正位置。它不只是法庭上的礼貌用语，而是一种时间秩序：国家可以调查、询问、保全证据，却不能把最后结论提前搬到过程开头。程序之所以重要，是因为处罚的强度必须跟随证明状态增长，而不能在事实尚未形成时先把终局强制用完。

有人会反驳：如果面对的是极端紧急情形，例如有人掌握即将造成重大伤害的信息，难道国家还不能施加强大压力吗？这类问题不能靠把“酷刑是否有效”偷换成“紧急状态是否允许特殊措施”来回答。真正需要的是另设严格的权限来源、必要性门槛、替代方案比较与事后审查，而不是把刑讯重新命名为“紧急调查”就获得免检。贝卡利亚的贡献正在这里：例外必须承担比常态更重的说明责任。

从SDE看，刑讯章让我们第一次看到一个“介生态被提前封顶”的危险。案件事实尚在显露，证据路径仍在形成，法官却用身体痛苦迫使某个版本提前定型。问题不仅是痛苦太多，而是定型的时机被权力强行提前。只要这一点成立，酷刑的认

识论错误与权力错误就被连在一起：国家既过早伤害了人，也过早宣布自己知道了真相。

还有一个容易被忽视的层次：刑讯把“调查措施”和“刑罚措施”混在了一起。调查本来服务于判断有无责任，刑罚则以责任已经成立为前提；一旦调查者可以用本身具有惩罚性质的剧烈痛苦去获取证据，制度就让尚未成立的结论提前取得了刑罚权限。贝卡利亚反复追问“若罪已确定何须刑讯，若罪未确定凭什么刑讯”，正是在拆这场时间倒置。

这并不要求审前阶段完全没有强制。为防逃跑、毁灭证据而采取必要限制，与为了让人开口而施加痛苦，不是一回事。关键仍是双重审查：强制服务于什么具体目的，是否有更轻替代手段；强制本身会不会污染稍后要评价的证据。越接近惩罚性的措施，越不能只靠“办案方便”证明必要。

从制度设计看，反酷刑不能只靠一句“禁止刑讯”。如果审讯过程没有可追溯记录、律师无法及时进入、身体伤痕无人独立检查、非法取得材料仍可进入判决，禁止就容易退化为纸面命令。贝卡利亚没有现代证据排除规则的完整工具箱，但他的权限提问为这些后来装置提供了方向：国家不能从自己的违法强制中取得证明优势。

最强的反驳仍然是“救急”：如果一个人可能知道即将爆炸的炸弹在哪里怎么办？这类极端思想实验之所以有诱惑力，是因为它把“不作为的损失”推到最大。但越是极端，越应区分救援性紧急措施与刑事证明制度。即便某种紧急权力在特定法制中被承认，它也不能因此反向成为日常取证许可，更不能把取得的信息自动转换成有罪证明。

所以酷刑章的核心不是“文明人比古人仁慈”。它是一堂权限秩序课：在事实未定时，国家必须忍受一部分不知道；在证据不足时，制度必须允许无罪或证据不足的结局。一个不能忍受不确定性的司法，最容易把强制当作填补知识缺口的捷径。

带回普通生活，最小原则同样清楚：调查一件事时，不能先把“你必须承认”写进程序。学校、单位、家庭都可以核查事实，但若把承认作为解除羞辱、恢复工作或停止惩罚的唯一条件，程序就不再是在找答案，而是在生产答案。

◆ 最强辩护：如果刑讯真的能救人呢?

最强的辩护者会说，本书把权限说得太整齐：现实中有些信息只存在于某个人脑中，时间又极端紧迫，如果施加压力能找到被绑架者、阻止爆炸或挽救生命，难道还要为了程序纯洁什么都不做？这个反驳值得正面保留，因为它揭示了“事后定罪”与“即时救援”并不是同一种国家任务。

回答也必须分层。第一，紧急救援能否拥有特殊权力，要由法律另行规定并接受极高门槛，而不能从普通刑事取证权自然扩张出来；第二，即使某种紧急措施被允许，它产生的信息进入后续定罪时仍要接受独立证据规则，不能让“救援需要”自动洗白“证明用途”；第三，制度必须防止把极端例外变成日常捷径。

因此，贝卡利亚的权限账不是一张僵硬的“什么都不能做”清单，而是要求每次扩权说明自己属于哪一种任务、依据哪条权限、什么时候结束、错误如何补救。真正危险的不是承认例外，而是例外无需另付理由成本。

还有一种更日常的制度诱惑：把“拒绝配合”本身当作有罪迹象。一个人在高压情境中保持沉默，可能来自恐惧、法律建议、记忆混乱，也可能确有隐瞒；若制度把沉默自动升级为不利证明，就会再次用程序反应代替实体事实。贝卡利亚的逻辑提醒我们，国家不能因为自己没有得到想要的答案，就把不给答案本身变成新的证据来源。

因此，审前强制的文明化不只在于减少肉体痛苦，还在于限制“合作换宽待、沉默加怀疑、承认才停止”的交易结构。只要被调查者必须通过自我归罪才能摆脱强制，程序就仍然保留酷刑逻辑的一部分，只是把身体疼痛换成了其他压力。

带走的话

先证明，再处罚：怀疑可以启动调查，却不能自动兑换成有罪。

小账

他看见了什么？——罪未定而施刑讯，是用处罚去生产证明。

还要追问什么？——救急例外如何另设门槛，不回流成日常取证许可？

第 21 章

疼痛真的能发现真相吗

酷刑最迷人的地方，恰恰是它看起来非常直接。一个人知道秘密，不肯说；让他疼，他就会说。相比漫长调查、复杂证据和反复核验，疼痛像一条最短的路。可贝卡利亚的厉害之处，就是把这条“最短路”拆成了另一幅画面：酷刑测出来的首先不是事实，而是一个人能承受多少痛苦。

同样的疼痛，对不同身体、年龄、疾病、恐惧和心理承受力的人不是同一回事。一个有罪但强壮的人可能坚持不招，一个无罪但脆弱的人可能为了让痛苦停止而承认任何事情。于是酷刑把“真假”问题悄悄换成了“耐痛能力”问题，却还把结果当成真相。

这就是工具参与结果。我们以为自己拿着一把透明尺子测对象，其实尺子本身在改变对象。痛苦会改变记忆、表达、注意力和意志，逼供者的提问方式会给出暗示，预先认定的故事会决定什么答案被视为“像真话”。越相信酷刑能发现真相，越容易忽视它正在生产真相的外形。

贝卡利亚在这里非常接近发生视角。事实本身当然不是因为法官想要什么就凭空出现：一个人做没做某件事，有真实世界里的发生。但“案件事实”不是现实事件自动复制进法庭。它要经过证据收集、证人叙述、记录、鉴定、讯问和规则筛选。程序越暴力，现实与案件之间的距离可能越大。

所以我们必须严格区分两句话。第一句“现实事实是程序制造的”是错的，容易滑向一切都是叙事。第二句“案件事实一定经过程序形成”则是关键。程序不是创造过去，而是决定哪些过去能够以什么形式进入判决。酷刑的危险，就在于它把形成过程伪装成发现过程。

◆ 疼痛不是测谎仪：贝卡利亚那道反讽的“数学题”

贝卡利亚在酷刑章里有一个极锋利的反讽：若要把痛苦当作真相仪器，仿佛应该先知道一个无辜者的肌肉力量、神经敏感度，再计算需要多少痛苦才能使他供认。这个荒诞的“数学题”揭开了刑讯逻辑的核心错误——疼痛强度与事实真假之间没有稳定、单调的函数关系。人招供，首先证明他此刻承受了某种压力，并不自动证明供词描述的现实真的发生过。

因此，本书必须把“现实事实”和“案件事实”严格分开。现实事实不是程序制造出来的；一把刀有没有刺入身体、钱有没有被拿走，不会因为法官换了笔记方式就改变。但是案件事实——法庭最终能够据以裁判的那个事实版本——一定经过证据取得、筛选、质证、推理与规则评价。程序不能创造过去，却会决定过去以什么证据形态进入判决。

这一区分还能再拆四层：第一层是现实发生；第二层是痕迹与记忆留下什么；第三层是侦查把哪些痕迹带进案卷；第四层是审判怎样评价其证明力与合法性。酷刑最危险的地方，是它同时污染第二、三层：它制造新的语言材料，又让这些材料带着“本人供认”的高权重进入第四层。工具不只是读取事实，而可能改造自己准备读取的对象。

现代证据法因此必须同时处理两本账：可靠性账问“这条材料可信吗”，权限账问“国家能否用这种方式取得并利用它”。一段通过非法强迫得到的话，即使碰巧为真，也不能因此倒推出取得方式正当；反过来，一条取得合法的证据也仍要接受真实性和证明力审查。把两本账分开，是避免“只要抓到坏人，过程都可以”的最低条件。

这并不意味着“一切都是叙事”。恰恰相反，只有坚持现实事实独立于我们的叙述，才有理由严厉要求程序别制造假材料；只有承认案件事实经过程序形成，才有理由检查谁控制了形成路径。发生学在这里不是取消真相，而是把真相进入制度的路径暴露出来。

贝卡利亚的反讽“数学题”还暴露了测量理论上的错误。一个合格测量工具，至少要让同一对象在可比条件下产生相对稳定的读数；而痛苦反应高度依赖个体身体、情绪、预期与审讯者行为。若输入条件本身不可控，输出“供认/不供认”就不能承担对事实真假的直接测量功能。刑讯不是一支误差较大的温度计，它根本没有稳定测量同一变量。

供词之所以危险，还因为它在人类判断中天然具有叙事优势。物证往往零碎、概率性强，供词却能给出一条完整故事；一旦故事出现，其他证据容易被围绕故事重新解释。于是刑讯制造的不只是一个句子，而可能制造后续证据排序的“中心”。这就是为什么独立印证比“供词细节很多”更重要：细节必须来自被讯问者独立掌握，而不能来自审讯过程泄露。

从证明结构看，可以把证据质量拆成四问：来源是否合法，信息是否可靠，与待证事实是否相关，不同证据是否真正独立。刑讯几乎可能同时破坏四项：来源越权，可靠性因疼痛动机下降，相关信息被审讯者的问题预设，所谓“多个细节”又可能同源于同一次讯问。只数证据数量，会把一个来源复制成很多份“支持”。

现代技术并没有让这个问题消失。算法可以把若干低质量线索压成一个高风险分，但分数外观的精确并不会自动修复输入数据的污染。如果历史数据包含强迫供述、选择性执法或偏见，模型只是把旧路径标准化。贝卡利亚用酷刑告诉我们的原则可以直接移过去：先检查数据怎样发生，再看计算结果多漂亮。

所以“案件事实经过程序形成”至少包含一个反身要求：程序必须留下检查自身的痕迹。讯问记录、证据保管链、质证、理由公开和救济，不是官僚负担，而是在告诉后来者：这个事实版本是沿哪条路径长出来的，你可以从哪里推翻它。没有可追溯路径，“事实”越整齐，反而越值得警惕。

本章因此给出一个比“不要刑讯”更宽的判断：凡是能够显著改变被调查者回答条件的工具，都不能被当成透明窗口。权力参与越深，证据评价就越要把工具本身纳入对象。

◆ 最强辩护：供述为什么仍然可以是证据?

反酷刑也不能滑到另一个极端：既然口供会被工具塑造，就认为一切供述都只是权力产物。人在自由、自愿、理解后果的条件下完全可能陈述自己亲历的事实，供述也可以与物证、证人证言相互印证。问题不在于“说话的人不可信”，而在于什么条件使这段话具有可检验性。

所以应把供述价值拆开看：自愿性决定它是否受到不当压力，来源独立性决定细节是否真由被讯问者提供，外部印证决定故事有没有离开语言本身进入现实痕迹。三项越完整，供述越像证据；越依赖审讯者喂给的信息和停止痛苦的交换，越像程序自己制造的回声。

这也解释了为什么现代程序宁愿牺牲一点侦查效率，也要保留律师、记录、质证与排除等保障。它们不是把真相挡在门外，而是在降低国家自己制造“真相外观”的能力。贝卡利亚没有这些成熟装置，却把问题的逻辑起点钉得很准。

把这层意思压成一句话：证据越能被取得工具本身改变，工具就越不能隐身。摄像、讯问技巧、鉴定模型、算法筛选都一样；不是说使用工具就不真实，而是结论必须连同工具条件一起接受审查。科学性不是“有设备”三个字，而是别人能够知道设备如何工作、误差在哪里、是否存在替代解释。

贝卡利亚在十八世纪没有现代认识论术语，却已经抓到一个原始版本：把痛苦当作真相发生器，是因为制度忘记了真相与反应之间还隔着人的身体和处境。今天的工具更复杂，这个中间层只会更需要被看见，而不会因为技术先进而消失。

◆ SDE小结：工具不是透明玻璃

贝卡利亚反对酷刑，不只是因为它残忍，还因为它会破坏认识。一个制度如果用错误工具寻找真相，越用力，可能离真相越远。

带走的话

疼痛不是测谎仪；凡能改变回答条件的工具，都不能被当成透明窗口。

小账

他看见了什么？——疼痛强度与事实真假之间没有稳定的函数关系。

还要追问什么？——供述在什么条件下仍能成为可检验的证据？

第 22 章

强者不招，弱者先招

如果酷刑真能稳定找到有罪者，那么支持它的人至少还有一种功利辩护：虽然残酷，但有效。贝卡利亚偏偏抓住了最致命的反例——痛苦并不会按照道德责任分配。身体不会因为一个人有罪就自动变脆，也不会因为无罪就自动变坚强。

于是刑讯出现一种荒诞的筛选机制：最能忍的人获得优势，最不能忍的人承担风险。一个无辜者如果比有罪者更怕痛，他就可能更快给出审讯者想听的答案。制度原本想筛出“谁做了”，最后筛出的却是“谁能撑住”。

这不仅是酷刑时代的问题。任何把某种可见表现直接等同于真实性的制度，都可能重演这个错误。面试中把自信当能力，课堂上把沉默当不懂，警务中把紧张当心虚，算法中把某些行为模式当高风险，都可能把“容易测的东西”偷换成“真正想知道的东西”。

贝卡利亚让我们看到，评价工具总会偏爱某些人、惩罚另一些人。一个制度若不检验这种偏差，就会把工具造成的差异误认为对象本来就有的差异。SDE所谓“土”和“路”，在这里不是抽象词：身体、审讯环境、权力关系就是土；疼痛—提问—回答—记录就是路；最后的“口供”只是这条差异序列中形成的一次显露。

这也解释了为什么现代证据制度如此强调相互印证、排除非法证据和程序权利。不是因为程序比真相更重要，而是因为没有可靠程序，我们甚至不知道自己得到的是事实，还是工具的回声。

制度最危险的错误之一，是把“容易测量”当成“真正重要”。酷刑把耐痛当真相；今天的新工具也可能把分数、表情和行为痕迹当成人本身。每一次测量，都要问工具在制造什么。

◆ 同一种刑讯，对不同身体不是同一种“测试”

刑讯还有一个常被“程序一视同仁”掩盖的问题：同样的疼痛手段落在不同身体上，实际强度并不相同。身体状况、年龄、疾病、恐惧阈值、对权力的依赖程度，都可能改变一个人何时崩溃。于是，刑讯不是让所有人面对同一把真理尺，而是在用个体承受能力决定谁先说出审讯者想听到的话。

这正是“强者不招，弱者先招”背后的结构。若有罪者更能承受，他可能坚持沉默；若无辜者更脆弱，他可能为了让痛苦停止而接受审讯者提供的故事。程序表面上对每个人“施加同样压力”，结果却把身体差异转成了证据差异。平等形式反而生产不平等结果。

更麻烦的是，错误供述并不总是简单的一句“我认罪”。在持续压力、诱导问题和信息污染中，被审讯者可能逐步吸收办案人员给出的细节，最后形成一套听起来内部一致的叙述。此时后来的办案人员再拿这套叙述去寻找印证，容易陷入自我强化：越多材料围绕既定叙事被收集，最初的强迫来源越不显眼。

因此，反对刑讯不等于反对一切讯问。合法讯问仍然可以核对矛盾、追问细节、展示证据；区别在于，讯问的目的应是让信息接受检验，而不是让身体痛苦成为结论生产线。只要“停止痛苦”本身成为最强激励，回答就同时服务于脱离痛苦这个新目标，真假不再是唯一变量。

把这一点带回普通生活也很清楚。老师如果反复逼学生承认一个尚未证明的错误，家长如果在孩子承认之前不停止羞辱，单位如果以停职相威胁要求“主动交代”，都可能得到高度一致的答案，却未必得到更可靠的事实。贝卡利亚的刑讯批判因此不是一件古董，它提醒任何调查者：权力越能改变回答条件，就越不能把回答当作透明窗口。

这里还有一个选择效应。刑讯会优先筛出“最不能承受的人”，而不是“最有罪的人”。如果制度随后把先供认者视为最可疑者，身体脆弱、心理易受暗示、语

言能力较弱的人就会在程序中系统性处于劣势。表面相同的手段，因为落在不同纠缠条件E中，会产生不同差异序列D，最后形成不同案件显露S。

这说明法律平等不能只理解为“对所有人施加同样压力”。真正的程序平等还要问：不同的人是否拥有大致相同的理解、表达和抵抗错误归责的机会。未成年人、认知障碍者、语言不通者或高度依赖权威者，需要额外保障，不是给他们特权，而是防止统一程序把差异变成错误证据。

错误供述还有长期回写。一次承认进入案卷以后，侦查人员、检察官、法官乃至公众都会在认知上更新对嫌疑人的看法；后来出现反证时，系统也可能倾向把反证解释成“狡辩”。所以纠错不能只在终点增加一道复核，而要尽早阻止污染源进入主叙事，否则每一轮回写都会提高撤回成本。

好的讯问制度恰恰要把“获得信息”和“让对方屈服”分开。开放式提问、避免泄露关键细节、对重要讯问全程记录、允许律师参与、对脆弱被讯问者增加支持，都是在降低工具对答案的塑形能力。它们共同守护的不是被告的舒服，而是事实生产线的可检验性。

因此，贝卡利亚的身体批判可以被重新表述为一种证据伦理：国家越强，越必须避免把自己的强度当作真相强度。强者沉默、弱者承认，不是个人性格趣闻，而是告诉制度：你选错了变量。

这一章的保底也必须写清。承认个体差异会影响回答，不等于任何供述都不可信；合法、自愿、能被独立证据印证的供述仍有证明价值。要拒绝的只是那种把“承受不住所以说了”直接等同于“因为是真的所以说了”的推理。

◆ 制度修复：把脆弱性从“个人缺陷”变成程序信息

如果制度知道不同人对压力的承受能力不同，就不能继续把这种差异当成被讯问者自己的运气。脆弱性应当成为程序需要识别的信息：年龄、语言能力、认知状态、疾病、创伤和依赖关系，都可能决定是否需要更早获得法律帮助、是否应缩短讯问时间、是否需要第三方在场。

这里的逻辑不是“脆弱者不负责任”，而是先保证事实形成过程尽量不把脆弱误写成有罪。责任判断发生在事实相对可靠之后；若取证阶段已经系统性偏向某些身体和人格，后面的责任讨论再精致，也可能只是给错误材料加上更漂亮的法律语言。

因此，“同样程序”并不是平等的终点。真正的平等，是不同的人都拥有足以抵抗错误归责的最低能力。贝卡利亚没有把这条现代程序保障清单写出来，但他的“强者不招、弱者先招”已经说明：司法不能把承受痛苦的差异当成真相差异。

这还改变了我们对“程序中立”的理解。程序不能通过假装所有人完全相同来保持中立，而要把与错误风险直接相关的差异纳入保障设计，同时避免把身份差异直接写成罪责差异。前者是纠错，后者是歧视；两者的分界在于：差异究竟用于提高事实可靠性，还是用于提前判断谁更像罪犯。

贝卡利亚的旧问题由此接到一个现代原则：越是国家自己造成的证据环境，越应由国家承担证明其可靠与公平的负担。不能先制造高压，再把高压下的反应解释成对象本性的证据。

> 带走的话
>
> 刑讯筛出的不是谁做了，而是谁撑不住；司法不能把承受痛苦的差异当成真相差异。

> **小账**
>
> 他看见了什么？——同样的刑讯落在不同身体上，强者不招，弱者先招。
>
> 还要追问什么？——脆弱性怎样成为程序信息，而不成为罪责标签？

第 23 章

司法机器从什么时候开始参与制造“事实”

说司法会“制造事实”，很容易被误解。好像只要程序一参与，现实就不存在了；好像同一件事可以随意讲成任何样子。这不是本书的意思。现实发生过什么，不会因为法官、律师或者SDE愿不愿意就改变。真正需要发生学解释的，是现实怎样变成法律能够使用的“案件事实”。

一场冲突发生在街头，现实里有无数细节：光线、距离、声音、先后动作、每个人的位置和注意力。进入案件以后，这些细节必须被选择。有人陈述，有录像，有鉴定，有笔录，有时间线。没有任何判决能把现实全部搬进法庭，只能把与法律问题相关的部分组织起来。

这一步组织不可避免，却必须受规则约束。证据资格决定什么能进来，证明标准决定需要多大把握，质证决定相反意见如何进入，裁判理由决定法官如何把碎片连接成结论。也就是说，案件事实不是“捏造”，而是一个受程序约束的重建结果。

酷刑特别危险，正因为它既参与事实形成，又假装自己只是被动发现。审讯者越相信答案早已藏在嫌疑人身体里，越可能把疼痛当钥匙，而忽略自己正在改变被讯问者的表达。贝卡利亚对酷刑的批判，因此天然地通向一条现代认识论：权力越参与知识生产，越需要公开它的工具和路径。

AI时代这个问题会更尖。摄像机、数据平台、行为画像和模型会产生海量证据，看上去比口供客观得多。但选择哪些数据、怎样训练模型、什么阈值算异常，仍然是路径。工具从皮鞭换成算法，不会自动变成透明。

所以，本书在这里立一个必须一直保留的区分：现实事实不是司法制造的；案件事实一定经过司法程序形成。前一句防止相对主义，后一句防止“工具中立”的幻觉。两句缺一不可。

司法的任务不是创造过去，而是以受约束的方法重建过去。程序不是挡在真相前面的麻烦，而是真相获得公共资格的道路。

◆ 米兰1630：程序怎样把恐惧写成“证据”

贝卡利亚身边的米兰知识环境提供了一个极强的历史物证。1630年瘟疫期间，卫生官员Guglielmo Piazza与理发师Gian Giacomo Mora被指控用所谓“涂油”传播瘟疫，遭受反复刑讯并被处死。Pietro Verri后来在《论酷刑的观察》中重读案卷，展示互相指控、翻供和刑讯怎样把一个荒诞故事越写越完整。Treccani关于Verri的传记还指出，他早在1760年代初已收集相关材料，并把材料提供给贝卡利亚使用；Verri在1770年代后期把论证整理为成熟著作。

这里必须守住思想史分寸：不能把Verri后来写成的每一句话倒装成贝卡利亚已经说过，也不能用曼佐尼十九世纪的《耻辱柱的历史》替贝卡利亚作证。这个案例的价值，是让我们看到同一座城市里真实存在的制度记忆：当司法把“必须找到涂油者”当作起点，刑讯就不再只是一个取证工具，而会反过来组织问题、回答和后续证据。

这个循环可以画成四步：先有灾难与恐惧；恐惧要求一个可归责对象；审讯通过刑讯逼出某个名字或叙述；新的名字又成为下一轮抓捕和刑讯的理由。到最后，每一步都似乎有上一份“口供”支持。系统不是凭空制造瘟疫，却可能制造“谁传播了瘟疫”的案件事实。

Verri的材料因此比一句“酷刑残忍”更有穿透力。残忍描述的是手段给身体造成什么；程序批判还要追问手段给知识造成什么。一个制度若让结论参与生产自己的证据，就会出现反馈闭环：越相信某个嫌疑，越容易收集只支持它的材料；材料越集中，又越觉得最初的怀疑得到证明。

这也是第21章区分现实事实与案件事实的历史落点。现实中的瘟疫存在；“涂油者”是否存在，是需要证据证明的命题。程序一旦把后者当作前提，就会把查证变成填空。贝卡利亚反对刑讯，最深处正是在保护“尚不知道”的状态：制度必须

允许一宗案件最后以“证据不足”结束，而不是把不确定性视为必须用强力消灭的漏洞。

从SDE看，1630年的教训不是“E决定一切”，而是S、D、E互相推动：社会恐惧与制度期待构成特征纠缠E；审讯、指认、再审讯形成差异序列D；“涂油者”这一案件显露S逐渐稳定；稳定后的S又回写E和D，让下一轮搜查更偏向同一叙事。真正要切断的，是这种没有反向检验的闭环。

Piazza—Mora案还说明，社会危机会改变司法对错误的容忍结构。瘟疫需要解释，城市需要恢复秩序，公众需要一个可以看见的原因；在这种高张力E中，“没有找到凶手”本身容易被视为制度失败。于是调查路径D被一种结果期待牵引：必须有人承担原因。刑讯在此并非孤立残暴，而是整个环境把不确定性压向单一答案的技术。

Verri为什么多年后还要重读这个案子？因为案卷保存了一条可以回看的发生路径。错误不是突然从某个“坏法官”脑中掉下来，而是由恐惧、医学无知、权力压力、审讯技术与相互告发逐层加固。把错误拆成路径，才能看见制度应在哪一步设置刹车，而不是把所有责任压给最后宣判的人。

曼佐尼后来再写“耻辱柱”，又把责任、历史记忆与法律判断推进了一步。但本书必须抵制一个方便的倒装：十九世纪的反思不能成为十八世纪贝卡利亚已经拥有的思想证据。它只能说明同一案件如何在后来的文化中持续回写，成为意大利法律与文学反思酷刑的共同材料。

这个案例还给“公开理由”增加了一层要求。仅仅公开最终判决不够；如果形成结论的路径不公开，社会只能看到一个干净答案，看不到答案曾经怎样被制造。现代审判要求理由、证据规则和救济路径，就是把部分形成过程变成公共可检验对象。

更重要的是，制度需要允许“退出错误故事”。一旦调查围绕某个叙述形成，必须有机制让反证真正降低对该叙述的信心，而不是永远被解释成更复杂的阴谋。

能够撤案、无罪、再审和赔偿，都是给系统准备退出门。没有退出门的调查，越努力越可能把自己锁进最初假设。

所以米兰1630不是本章的戏剧性插曲，而是一个结构压力测试：当E中的恐惧升高、D中的确认路径变窄，S就可能被过早锁定；若锁定后的S又不断回写E和D，错误便拥有自我延续能力。真正的法治，不只是避免第一次错，还要降低错误自我繁殖的能力。

◆ F6检验：错误判决怎样写进环境

第80章的F6要求本书证明：SDE不能只把既有证据法语言换个名字，它至少应提出一条会改变取证方向的可检验问题。1630年米兰“涂油者”案提供了一个机会。判决不只留在卷宗里；在Mora房址上竖起的“耻辱柱”把司法判断写进城市空间，让后来的人在经过现场时继续接收“这里曾有罪恶”的官方叙事。

由此可以提出一个不同于“什么时候出现新证据”的问题：错误判断被纠正的速度，是否还受它已经写入环境的物、纪念物、档案口径与制度记忆影响？耻辱柱直到1778年才被拆除；Verri对这一案件的系统反酷刑批判后来才公开出版，Manzoni又在十九世纪重写这段历史。这些时间点不能证明“拆柱导致纠错”，却证明错误判决具有一种独立于原始证据的环境持续性。

F6因此得到一个有限的正面结果：SDE的三方程至少逼研究者把“判决后的环境残留”单列为变量——S一旦稳定，可能通过E的物质和制度载体继续影响后续D。传统证据法主要审查证据怎样形成判决，这一问则把视线继续推到判决怎样改变后续证据环境。这个增量仍然需要更多案例检验，不能由一个耻辱柱直接升级成普遍因果律。

◆ F6扩样：卡拉斯案对照，以及三类近邻

单个耻辱柱只能提出问题，不能回答问题。要让“环境持续性”从一个好看的比喻变成可检验的变量，至少要做两件事：先说清它和已有研究的关系，再找一个条件不同的案子对照。

先看近邻。政治学的路径依赖研究早已说明，制度一旦起步，会因收益递增而越来越难以回头（Pierson， 2000）；记忆研究从哈布瓦赫起就指出，社会通过纪念物、仪式与叙事框架保存对过去的解释；错案研究中的“隧道视野”则解释，调查者一旦锁定嫌疑，会系统性地收集支持材料、忽视反证（Findley & Scott, 2006）。这三类研究分别覆盖了制度惯性、集体记忆与认知偏差。本书所谓“环境持续性”若要保有一点自己的东西，只能是一个更窄的问题：判决被物质化、官方化地写进公共环境以后，这些“铭刻物”本身会不会拉长纠错所需的时间。它不讨论调查者怎样犯错，那是隧道视野；也不讨论一般制度为什么难改，那是路径依赖；它只讨论判决之后留在环境里的那层“物”。

再看对照。1761年秋，图卢兹新教商人让·卡拉斯的长子死在家中，卡拉斯被控为阻止儿子改宗天主教而将其杀害，1762年3月被处以车轮刑。伏尔泰随即发起持续的公开申诉，1763年写成《论宽容》；1765年3月，王室复审机构宣告卡拉斯无罪，距处决约三年。这起案件没有留下任何把“罪行”铭刻在城市空间里的纪念物，却有一位当时欧洲最有影响的作家持续为之奔走。

把两案并排：米兰案的判决被铭刻为耻辱柱，此后一个多世纪里没有一位有足够影响、又能公开发声的申诉者，Verri的反驳写成后也长期没有出版，耻辱柱从1630年立起，到1778年才被拆除，前后约一百五十年；卡拉斯案没有铭刻物，有强大的外部申诉，约三年平反。结果与“铭刻物拉长纠错时间”的预期方向一致。

但必须同时承认：两案中，“有无铭刻物”和“有无强大申诉者”这两个条件恰好一起变化，两地的司法结构、宗教处境与时代也都不同。两个案例无法把铭刻物的作用与申诉力量的作用分开，更谈不上因果。要分开它们，至少还需要两类案例：一类有强大申诉却也有铭刻物，一类既无铭刻物也无申诉者。每个案例都应记录同样四项：判决以什么形式写进环境；铭刻物何时撤除；第一份独立反驳何时公开；正式纠错何时完成。

因此，F6的状态从“有限正面”改记为“有限正面，待分离”。“环境持续性”仍是本书保留的唯一SDE解释增量候选，但它现在有了一个清楚的失败方式：

如果在更多案例中，控制申诉力量以后，铭刻物的有无与纠错时间无关，这一增量就应撤回。

◆ 历史保底：1630案不能被倒装成贝卡利亚自己的完整论证

把Piazza—Mora案写进贝卡利亚卷，还要再加一道史料纪律。Verri对案卷的系统整理主要在后来完成，曼佐尼的文学—法律反思更晚。我们能够确定的是：这一案件及相关材料属于米兰启蒙圈可接触的历史记忆，Verri早期已经收集材料并与贝卡利亚有密切思想合作；不能确定的部分，不能替贝卡利亚补写成他亲自完成的案例研究。

这条限制本身很有方法意义。思想“发生在一个圈子里”不等于圈子里每个人拥有同一套论证。共同材料可以构成E，却不能把不同作者的S直接并账。贝卡利亚的《论犯罪与刑罚》、Verri的《论酷刑的观察》、曼佐尼的《耻辱柱的历史》应当被看成同一历史材料在不同时间、不同问题意识中形成的三次显露。

因此本章保留这个案例，不是为了给贝卡利亚增加一份他没有写过的证据，而是为了给读者看见：一套错误司法怎样被后人一次次重新读取。发生史真正有价值的地方，正是把“谁在什么时候看见了哪一层”分清，而不是为了论证漂亮把后来的眼睛借给前人。

从思想传播角度看，这个案例还揭示“材料—作者—制度”的三层不同。案卷可以长期存在，却无人重新解释；Verri把旧材料组织成反酷刑论证，贝卡利亚把酷刑放进刑罚权与程序正当性的普遍问题，后来的读者再把两者带入新的制度争论。材料不是自动会说话，思想的发生就在于有人重新决定什么差异值得保留、怎样把它们连成一条路。

这也是本书为什么要给历史材料标明时间。没有时间标记，后人的成熟批判会被偷偷塞回前人的脑中，思想史就变成一条倒着修建的高速公路。真正尊重贝卡利亚，不是让他提前拥有所有现代答案，而是准确说明他在哪一步把门打开，后人又在哪一步继续走。

带走的话

现实事实不是司法制造的，案件事实一定经过程序形成；两句缺一不可。

小账

这一章看见了什么？——米兰1630年案里，刑讯把恐惧一步步写成证据。

接缝在哪里？——铭刻物与申诉力量尚未分离，环境持续性仍待检验。

第 24 章

贝卡利亚在这里离发生学有多近

写到这里，最容易出现两种夸张。第一种说，贝卡利亚已经是一个完整的发生学思想家，因为他看见制度、工具和环境影响结果；第二种说，他只是十八世纪刑法改革者，SDE不过是在用自己的词给他重新包装。两种说法都太满。

更准确的判断是：贝卡利亚多次亲眼看见“发生”。他看见处罚效果取决于确定性和及时性，而不是强度一个变量；看见酷刑会生产错误供述；看见秘密制度会改变人的关系；看见法律过多会制造更多犯罪；看见教育会改变行为条件。这些都说明他没有把社会世界理解成一堆静止对象。

但这些看见还没有被统一成一套“所有显露都有差异序列和特征纠缠，并会回写”的一般理论。很多时候，他仍然需要一个较稳定的立法者位置、一个可以被公共幸福计算的总体视角，以及一个作为规范起点的自由成员。我们后面真正要做的，不是宣布这些位置错误，而是逐一问：它们什么时候只是必要的制度脚手架，什么时候会变成不再接受追问的地基?

第三编和第四编还带来一次重要改判。前期研究时那句“国家发生，主体先在”开始显得过粗。至少在经验层面，贝卡利亚已经看见人的选择会被痛苦、恐惧、教育和制度塑造；因此不能说他完全把人当成固定原子。更谨慎的说法应当是：他在规范上需要一个自由主体来限制国家，在经验上又承认实际行动者会发生变化，而两层之间怎样连接，还没有成为他理论的中心。

这就形成全书后面的一个重要任务：不把“主体发生”误写成“人没有责任”，也不把“规范自由”误写成“现实能力人人相同”。一个人可以在规范上平等地拥有权利，同时在经验上拥有非常不同的知识、资源、心理和选择条件。刑法既不能因为后者而取消前者，也不能因为前者而假装后者不存在。

◆ 四座桥：从程序走向刑罚机制

第一座桥通向严厉性：如果程序已经说明国家不能用任何方法发现真相，那么处罚阶段同样不能用任何强度制造秩序。第二座桥通向确定性：贝卡利亚真正相信的不是疼痛越大越有效，而是规则与后果之间的连接越稳定越有力量。第三座桥通向死刑：当一种处罚不可逆，错误就不再只是概率问题，而成为制度无法回写的问题。第四座桥通向主体：威慑、教育、身份与社会关系将进一步证明，经验中的人一直在制度里发生。

◆ 本编总判词

贝卡利亚反对酷刑最深的一层，不是“文明战胜野蛮”，而是他看见：国家用来寻找事实的工具，会参加事实如何进入法律的过程。现实不是程序制造的，但任何可处罚的“案件事实”，都必须走过一条能被质疑、能被复查的公共道路。

> 带走的话
>
> 贝卡利亚多次看见了发生，却还没有把这些看见统一成一套一般理论。

> **小账**
>
> 他看见了什么？——工具、秘密、过度禁止与教育都在改变结果。
>
> 接缝在哪里？——规范上的自由主体与经验中变化的人，怎样相接？

第 五 编

最狠的刑罚为什么未必最好

编 序

第五编五章。第25章攻击“强度就是效果”的等式；第26章讲确定性比严厉更重要；第27章讲及时性和每一段等待的理由；第28章把威慑看成一条路；第29章说明刑罚第一次变成行为机制研究。

本编是第三条判决中“规则改变行为”的主场：刑罚被拆成强度、确定性、及时性、比例等变量，效果取决于整条路。第29章又提出“自由是边界，效用是边界内的选择标准”，为第四条判决和死刑编做准备。

读这一编时，可以拿交通罚款对照：一条罚得很重却几乎没人查的规则，和一条罚得轻却每次都执行的规则，哪一条更管用？

第 25 章

人人都相信“罚得重才害怕”

一谈犯罪，人们最自然的反应往往是把处罚往上加。盗窃多了，就加重；诈骗多了，再加重；出了恶性案件，舆论里最容易出现的句式也是“判得太轻，所以才有人敢犯”。这套直觉有一个非常简单的模型：人怕痛苦，痛苦越大，越不敢犯罪。只要接受这个模型，刑罚政策似乎就像拧音量旋钮——犯罪声音太大，就把刑罚音量继续拧高。

贝卡利亚真正厉害的地方，是他没有停在“刑罚要轻一点”这种道德劝告，而是直接攻击这个模型的因果结构。他指出，刑罚的目的不是让已经发生的犯罪消失，因为过去不可能被取消；也不是让国家获得报复的满足。刑罚若有公共正当性，只能面向未来：阻止犯罪人再次伤害社会，并使其他人不愿做出同类行为。既然目标在未来，评价刑罚就不能只问“有多疼”，而要问“它到底怎样影响未来的行为”。

这一步看起来只是把报应换成预防，实际改变了整个问题。只要刑罚目的被放到未来，强度就不再天然等于效果。医生给药不是剂量越大越好，灭火也不是水越多越好；药物过量会伤害身体，水用错地方可能扩大损失。刑罚同样如此。一个制度如果只会不断加重，却不知道加重通过什么路径改变行为，就只是把国家暴力本身当成答案。

贝卡利亚在“刑罚的温和”一章里提出一个非常重要的门槛：为了达到预防目的，刑罚只需要使犯罪带来的预期不利超过其吸引力；超出这一需要的部分，就不再增加正当性。这里的“只需要”极有力量，因为它第一次把刑罚从“能多重就多重”改成“足够即可”。国家不是在举行一场痛苦竞赛，而是在寻找实现公共目标所需的最低强制。

但如果只把这句话理解成一张简单的收益—成本表，也会误读贝卡利亚。前面几章已经看到，他并不相信人的激情可以被几何精确计算。他关心的是结构，而不

是小数点。一个人面对犯罪机会时，真正进入判断的至少有几样东西：收益是否立即可得，被发现的可能性有多大，程序是否会真的走到底，刑罚是否会执行，执行要等多久，以及他对这些信息是否相信。法条上写着“最高十年”只是其中一格。

这里可以用最普通的交通规则来理解。假如一条路明文规定超速一次罚五千元，但一年中几乎没人测速，驾驶者很快会学到：这条规则虽然很重，却很远。另一条路罚款只有二百元，但摄像头稳定工作、规则清楚、申诉渠道存在、处罚很快到达。哪一种更能改变日常驾驶？答案并不取决于罚款数字谁大，而取决于整条规则链条在行为者心里有没有形成稳定联系。

贝卡利亚由此把“严厉”拆成了多个变量。严厉只是强度；确定性回答“会不会真的发生”；及时性回答“什么时候发生”；持续性回答“影响是一瞬间还是反复进入记忆”；比例性回答“不同犯罪之间能不能保持梯度”。一旦拆开，我们会发现，国家把全部希望压在强度上，常常意味着另外几项做得不好。侦查不稳定、程序缓慢、执行随机，于是只好把纸面刑罚越写越重，希望靠恐怖数字弥补制度路径的松散。

这也是为什么极重刑罚有时反而制造逃避。贝卡利亚注意到，刑罚越可怕，行为人越可能为了逃避它冒更大风险；司法者、被害人或陪审者也可能因为刑罚过重而不愿推动定罪，最终形成“法律很凶，执行很少”的奇怪结构。强度上升如果同时把确定性压低，整体威慑并不会按法条数字同比增加。

更深的一层是社会会适应残酷。贝卡利亚用一种近乎心理学的直觉指出，人会对反复出现的严厉景象形成习惯，极端痛苦不可能无限增加震慑。一个社会把残酷变成日常，可能得到的不是更温顺的公民，而是更习惯残酷的公民。刑罚因此不只是作用于犯罪人，它也在训练旁观者对暴力的感觉。国家如何处罚人，会回写社会怎样理解暴力。

这一点值得放进SDE的“回写”。立法者原本以为自己只是用强力压住犯罪，可刑罚方式本身会成为社会经验的一部分。公开酷刑、羞辱性处罚、极端处置如果

长期存在，人们可能逐渐把暴力视为解决冲突的正常语言。贝卡利亚已经隐约看见：国家的手段不是透明的工具，它会改变那片需要治理的土。

因此，本章不是替“轻刑”做简单宣传。刑罚太轻而完全失去约束力，同样不能完成贝卡利亚的目标。真正要保住的是一句更精确的话：刑罚强度必须服从于其功能，而不能反过来让“越重越好”成为无需证明的信仰。重到什么程度、为什么重、增加的这一层痛苦到底增加了什么效果，都应当被说明。

> 带走的话
>
> 贝卡利亚真正推翻的不是重刑本身，而是“强度就是效果”这个最容易被相信的等式。刑罚从此不能只展示牙齿，还必须证明牙齿为什么有用。

> **小账**
>
> 他看见了什么？——刑罚面向未来，强度并不天然等于效果。
>
> 还要追问什么？——增加的那一层痛苦，究竟增加了什么效果？

第 26 章

重不重要，为什么“跑不掉”更重要

如果第25章只是告诉我们“罚得重不等于有效”，问题还没有完成。贝卡利亚接着给出一个更具体的替代变量：确定性。与其让法律偶尔落下一记雷霆，不如让较温和的后果稳定地跟在违法之后。他的著名判断是，适度而确定的刑罚，往往比可怕却伴随逃脱希望的刑罚更能影响人的行为。（原典锚点：剑桥版第27章“刑罚的温和/宽缓”。）

◆ 现代实证检验：贝卡利亚的“确定性”经得起多大程度的外部验证

Daniel S. Nagin 2013年对二十一世纪威慑研究的综述，为贝卡利亚“确定性比严厉性更重要”的直觉提供了难得的外部检验。综述认为，关于处罚确定性的威慑证据比关于处罚严厉性的证据更一致；但它同时加了一个重要限定：现有证据主要指向“被抓获的概率”，并不能简单扩展成“只要法律后果更确定，威慑就一定更强”。

这个限定很重要。它阻止我们把贝卡利亚升级成已经被现代统计“完全证实”的先知。现代研究支持的，是对“不断加重刑期就能显著提高威慑”的怀疑，以及对执法可预期性的重视；至于不同制裁组合、被抓概率、起诉概率、定罪概率与社会信任怎样共同作用，仍存在大量研究空白。

因此，本章的判断从口号变成可证伪命题：若未来高质量研究反复表明，在控制抓获概率后，单纯提高刑罚严厉度具有稳定、强而普遍的额外威慑作用，本书就必须收回“严厉性通常不是主杠杆”的概括。现在能说的更谨慎：贝卡利亚把注意力从痛苦强度移向可预期的行为路径，这一转向与现代实证总体方向相容。

“确定性”很容易被误解成国家必须百分之百抓到每一个人。贝卡利亚的重点其实更接近一种可预期联系：一个人不能长期形成这样的经验——多数时候规则只是写在纸上，只有极少数倒霉者被突然严惩。如果真实经验是“很可能没事”，那

么再重的刑罚也会被希望稀释。人在计算时，不会只看最坏后果，还会看最坏后果离自己有多远。

一个校园例子最直观。学校规定抄袭一经发现立即取消课程成绩，但老师几乎从不检查作业，学生很快就会知道这是一条低确定性规则。另一门课处罚较轻，只要求重做并扣分，可教师每次都认真核查、说明理由、允许申辩，学生反而更容易形成稳定预期：抄袭会带来真实成本。制度是否有力量，不只看它能够做多重的事，还看它是否持续做自己公开说过的事。

因此，确定性首先是一种“承诺能力”。法律告诉公民：你做A，会发生B。若A之后经常不发生B，法律其实在进行反向教育——它教人学会区分“写着的规则”和“真的规则”。时间一长，真正支配行为的不是法典，而是大家从经验中总结出来的执行概率。这个经验法比任何宣传都强。

这里能看见一个重要回写：执法选择会改写法律意义。纸面规则保持不变，但如果长期选择性执行，人们对它的实际理解会变化。某些规则会被当作真正底线，另一些规则会被当作“只有碰上检查才算”的临时障碍。法律不只通过文字塑造行为，也通过执行历史塑造行为。贝卡利亚虽然没有现代制度经济学术语，却抓住了这一点。

不过，确定性也有一条危险边界。为了让处罚“跑不掉”，能不能降低证明标准？能不能扩大监控？能不能牺牲程序权利，让抓捕和定罪更快？当然不能。前文刚刚花了整整一编说明，尚未定罪者不能被提前写成罪犯，证据工具也不能为了效率扭曲真相。所以贝卡利亚式确定性必须与正当程序同时成立：确定的是经过合法证明后的后果，而不是怀疑一出现就自动处罚。

这可以写成一个非常重要的双门结构。第一道门是证明：国家必须证明这个人实施了法律规定的犯罪，程序允许怀疑失败。第二道门是执行：一旦责任经过正当程序成立，法律后果就不应再由关系、金钱、身份和临时恩宠随意改变。前一扇门要求谨慎，后一扇门要求稳定。没有第一扇门，确定性会变成高效冤错；没有第二扇门，正当裁判会变成可交易的纸。

贝卡利亚谈到赦免时，正好暴露了他对确定性的极端重视。他认为，一个完善而温和的刑法体系不应经常依赖个案恩赦来表现仁慈；真正的仁慈应该进入立法本身，让法律整体温和，而不是先制定过重刑罚，再靠君主偶尔饶恕。这个判断今天仍然很有穿透力：制度若必须靠“好心人破例”才能保持人性，往往说明规则本身有问题。（原典锚点：剑桥版第46章“赦免”。）

这也解释了为什么贝卡利亚对“善良的统治者”并不满足。一个人今天碰到仁慈法官、明天碰到严厉法官，这不是自由；一个人今天被赦免、另一个同样处境的人没有被赦免，也很难形成平等预期。贝卡利亚宁可把仁慈写进普遍规则，也不愿让人命依赖偶然的恩典。这里仍然是他一贯的理念自由：把个体命运从人格化权力中抽离。

SDE在这里会追问一个新问题：确定性究竟是D的稳定，还是S的僵化？答案取决于它稳定的是什么。如果稳定的是“依法证明—公开裁判—一致执行”这条路径，确定性让规则可预期；如果稳定的是一个错误判断且缺乏纠错机制，确定性就会把错误放大。所以真正成熟的确定性必须包含复核与改判。可预期不等于永不修改，而是修改也要有公开路径。

普通人在生活里其实很懂这个道理。家庭规则最伤人的往往不是太严，而是今天这样、明天那样；领导心情好就算了，心情差就重罚；同样迟到，有人被处分，有人因为关系被放过。此时大家学到的不是守规则，而是猜人。贝卡利亚最想消灭的正是这种“猜人社会”。

> **带走的话**
>
> 刑罚真正的力量不在于它能把人吓到多远，而在于行为、判断和后果之间能否形成稳定连接。确定性不是让国家更快处罚所有嫌疑人，而是让经过正当证明的法律后果不再靠运气。

小账

他看见了什么？——选择性执法会教人区分写着的规则和真的规则。

接缝在哪里？——追求“跑不掉”，不能以降低证明标准为代价。

第 27 章

三年以后再罚，还有多少意义

刑罚还有一个很容易被忽视的变量：时间。即使一个刑罚强度适当、最终也会执行，如果犯罪发生后三年、五年、十年才到来，它对人的行为意味着什么？贝卡利亚明确提出，刑罚越迅速、越接近犯罪，通常越公正也越有用（原典锚点：剑桥版第19章“刑罚的及时性”）。这里的“迅速”不是草率审判，而是在满足必要程序的前提下，尽量减少没有理由的等待。

为什么先说“更公正”？因为等待本身可能已经成为一种惩罚。一个尚未被最终认定有罪的人如果长期羁押，长期处于不知道命运的焦虑中，程序就把本应在判决之后出现的痛苦提前扩散到了判决之前。贝卡利亚强调，审前羁押只能服务于必要的安全和证明需要，而且应尽可能短。这与前文关于“时间顺序”的讨论完全接上：不能让处罚偷偷跑到证明前面。

为什么又说“更有用”？贝卡利亚借助观念联结的思路认为，行为与后果之间间隔越短，人在心里越容易把两者连在一起。小时候摸到烫杯子，疼痛立刻到来，很快就学会避开；如果今天做一件事，五年以后才出现一个完全不同的后果，心理联系会弱得多。刑罚当然比条件反射复杂，但时间距离确实会影响规则是否被理解成行为的结果。

这给我们一个很重要的D序列。犯罪不是一个点，刑罚也不是另一个点，中间至少有发现、调查、起诉、审判、执行多个环节。任何一环长期停滞，都可能把“行为—后果”的联系拉断。贝卡利亚的及时性因此不是单纯追求速度，而是在保护整条因果路径不要散掉。

但这里必须加上一道现代反问：快是不是越快越好？当然不是。程序越快，如果压缩的是辩护、核验、证据审查和上诉时间，就可能以“效率”之名破坏准确性。一个十五分钟就判完的案件也许极其迅速，却不能因此更正义。贝卡利亚自己把迅速与必要程序联系起来，他要减少的是无意义拖延，而不是把证明过程取消。

于是及时性出现两个方向相反的风险。太慢，审前焦虑和羁押可能变成无判决惩罚，行为与后果联系变弱；太快，事实尚未充分显露就被封顶，错误可能无法纠正。真正成熟的制度不是简单选择快或慢，而是问：每一段时间在做什么？这一天是在等待必要鉴定，还是卷宗躺在桌上没人处理？这一个月是在保障辩护，还是程序惯性？只有把时间拆成路径，速度才有法律意义。

这一点对普通机构同样适用。公司发现员工严重违反制度，如果半年不处理，团队会形成“其实没关系”的经验；半年后突然追责，又会让人觉得规则只是权力随时拿出的旧账。学校处分学生亦然。及时回应不是为了让管理者马上发火，而是为了让规则与事件保持可理解联系。

时间还影响人的身份。一个人犯下一次轻罪，经过漫长程序后可能已经工作、成家、重新建立生活；刑罚很晚才到来，法律面对的已经不是完全相同的社会位置。这并不自动取消责任，却提醒我们：程序拖延本身改变了E——人的关系、生活、经济和心理条件。刑法不能假装时间只是日历上的空白。

从SDE看，贝卡利亚在这里比单纯“威慑论”更有意思。他不是只问一个结果S够不够重，而是在看D的长度怎样改变S的意义。相同刑罚，紧接行为发生与多年后发生，虽然法条文字一样，却不是同一种社会事件。时间参与了刑罚的效果，也参与了刑罚的正义。

这也为后面死刑问题埋下一根钉子。死刑是一种强度极高、不可逆的终局结果。为了避免错误，程序往往需要极高谨慎；可程序越长，贝卡利亚强调的及时关联越弱。极端刑罚于是可能天然拉出一个张力：越想保证准确，越难保持迅速；越想迅速，越无法承受不可逆错误。这不是他原文完整展开的论证，却是沿着他的两条原则自然出现的后续问题。

带走的话

时间不是刑罚之外的包装。犯罪与后果之间隔多久，会改变程序的正义，也会改变刑罚的意义。真正的及时不是快判，而是让每一段等待都有必须存在的理由。

小账

他看见了什么？——等待本身会变成未判决的惩罚，拉断行为与后果。

接缝在哪里？——死刑越求准确越难迅速，越求迅速越难承受错误。

第 28 章

威慑不是一个点，而是一条路

现在可以把第25—27章放在一起了。过去最简单的威慑图像只有两个点：犯罪——刑罚。刑罚越重，犯罪越少。贝卡利亚实际上把这条直线拆开了。一个潜在行为人面对的不是抽象“刑罚”，而是一整条可能的路：做了以后会不会被发现，被发现后会不会形成可靠证据，有没有机会靠身份或金钱逃脱，判决要多久，最终处罚是否会稳定执行，处罚本身又以什么方式进入他的经验。

这条路里任何一处断掉，最终的重刑都可能失去效果。抓不到，重刑只是墙上的海报；抓到了却因为程序任意导致大量撤销，规则仍然不稳定；判了却经常靠关系改变，法律再次变成人情；拖延太久，行为与后果失去联结。威慑因此不是“刑罚强度”一个变量，而是多环节共同形成的可预期性。

用SDE的语言说，这一章主角是D——差异序列。贝卡利亚在这里不像一个只会设定终点的立法者，更像在观察一条路径怎样影响人的选择。路径每走一步，行为人、执法者、旁观者都会更新对规则的认识。某项法律第一次出台时，人们只知道纸面文字；几年以后，人们真正相信什么，更多来自对执行历史的观察。D不断回写E，E又影响后文D。

这能解释一个常见现象：为什么有些社会规则不重，却人人认真遵守；另一些规则写得很重，却长期被视为“看运气”。差异往往不在法条修辞，而在路径是否一致。一个稳定但温和的制度，会慢慢积累可信度；一个偶尔雷霆万钧、平时几乎不工作的平台，会积累另一种知识——只要避开抽查即可。

但贝卡利亚的行为模型也必须接受攻击。他常用“人会根据痛苦、利益、概率调整行为”的方式解释威慑，这确实有很强的现实直觉，却可能把人的行动压得过于单一。很多犯罪并不是坐下来拿纸计算收益后发生的。冲动、成瘾、群体压力、身份认同、报复、绝望、情绪失控，都可能让远期刑罚信息暂时失去作用。

因此，贝卡利亚的模型在什么条件下最强？很可能是在那些存在一定计划空间、行为人能够感知规则、违法收益和风险可以比较的场景中。比如经济型犯罪、走私、部分财产犯罪、重复性违规等，确定性与收益结构可能更直接影响选择。到了极端激情犯罪、严重精神障碍相关行为或高度组织化暴力，单纯威慑模型的解释力就会下降。

这就是本卷必须写的“保底”。不能因为威慑不能解释所有犯罪，就宣布贝卡利亚错了；也不能因为威慑在很多领域有效，就把它当成人类行为总理论。一个理论真正成熟，不是声称无处不适用，而是知道自己的工作区间。贝卡利亚最强的是把刑罚从情绪报复带向可检验的制度效果，而不是已经解释了每一种犯罪心理。

更深的问题是：刑罚是否只能通过恐惧起作用？一个人不盗窃，可能因为害怕处罚，也可能因为认同他人的财产权，珍惜自我身份，不愿伤害关系，或者根本生活在不需要靠盗窃生存的环境中。贝卡利亚后面谈教育和法律清晰时，其实已经走出“纯恐惧模型”。规则可以塑造习惯、信任和公共道德，而不仅是提高成本。

这意味着威慑本身也有层次。最底层是害怕痛苦：“抓到会罚。”再上一层是可预期秩序：“规则是真的，每个人都这样。”再上一层是规范认同：“我知道为什么不该这样做。”如果制度永远只靠第一层维持，它需要不断提供外部压力；如果第二、第三层形成，刑罚反而可以更少地出现。贝卡利亚所谓“最好的预防”最终走向教育，正是在给这条路打开出口。

所以，把贝卡利亚称为“威慑主义者”并没有错，却不够。更准确地说，他把刑罚看成改变未来行为的一种制度信号，而信号是否有效，取决于它在时间、确定性、比例和社会经验中怎样被接收。这个模型已经比“重罚就行”复杂得多，也为现代犯罪治理留下了一个问题：我们到底是在制造恐惧，还是在制造可信规则？

> **带走的话**
>
> 威慑不是一记重锤，而是一条从规则、发现、证明、裁判到执行的路。路断在哪里，重锤就会落空；路越可信，刑罚越不需要靠残酷证明自己存在。

小账

他看见了什么？——威慑是发现、证明、裁判、执行连成的一条路。

接缝在哪里？——冲动、成瘾、绝望之下，计算式模型解释力下降。

第 29 章

现代刑罚第一次变成行为机制研究

写到这里，我们可以重新评价贝卡利亚的创新。他当然是刑法改革者，但如果只说“他主张刑罚温和”，仍然低估了他。他真正把一个旧问题改成了新问题：过去问的是“罪犯配得上什么痛苦”，他更愿意问“什么制度安排会产生什么行为结果”。前一个问题主要是道德分配，后一个问题开始具有机制研究的样子。

机制研究最重要的一步，就是把一个笼统变量拆开。贝卡利亚没有把“刑罚”视为一个黑箱，而是区分强度、比例、确定性、及时性、持续性和公开性。我们不必把这些概念硬说成现代实验科学，但它们已经使政策讨论从“更狠一点”进入“哪个环节真正起作用”。这一步在思想史上非常关键。

如果把前几章压成一张表，可以得到这样一条链：犯罪诱因出现——行为人形成预期——实施或放弃——国家发现——证据形成——裁判——处罚执行——公众观察——后文预期改变。每一环都不是孤立的。今天一次选择，会变成明天那片土的一部分。法律因此不是只在终点出现，它从一开始就通过预期参与行为。

这一结构最值得SDE注意的，是它已经包含明显回写。处罚执行后，不只是犯罪人受到影响，旁观者也更新经验；执法机关会根据结果调整策略；立法者会根据执行困难修改规则；社会对某种行为严重性的理解也可能改变。虽然贝卡利亚没有把这些写成统一方程，但他对法律、行为、社会意见之间循环关系的敏感，已经远超一种静态刑罚表。

可也正是在这里，幸福／效用视角开始向自由中心发动最强攻击。前面我们一直说贝卡利亚用个人自由限制国家：只让渡最少部分，刑罚不得超过必要。但到了刑罚效果章节，他又反复问怎样以较少痛苦取得较强预防，怎样让法律更有效地改变行为。这里的判断标准似乎越来越像公共幸福，而不是自由本身。

于是必须问：如果某种限制个人自由的措施能够显著提升总体安全，它是否就因此正当？贝卡利亚并没有给国家无限效用授权，因为必要性、法律预定和最少让渡仍在。但两条线确实同时存在：一条说“国家只能拿走被授权的最少自由”，另一条说“选择能以较少痛苦取得更大社会效果的办法”。自由像入口闸门，效用像闸门之后的优化标准。

这可能比简单宣布“贝卡利亚是自由中心主义”更准确。自由决定国家能不能进门，幸福／效用决定进门以后应该选哪种手段。如果这个双层结构成立，那么自由仍然可能是中心，因为效用计算只能在自由限定的范围内进行；但如果原典显示公共安全一旦被认为极重要就能突破自由边界，那么中心就可能发生移动。死刑问题将提供最强测试。

为什么？因为死刑把两条线推到极端。生命是个人能够失去的最大利益，公共安全则是国家最容易诉诸的最高理由。如果在这里，贝卡利亚仍坚持国家首先证明自己拥有权利，再证明死亡确实必要，他的自由中心就得到加强；如果他主要因为死刑“不够有效”才反对，那么幸福／效用的地位会大幅上升。下一编就是这场审判的主场。

还有一条不能忽略的反驳：贝卡利亚用行为后果评价刑罚，会不会把人降成可以操纵的对象？如果立法者只是问“怎样让人最容易被控制”，刑法就可能变成行为工程学。对此，贝卡利亚自己的限制是重要的：人不是统治技术的材料，刑罚只能在必要、合法和最少侵害的范围内使用。也就是说，他的行为机制研究从一开始就被权利问题夹住，不能彻底变成纯治理技术。

这个夹层恰恰是本书最有价值的地方。现代社会拥有比十八世纪强得多的预测、统计和算法能力。我们越来越能知道某种提醒、摄像头、风险评分、处罚频率怎样改变行为。如果只继承贝卡利亚的“有效性”而忘掉他的“权力从哪里来”，技术会把人变成可优化参数；如果只谈权利而拒绝研究制度效果，又可能让漂亮原则停在纸面。两边都不能少。

带走的话

贝卡利亚把刑罚从“应当让人多痛”推进到“制度怎样改变未来行为”。但他没有因此把人交给纯粹治理技术；自由是边界，效用是边界内的选择标准。死刑将检验这两层到底谁更根本。

小账

他看见了什么？——刑罚被拆成强度、确定性、及时性等可研究的变量。

接缝在哪里？——自由是闸，效用是闸后标准；公共安全能否冲开闸门？

第 六 编

国家能不能杀死一个公民

编 序

本编六章。第30章问国家从哪里得到杀人的权力；第31章问一个人能否把命交给国家；第32章问即使有权，死刑是否必要、有效；第33章让康德以应得出场反驳；第34章掀开两人各自的地板；第35章把死刑之争拆成四个制度连接。

本编是判决一的主场。死刑是最大且不可逆的强制，贝卡利亚不先谈残不残忍，而是先要国家出示授权，再证明非此不可。刑罚权的出生证明，在这里接受最严的查验。

读这一编，可以带着一个日常问题进去：物业管大门，能不能因此进你家？单位要把一个人永久拉黑，凭的是哪一次授权，错了还能不能改回来？

第 30 章

先别问死刑残不残忍，先问国家从哪里得到这项权力

谈死刑最容易从感情进入。有人看到恶性犯罪，第一反应是“这种人就该死”；有人看到国家处决生命，第一反应是“无论怎样都太残忍”。两种反应都真实，却都不是贝卡利亚最先问的问题。他先把情绪放到一边，问了一句极冷静的话：国家到底从哪里得到杀人的权利？（原典锚点：剑桥版第28章“死刑”。）

这是《论犯罪与刑罚》中最有力量的起手式之一。贝卡利亚前面已经建立了刑罚权来源模型：人为了摆脱彼此不安全的状态，把自己自由中的最小必要部分放进共同体，由这些小份额构成国家合法强制的基础。既然如此，任何刑罚都必须能够回到这个授权结构里。国家不是因为强大就拥有权利，权力必须能说明自己的来源。

死刑因此不是“最重的一档刑罚”那么简单。它逼社会契约理论回答一个边界问题：一个人真的曾经把“别人可以杀死我”的权力交给国家吗？如果每个人只愿意交出保障共同安全所需的最少自由，那么生命这种最大利益是否可能被包含在那份最小让渡之中？贝卡利亚的答案是否定的。

这里必须看清他的论证方式。他不是先说“生命神圣，所以死刑不行”，而是从国家权力来源内部发动攻击。哪怕你承认社会契约、承认刑罚、承认国家有强制权，也不能直接跳到“所以国家有权杀人”。每向前多走一步，都要重新证明授权范围。死刑要求最大强制，因此也应承担最大的来源说明。

这个问题可以用物业钥匙比喻。小区居民为了安全，把大门管理权交给物业，允许物业登记来访者、管理停车、维护公共秩序。这个授权当然可能给居民带来限制，但不能因为“你已经授权我管理大门”，就推导出“所以我有权进入你家拿走

任何东西”。权力不是一张空白支票。每一种更深侵入都需要新的理由。贝卡利亚对死刑的追问就是：刑罚权这张票据上，究竟有没有写到生命？

SDE在这里看见贝卡利亚最强的一次“发生化”。他没有把国家的生杀权当作王冠上天然附带的宝石，而是把它拆回生成路径：谁给的、给了什么、为了什么、给到哪里为止。死刑因此从一个“国家一直这样做”的历史事实，变成一个必须重新证明的规范结论。传统不再自动等于权利。

但这一论证也立刻暴露困难。社会契约真的是历史上每个人实际签过的协议吗？当然不是。贝卡利亚用“让渡”说明正当性，更像一种规范模型，而不是档案事实。既然如此，问“我有没有真的同意国家杀我”也不能只按字面合同理解。辩护者可能说：一个制度为了保护所有人的生命，必须拥有对极端威胁采取终极手段的权力，这种权力可以由共同生活的必要条件推出，而不需要每个人逐项签字。

这正是本书必须给贝卡利亚安排的最强反驳。否则“没人会同意别人杀自己，所以死刑没有授权”太容易被击破。真正的问题不是心理上愿不愿意，而是：一个把个人视为自由平等成员的共同体，是否能够把彻底消灭其中一员作为正常刑罚工具？如果能，条件是什么？如果不能，限制来自哪里？这才是社会契约论的深水区。

贝卡利亚自己并没有完全把门焊死。他承认，在国家处于失序、内战或某个人即使被剥夺自由仍能严重威胁共同体存续的极端情况下，杀死一个人可能被视为必要（原典锚点：剑桥版第28章“死刑”）。这一点非常重要，因为它说明他的绝对句式其实留下了非常态例外。死刑在正常法治状态下不是权利，但极端生存威胁可能被重新描述成战争或紧急防卫。

这使“死刑不是权利”变得更复杂。贝卡利亚似乎在区分两种情境：一种是正常刑法中的处罚，一种是共同体处在生存危机中的敌对行动。前者受社会契约和最少自由让渡限制，后者则接近战争逻辑。这种区分后来会引发很多问题，但至少说明他没有简单说“任何时候国家都绝不能导致一个人死亡”。

于是，本章真正的结论不是替死刑争论直接投票，而是改变证明顺序。一个国家如果要剥夺生命，不能只说“这个人罪大恶极”；它还必须证明：为什么这种权力属于国家，为什么在正常制度中仍然需要它，为什么没有其他足以保护社会的办法。贝卡利亚把死刑从情绪判断送进了举证程序。

> **带走的话**
>
> 贝卡利亚反死刑最锋利的一刀不是“死亡太残忍”，而是“请先出示授权”。国家越想使用不可逆的强制，越不能靠传统和愤怒替自己补签权力来源。

> **小账**
>
> 他看见了什么？——死刑先要回答权力来源：生命是否写在刑罚权那张授权票据上。
>
> 接缝在哪里？——他为内战、失序留下例外，正常刑罚与战争逻辑怎样分界仍未说清。

第 31 章

一个人能不能把自己的命交给国家

贝卡利亚的社会契约论证到了死刑这里，碰上一个有趣的悖论。他一方面说，国家权力来自个人自由的最小让渡；另一方面又说，人并不是自己生命的主人，否则他才可能把杀死自己的权利转交给社会。也就是说，为了否定国家杀人的权利，他同时使用了两条线：我们没有愿意交出生命，而且我们甚至未必拥有一项可以把生命交出去的完整处分权。

这使问题从政治哲学进入更深的主体哲学：一个人的生命究竟是不是“他的财产”？如果完全是，那么似乎可以像转让财产一样授权别人处置；如果完全不是，那么自由的边界又在哪里？贝卡利亚没有把这个难题展开成独立生命伦理学，但他敏锐地看见：生命不能被简单想象成口袋里最大的一枚自由硬币。

这里要防止一种过度现代化的读法。我们今天会想到自杀、安乐死、器官处分、危险职业等复杂问题，但贝卡利亚的直接任务只是说明国家刑罚权不容易从契约中推出。不能因为后来生命自主理论更复杂，就要求他预先回答全部。我们应当先守住他的论证功能：生命与一般自由份额不同，国家不能通过模糊的“你加入社会就同意一切”把它一并拿走。

用一个合同例子。一个人加入某个平台，同意平台可以处理他的账户信息、在违规时限制功能。平台不能因此写出一句“既然你同意平台规则，就视为同意我们永久删除你所有数字资产”。越接近不可逆、越重大、越超出通常目的的处置，越需要清楚而独立的授权。贝卡利亚对生命的处理，就是把这种“重大性—授权”关系推到极限。

但社会契约论的反方仍然会说：刑罚权从来不是犯罪人事后同意的结果。一个杀人者当然不会在被捕时签字同意处罚。国家的权力来自每个人作为共同体成员对一般规则的认可，而不是个案里的即时愿望。因此，不能用“没有人愿意被杀”直接否定死刑。否则任何刑罚都能被同一逻辑否定，因为也没有人愿意被监禁。

这个反驳很强，贝卡利亚真正需要依靠的就不只是“我不愿意”，而是“最少必要让渡”。监禁为什么可能被授权？因为在一定条件下，它可能是保护共同自由所需的较小强制；死刑为什么难以授权？因为它是最大且不可逆的剥夺，如果存在其他足以实现保护目的的办法，就超出了最少必要。这样一来，权利来源与必要性重新合流。

这也说明死刑问题不是一个孤立章节，而是前面全部理论的压力测试。社会契约负责说明国家从哪里获得强制；必要性负责说明为什么一定需要这种强制；比例负责说明它是否超过犯罪和社会防护所需；确定性与效用负责说明这种手段是否真的比替代方案有效；程序则负责处理不可逆错误风险。死刑把贝卡利亚所有工具同时拉到桌上。

从SDE看，这里出现一个值得保留的“中心主义反证”。如果贝卡利亚最终反对死刑只是因为它“不够有效”，那么幸福／效用可能才是根，自由只是一种修辞入口。可是他偏偏首先追问权利来源，而且在效用讨论前先否定了“国家天然拥有生杀权”。这说明自由至少不是装饰，它在逻辑顺序上承担了前置闸门。

不过，自由中心仍然不是绝对的。因为贝卡利亚在极端国家危机里承认消灭某个成员可能必要。这意味着当共同体生存本身受到威胁时，公共安全会压过平常自由边界。于是我们可以暂时形成一个更细的判断：在正常法治状态，个人自由是国家刑罚权的前置限制；在非常态生存危机，贝卡利亚允许必要性重新打开例外。中心不是没有例外，而是决定例外需要多高门槛。第56章还要为这个判断补上第三个问题：例外开在哪一类情形，由谁判定。

这类结构在普通生活中也能理解。父母有监护权，不等于可以为了“孩子好”决定一切；公司有管理权，不等于可以无限侵入员工私人生活；学校有纪律权，不等于可以用任何手段维持秩序。权力的来源总要和目的、范围一起解释。越严重的干预，越不能只拿“你已经接受管理”四个字作空白授权。

贝卡利亚的价值因此不只是给死刑一个结论，而是给所有强制权力立下一条审查习惯：每当权力说“你已经把这个交给我了”，都要继续问——什么时候交的？

为了什么交的？交了多少？有没有更轻的办法？这四问直到今天仍然是法治最重要的肌肉记忆之一。

> **带走的话**
>
> 贝卡利亚并没有靠“没有人想死”简单赢下死刑争论。他真正建立的是一条更严的规则：越不可逆的国家强制，越必须证明它确实包含在共同生活所必需的最小授权里。

> **小账**
>
> 他看见了什么？——生命不是最大的一枚自由硬币，越不可逆的处置越需要独立授权。
>
> 接缝在哪里？——「没人愿意被杀」挡不住「也没人愿意被关」，须靠最少必要让渡补强。

第 32 章

即使国家有权杀，杀人真的有用吗

贝卡利亚没有满足于权利论证。他很清楚，一个反对者完全可以说：即使社会契约理论证明得不够漂亮，只要死刑确实能保护更多生命，国家就仍有现实理由保留它。因此他第二次换战场——先把“有没有权”暂时放到一边，再问“有没有用、有没有必要”。这一步正好把前一编的刑罚机制全部带进死刑。

最先被拆的是强度神话。死刑当然是强度最高的刑罚之一，但贝卡利亚已经论证过，震慑效果不能只按强度计算。他认为，极端而短暂的痛苦未必比持续、反复进入社会经验的处罚更有影响。死亡是一瞬间的终结，而长期剥夺自由则可能在旁观者心里形成持续印象。无论我们今天是否接受他对终身苦役的具体比较，这个论证结构非常清楚：最重，不等于最有效。

这里必须对贝卡利亚保留批判距离。他为了证明死刑不必要，曾把终身苦役的持久痛苦描述成甚至可能比死亡更令人畏惧。今天的我们不能把这部分直接当作人道主义结论。反死刑不等于赞美终身折磨。恰恰相反，这暴露出他仍然强烈依赖威慑逻辑：只要一种非死刑手段能持续制造足够恐惧，就可以替代死亡。现代刑罚理论还必须继续追问人的尊严、改造可能和刑罚上限。

但他的核心策略仍然重要：如果存在足以保护社会的较轻替代方案，死刑的必要性就被削弱。必要性不是问“死刑有没有一点效果”，而是问“是否非它不可”。一项措施可能有用，却仍然不必要。比如用铁门可以防止孩子进入危险房间，但普通门锁已经足够，铁门就不是必要。国家尤其不能把“有一点额外效果”直接转换成“可以使用最大强制”。

贝卡利亚还指出，公开处决可能变成另一种景观。国家原本想通过死亡展示法律力量，旁观者却可能产生同情、厌恶、麻木，甚至把国家暴力本身当成模仿对象。也就是说，刑罚的社会效果并不由立法者单方面决定。国家想发送“不要杀

人”的信号，实际收到的可能还有“国家可以在庄严仪式中杀人”。信息在传播中会变形。

这正是回写问题。死刑不是一项只落在被执行者身上的措施，它还进入社会对生命、暴力和国家的理解。贝卡利亚的著名反问大意是：法律一方面谴责杀人，一方面又通过公开处决杀人，这种示范是否自相矛盾？我们不需要照抄他的修辞，只要抓住结构：法律手段会不会损害法律想培养的规范？

另一个问题是确定性。越严厉的刑罚，司法系统中的不同参与者越可能犹豫。被害人可能不愿推动极端后果，审判者可能寻找减轻路径，执行者可能迟延，政治权力也可能通过赦免介入。贝卡利亚认为，过度严厉会反过来制造不处罚。这样，强度与确定性出现反向关系：法律写得越凶，实际执行越不稳定。

死刑把这一悖论推到最高点。因为它不可逆，程序自然要求更谨慎；程序越谨慎，时间越长；时间越长，及时性越弱；社会越担心错杀，就越可能设置更多复核。于是极端强度往往带来极端程序负担。这不是反对程序，而是说明“最严厉”从来不是一个单独变量，它会改变整条D序列。

更不能忽视错误成本。贝卡利亚时代没有今天成熟的DNA鉴定、错案研究和程序救济资料，但只要承认审判可能出错，死刑就与其他刑罚出现根本差异：监禁错误至少在原则上还有停止、释放、赔偿和名誉恢复的可能，死亡则把纠错通道彻底关闭。不可逆性使证明责任不只是更高，而是承担一种无法完全修复的风险。

从SDE看，这里可以提出一个“回写死法”候选：一个制度如果把最终S设成绝对不可回写，它就失去了用新信息纠错的能力。死刑正是最典型的终局S——一旦执行，后来出现的证据只能改写社会评价，不能改写被执行者的生命状态。当然，这并不能单独证明所有死刑都错误，但它说明死刑把制度的反身修正能力压到最低。

这也把SDE自己拖进来。我们常说“回写”是发生系统的生命力，但是否所有社会决定都必须可逆？当然不是。很多行为本来就不可逆，法律也必须在不完美信息中作决定。真正的问题是：当一项决定不可逆时，制度应否提高前置证明、缩小

适用条件、优先考虑可修正替代方案。也就是说，发生学不是要求世界永远不定型，而是要求定型程度与错误成本相匹配。

到这里，自由中心与幸福中心的第一次交锋可以暂时落判。贝卡利亚反死刑不是只有一条线。自由告诉他：国家必须先证明自己拥有这项权力；效用告诉他：即使有权，也必须证明死亡确实比替代方案更必要、更有效。两条线不是互相取消，而是形成“双钥匙”：权利来源打不开，不能用；必要效用打不开，同样不能用。

这比把贝卡利亚简单归类为“功利主义反死刑”或“自然权利反死刑”更接近他的复杂性。他站在一个过渡位置：既从契约和自由限制国家，又用心理与社会效果检验刑罚。正因为两套语言同时存在，他才成为后来的报应论、功利论、权利论都必须回应的人。

◆ 理论深化　七个问题把第五、六编钉牢

一、严厉与确定为什么不是简单替代关系

不能把贝卡利亚写成‘只要确定，刑罚多轻都行’。刑罚仍须保持足以回应犯罪严重性和预防需求的强度，否则规则会失去规范表达与行为约束。真正的关系是：在达到必要效果以后，额外强度不应被当作自动增益；而提高确定性往往能减少对极端强度的依赖。二者不是零和，而是共同组成效果结构。

二、及时性必须与准确性相乘

如果只追求速度，程序可能把尚未显露的事实提前封顶；如果只追求无限谨慎，刑罚又会与行为脱节。可以把本编形成的规则写成：有效及时性＝必要程序内的最短时间，而不是日历上绝对最短。每一段延迟都要有功能，每一次提速都不能砍掉证明。

三、赦免为什么暴露了‘重刑—仁君’结构

贝卡利亚对频繁赦免的怀疑，不是反对怜悯，而是反对先制定残酷规则，再把人的命运交给统治者个人善意。若一个制度必须靠例外才显得人道，例外本身就是对一般规则的诊断。真正的仁慈应优先写进刑罚幅度、程序保障和普遍规则，而不是保留给少数幸运者。

四、死刑争论里‘权利’与‘必要’为什么必须分开

一项措施有用，不等于有权使用；一项权力存在，也不等于每次使用都必要。贝卡利亚最值得保留的结构，是把‘授权’和‘效果’分成两道门。现代公共权力审查中，这种分层仍然重要：先问有没有法定与正当权源，再问具体手段是否必要、比例和可替代。

五、终身苦役不能被现代读者直接接收

贝卡利亚为了反死刑而强调长期劳役的持续威慑，有其十八世纪背景，也暴露出他把刑罚价值大量放在预防效果上的局限。今天若继续这一论证，必须加入人的尊严、改造可能、过度惩罚和长期监禁本身的伤害。保留贝卡利亚的‘非死刑替代’结构，不等于保留他对替代刑罚痛苦程度的全部判断。

六、不可逆性为什么是新的SDE问题

SDE强调回写，但现实决策经常不可逆。死刑迫使我们把‘可回写性’从一般优点变成风险变量：越不可逆的决定，越需要更高的前置证据、更窄的适用范围、更强的复核和更认真地比较可逆替代方案。发生学并不拒绝定型，而是拒绝在高错误成本下轻率封顶。

七、自由中心暂时没有被效用中心推翻

本编最重要的中心诊断结果是：效用非常强，却暂时仍在自由闸门之后。贝卡利亚先问国家从哪里获得刑罚权，再问怎样处罚更有效；到死刑，他同样先问权利，再问必要与效用。因此目前更合适的结构不是‘自由或幸福二选一’，而是‘自由定边界，效用做优化’。后面进入预防、教育和公共幸福时，这个判断还要继续接受攻击。

◆ 本编增厚　四个深描：把刑罚机制真正压进SDE

深描一：为什么‘严厉—确定—及时’不是三个并列口号

如果只把贝卡利亚的刑罚论总结成三个词——温和、确定、及时——仍然会把他的思想写扁。三个变量不是放在一张清单上，而是彼此牵动。刑罚越严厉，证明和执行往往越谨慎，案件可能更慢；越依赖个案赦免，执行越不确定；越想追求迅速，如果削弱程序，又会降低事实判断的可靠性。也就是说，这不是三个可以同时无限拉高的旋钮，而是一组有张力的参数。真正的制度设计必须承认它们会互相改变。例如，假设某种犯罪法定刑极重。为了避免把一个人送入极端处境，警察、检察官、法官都会自然要求更强证据，辩护也会投入更多资源，案件时间因而拉长。若证据处在边缘，裁判者还可能通过解释、降格或量刑回避极端结果。纸面上看，国家提高了强度；路径里却同时发生确定性下降和及时性下降。单看法条，似乎更严；看整条D，未必更有效。反过来，若刑罚相对温和且规则稳定，制度更容易做到持续执行。执法者不必每次面对‘是否毁掉一个人一生’的巨大心理负担，公众也较少因处罚过重而抵触，程序能把注意力集中在事实与规则本身。于是温和不是与秩序相反，而可能是确定性的条件。贝卡利亚最深的地方，正在于他把‘人道’和‘有效’从对立面改成了可能互相支持的关系。但及时性又给这个关系增加限制。一个制度为了确保每次都处理，可能堆积大量案件；案件越多，程序越慢。若只追求快速，又可能通过简化程序牺牲准确性。于是最成熟的表达不是‘刑罚要轻、要快、要确定’，而是：在正当证明的约束下，以最低必要强度形成稳定而不

过度迟延的后果。这里已经出现一个近乎系统工程的问题——不是找到最大值，而是在多种价值与能力约束中寻找可持续组合。SDE的意义在这里不是给三个变量再贴一个新标签，而是把它们重新放回发生过程：S是最后显露出来的刑罚结论，D是从行为到发现、证明、裁判和执行的路径，E是司法资源、社会信任、执法能力、贫富差异、法律文化等条件。只改S而不改D、E，常常只能得到更重的纸面刑罚；D、E若没有能力支撑，S越重反而越容易空转。因此，所谓刑罚改革不能只盯着刑法典的数字。增加侦查能力、改善程序透明、减少选择性执法、缩短无意义积压、提高辩护质量、让判决理由可理解，这些看起来不像‘加刑’，却可能比加刑更直接地改变规则可信度。贝卡利亚两百多年前把目光从刑场移向制度链条，这正是他仍然现代的原因。

深描二：赦免为什么是一面照制度的镜子

贝卡利亚对赦免的警惕容易让现代读者不舒服。宽恕不是好事吗？如果一个人有特殊苦难、已经悔改，或者刑罚明显过重，为什么不能给他机会？问题在于，贝卡利亚针对的不是所有宽缓机制，而是一个特别的政治结构：法律先把刑罚写得残酷，然后由君主以私人恩典的姿态偶尔放人。这样，仁慈不属于法律，而属于人。一旦如此，公民面对的就有两套规则。第一套写在纸上，告诉你应受何种处罚；第二套藏在人际关系与政治权力里，决定你有没有机会例外。越严厉的法律，越需要例外来维持人性；例外越多，越说明纸面规则无法稳定执行。久而久之，大家会把注意力从‘我是否违法’转向‘我能不能找到被原谅的路径’。规则又退回猜人。贝卡利亚因此提出一个很有现代意味的想法：真正的仁慈应当由立法者提前放进法律，让刑罚本身温和而适度，而不是让执行者随意把已经规定的刑罚拿掉。这不是不要同情，而是要把同情制度化。今天我们会用量刑情节、缓刑、假释、减刑、未成年人特殊制度等更复杂的工具来实现差异化，而不必把一切寄托在一个最高权力者的临时恩赐上。SDE可以把赦免看成一种‘回写接口’。当一般规则碰到具体个案，发现规则过重、过粗或者没有容纳特殊情况，体系需要一个回写通道。问题是，这个通道是制度化、可说明、可复查的，还是完全人格化的？前者让个案经验有机会推动规则成长，后者只是用例外掩盖规则本身的缺陷。这又提醒我们，不要

把贝卡利亚的‘确定性’误写成绝不变化。一套真正稳定的制度，必须允许规则通过合法程序修改，也允许在明确标准下处理特殊情况。确定性不是每个人永远得到同样结果，而是差异本身也要有规则。一个制度若只能靠秘密电话、身份关系和临时批示创造差异，就没有把回写变成公共路径。普通人的组织生活里也一样。公司制度如果动辄写‘一律开除’，最后又靠领导逐个求情，真正运行的规则就不在员工手册，而在领导办公室。学校如果处分标准极重，最终靠校长‘酌情给机会’，学生学到的也不是规则，而是谁更值得同情。贝卡利亚不愿意把自由放在这种恩典结构里。

深描三：死刑把‘错误’从概率问题变成终局问题

任何司法制度都不可能把错误概率压到绝对零。证据可能遗漏，证人会记错，鉴定技术会更新，办案者会受偏见影响，新的事实也可能在多年后出现。通常情况下，制度用上诉、再审、纠错、赔偿等机制承认自己会错。也就是说，成熟法治并不是宣称‘我永远正确’，而是设计‘如果我错了，怎样回来’。死刑之所以在这一结构里特别突出，不只是因为它最重，而是因为它切断了最重要的回写。判决可以改，记录可以更正，国家可以道歉，但生命状态不能恢复。新的证据再强，也只能证明‘当时错了’，不能把人带回来。于是错误从一般概率风险变成不可逆的终局风险。这并不意味着只要一种决定不可逆，就绝对不能作出。外科手术、战争防卫、紧急救援甚至日常人生选择都可能不可逆。真正的制度原则应当更精确：不可逆性越高，前置证明负担越重，可替代方案比较越重要，决定权越需要分层，复核越应充分。贝卡利亚虽然没有用这套语言，但他把死刑先送进‘必要性’检验，已经给出了同方向的入口。这里也能检验SDE自身。发生学常强调回写，容易让人产生一种错觉：仿佛只要保持开放、不封顶，一切问题都能解决。现实不是这样。制度必须在有限时间里作出结论，很多结论一旦执行就无法完全撤回。SDE若要成为严肃方法，就必须研究‘不可逆定型’而不是回避它：什么时候必须定型，定型前需要什么门槛，定型后还能留下哪些补救。死刑因此给全书增加了一种新的死法诊断：不是‘没有变化’，而是‘错误发生后无处回写’。这可以暂叫‘终局回写死’。它不自动判决死刑制度必然不正当，却要求任何支持者正面回答：当制度知

道自己并非绝对不会错时，为什么仍允许一种无法纠错的结果？反对者同样需要回答：如果出现持续而无法通过其他手段控制的极端威胁，国家是否仍绝对不能采取致命行动？贝卡利亚自己通过‘正常刑罚／非常态战争’的区分试图处理后一问题。他并没有把所有致命强制放进同一盒子。在稳定法律秩序中，他否定把死亡当作普通刑罚的必要；在共同体处于失序、某人仍能威胁国家存续的极端情境中，他留下了例外。这种区分是否成功，后面与康德以及现代刑罚理论对撞时还要继续审。

深描四：本编的九格局部扫描——自由为什么仍在中轴

把第25—32章放回九格，可以看见自由、幸福、创造三种意义目标并不是平均出现。理念界里，确定性和赦免问题首先关心规则能否稳定限制权力；现实界里，刑罚强度、身体痛苦和死亡成为国家强制的物质形态；自我界里，恐惧、习惯、希望、时间感和风险判断决定刑罚怎样进入人的选择。三界都活着，但意义轴上仍有一条最明显的中线：自由。为什么不是幸福？因为贝卡利亚确实不断讨论公共安全、较少痛苦和更大预防效果，幸福／效用很强。然而这些优化从来不是在无限空间里进行。国家不能为了多一点安全就任意增加刑罚，不能为了更高破案率就使用酷刑，也不能因为死刑可能有某种震慑就跳过权利来源。效用必须先穿过自由与合法性的门。为什么不是创造？因为贝卡利亚也看见法律会制造新的行为结构：过度立法会制造犯罪身份，执行历史会制造预期，公开刑罚会制造社会对暴力的态度。但他很少把‘创造新的社会可能’本身放在刑罚的首要目标。创造更多作为结果和副作用出现，不像自由那样反复承担边界，也不像幸福那样承担效果评价。因此，本编的局部九格可以暂时写成：理念界最强的是规则自由——罪刑法定、确定执行、反对私人恩典；现实界最强的是身体自由——强度受必要性限制、死刑必须额外说明；自我界最强的是选择自由——刑罚通过确定性、及时性、持续性影响判断，但人的选择又受到激情、希望、习惯和处境塑造。三处自由并不是同一种自由，却能沿一条轴连起来。这也让‘主体先在’的批判再次收窄。贝卡利亚并不否认人的选择会被环境塑造，相反，他的威慑理论几乎完全建立在环境会改变选择上。真正可能先在的，是规范资格：无论一个人经验上怎样形成，国家都必须先把

他当作拥有某种不可随意吞没的自由份额的成员。这个先在如果是规范护栏，就不能简单当作理论错误；只有当它阻止我们理解现实中的自由能力如何被制度生产和损害时，SDE才有继续补足的空间。所以本编结束时，母题没有变得更轻，反而更准确：贝卡利亚的伟大是把国家权力发生化，并用自由为发生设置边界；他的未竟之处，不是完全看不见人的发生，而是尚未把‘规范上必须承认的自由’与‘经验中不断生成的自由能力’系统地接成一体。后面的教育、预防和社会环境章节，正要测试这座桥能不能在他自己那里找到。

深描五：从贝卡利亚回看‘刑罚国家’与‘规则国家’的分野

第25—32章还可以压出一个更大的政治—法律结构，但这里不需要把它写成政治立场，而是作为制度类型来理解。所谓‘刑罚国家’，并不是刑罚多的国家，而是一个主要靠展示惩罚能力来证明秩序存在的制度：出了问题先问怎么罚得更重，规则执行不稳定时用更高强度补偿，公众安全感也被不断绑定在更重处罚上。所谓‘规则国家’，则更在意另一组问题：规则是否事前公开、事实是否可靠证明、后果是否稳定、不同案件是否一致、错误是否能够纠正。两者都会处罚，区别在于秩序的信用究竟建立在痛苦强度，还是建立在规则可预期性上。贝卡利亚显然想把制度从前一种结构往后一种结构推。他反对酷刑，不是因为国家从此不能强制；他反对重刑迷信，也不是因为社会不需要防卫；他反对频繁赦免，更不是因为法律不应有人情。几条线放在一起，真正目标是把公共秩序从‘统治者有多可怕’改成‘规则有多可信’。这一步解释了为什么罪刑法定、确定性、及时性和温和刑罚在他那里能组成同一套工程。规则国家当然也会失败。规则可能本身不公，执行可能形式平等却实质偏斜，技术可以让错误更稳定，制度也可能把少数人的处境长期忽略。贝卡利亚没有替我们解决这些问题。但他的贡献是先改变秩序的证明方式：国家不能再只靠展示暴力证明自己有效，而必须把权力放进公开的理由、程序和一般规则中。这也说明为什么本书一直不愿把贝卡利亚写成‘刑罚越轻越好’的简单人道主义者。他真正追求的是让强制变得有来源、有边界、有路径、有理由。必要的时候可以处罚，但处罚必须是规则秩序的最后手段之一，而不能成为规则失灵后的情绪补偿。如果把这一点带到今天，最值得警惕的不是社会要求严惩某一种犯罪

本身，而是任何时候只要问题复杂，我们就本能地把‘提高刑罚’当作唯一可见动作。很多风险真正需要的可能是更好的发现机制、更快但更可靠的程序、更少的制度漏洞、更均衡的资源、更清楚的规范，甚至是教育与社会支持。加重刑罚最容易被看见，却未必是最深的治理。贝卡利亚的现代性，就在于他迫使我们把目光从刑罚终点重新移回整条发生路径。

◆ 给普通人的三个现场　把贝卡利亚从刑场带回日常

场景一：公司罚款很重，为什么员工还是照样违规

制度写着一次严重违规扣半月奖金，但检查极少、领导熟人可以例外、执行时有时无。员工真正学到的并不是‘违规代价巨大’，而是‘关键在于别被特定的人看见’。这就是第26章的确定性问题：纸面强度再高，也抵不过长期形成的低执行预期。

场景二：学校为了震慑作弊，第一次就开除

重罚可能显得立场鲜明，却也可能使教师面对边缘案件时不愿上报，同学不愿作证，家长全力争夺程序漏洞，最终形成‘开除很重，但真正落下很少’。如果处罚有梯度、程序明确、每次作弊都稳定处理，反而可能形成更可信规则。

场景三：一个不可逆的人事决定

单位公开辞退并永久列入行业黑名单，后来发现证据有误。即使道歉，职业与名誉也难完全恢复。这不是死刑，却有相似结构：结果越不可逆，前置证明越应严格，替代措施越应认真比较。SDE的‘回写’提醒我们，制度必须把纠错能力当成设计变量。

◆ SDE对账　从‘强度’到‘路径’，再到死刑双钥匙

一、本编新增的“看见”

至少新增五处：贝卡利亚看见严厉会反过来降低确定性；看见时间距离会改变犯罪与刑罚的联结；看见赦免可能暴露法律本身过重；看见死刑不仅要接受权利审查，还要接受必要性和效用审查；看见刑罚方式会回写公众对暴力和法律的感受。这些都说明他不是简单的静态契约论者，而已经在多个位置观察制度与行为的相互塑造。

二、本编对“收回”候选的影响

‘先放人，再限权’仍然是最强总候选，但本编让它变得更难下。因为贝卡利亚并不把经验中的人理解成完全静止：他明确关心习惯、激情、时间、概率、社会意见怎样改变行为。所谓主体先在，越来越应被限定为规范起点，而不能扩大成‘他认为真实的人完全不受环境塑造’。后续如果要正式判中心主义，必须抓到规范主体为何不能被自身的发生条件回写。

三、自由中心与幸福中心的第一次比分

本编暂时支持‘自由定边界、幸福做优化’。依据不是抽象标签，而是论证顺序：刑罚权先从自由让渡说明来源，之后才讨论怎样有效；死刑先问国家是否拥有该权利，再问是否必要与有用。因此自由在结构上仍较像入口，幸福更像进入合法范围后的目标函数。但这只是暂判，后面的预防犯罪、教育与公共幸福章节可能改变比分。

四、D第一次成为主角

前面多次使用‘路’作比喻，本编则真正让D承担解释功能：犯罪—发现—证明—裁判—执行—社会观察—后文预期。刑罚效果不是某个固定S单独造成，而是在

序列中发生。若把整条路径压成‘刑罚十年’，就会把真正产生行为差异的机制全部藏掉。

五、死刑给SDE提出的反身难题

如果SDE把‘能回写’当作活体系标志，会不会因此预设一切决定都应可逆？现实中不存在这种世界。更好的表述应是：体系必须把不可逆性作为风险成本纳入前置判断，不能把不可回写本身自动判死。死刑因此不是用一个SDE口号就能结束的议题，而是逼SDE说明自己如何处理终局决定。

桥一：第33章进入康德

后文应从‘康德为什么不能接受贝卡利亚’开始。康德将把报应论推到正面：如果刑罚只为未来效用服务，是否把犯罪人当成了别人安全的工具？这将是贝卡利亚最强的近邻攻击之一。写作时不能把康德写成‘残酷派’，而要把他维护人的责任主体地位这一理由说足。

桥二：第34—35章完成死刑编

在康德—贝卡利亚对撞之后，要用SDE判断两人各握住什么：贝卡利亚握住权力来源、必要与制度效果；康德握住责任、应得与不得把人纯粹工具化。然后回答‘死刑争论真正争的不是善良，而是国家、个人与正义的关系’。

桥三：第36章以后转向法律如何制造犯罪

死刑编结束后，不要立即跳到中心主义总判，而应沿原提纲进入‘法律不仅抓犯罪，也制造犯罪’。过度禁止、秘密控告、污名身份、刑罚回写社会，都是贝卡利亚已经看见的发生线索。这一编会决定我们是否仍有资格说他的回写不足。

桥四：继续保留改判权

本书当前最有价值的不是把‘国家发生，主体先在’喊得越来越响，而是不断给它找反证。如果后续发现贝卡利亚通过教育、社会环境、习惯与制度设计已经给经验主体充分发生空间，就要把批判继续收窄。要求抓到真正接缝，而不是为了母题牺牲原典。

> 带走的话
>
> 贝卡利亚对死刑设置了两把锁——国家先证明“我有权”，再证明“非这样不可”。死刑之所以成为全书最强试金石，是因为它迫使自由与效用同时交卷。

> **小账**
>
> 他看见了什么？——最重不等于最有效：严厉压低确定性，公开处决教人暴力，死亡关闭纠错。
>
> 接缝在哪里？——他以终身苦役更可怕来替代死刑，仍靠威慑逻辑，尊严与刑罚上限没进论证。

第 33 章

康德为什么不能接受贝卡利亚

如果只看结论，贝卡利亚和康德在死刑问题上几乎站在桌子的两端。贝卡利亚把死刑送上审判席，要求国家证明这种不可逆的刑罚究竟从哪里取得权利，又是否真的必要；康德则在《道德形而上学》的法权论部分坚持，对于谋杀，正义要求一种与罪责相称的报应，不能因为社会效果、犯罪率或其他人的幸福而把刑罚改成别的东西。

最容易把这场争论讲坏的办法，是把它翻译成“一个有人道，一个没人道”。这样写，康德只剩下冷酷，贝卡利亚只剩下同情，哲学问题反而消失了。康德真正担心的，不是社会罚得不够重，而是一个有责任能力的人被国家当成实现别人安全的手段。如果国家对甲施加痛苦，理由只是“这样乙、丙、丁看了以后就不敢犯罪”，那么甲这个人就有被工具化的危险：他承受刑罚，不再主要因为他做过什么，而是因为他的痛苦能对别人产生效果。

用一个日常例子看得更清楚。学校抓到一个作弊学生，为了“杀一儆百”决定把他开除。管理者说，开除一个人能让全校一千人都不敢再作弊，所以总体效果很好。康德式追问会立刻出现：这个学生该承担什么后果，应当根据他的行为与责任来决定，还是根据其他九百九十九个人需要多大震撼来决定？如果后者可以无限扩张，那么只要效果足够好，似乎就可以把任何一个人的处罚加重到足以吓住别人。这里出现的危险，正是把被处罚者从“责任主体”变成“教育别人用的道具”。

所以，康德刑罚论最硬的一条线不是“报复”，而是“应得”。行为已经发生，责任已经成立，刑罚的正当根据首先必须从这个人的行为和罪责内部产生，而不能从第三人的利益里借来。也因此，他才说刑罚不能仅仅作为促进犯罪人本人或社会其他利益的手段；一个人必须先因为犯罪而成为可罚者，之后才谈社会可能从刑罚中得到什么附带好处。

这与贝卡利亚的方向恰好相反。贝卡利亚最著名的转身，是把刑罚从“为已经发生的恶付账”拉向“防止新的恶发生”。刑罚的目的不是让敏感的生命单纯受苦，也不是把过去的罪抹掉，而是阻止犯罪人再次伤害社会，并阻止其他人重犯类似行为。于是他的目光天然向前：这项刑罚能不能产生必要的预防效果？有没有更轻却同样有效的办法？如果没有额外效果，额外痛苦凭什么存在？

两个人因此像站在时间箭头两端。康德先向后看：他做了什么，因此应得什么。贝卡利亚先向前看：这样处罚以后，会发生什么，因此是否必要。一个把“罪责”放在入口，一个把“必要的社会防卫”放在入口。若只用一句“报应论对功利论”概括，还不够，因为他们真正冲突的是刑罚正当性的起点。

这时，死刑成为最尖的试金石。贝卡利亚说，国家从个人为了安全而让渡的有限自由中取得权力，个人不会把自己的生命完整交出去；即便暂且不谈这一权利来源，死刑在通常社会状态下也很难证明比长期、确定的自由剥夺更有必要。康德则认为，对故意杀人这样的罪，正义的比例关系本身要求生命对生命；如果因为有用或无用而改变，就把正义交给效果计算。

康德甚至设计了一个极端思想实验：假如一个岛上的公民社会决定解散，所有人第二天各自离去，最后一个已经被判决的谋杀犯怎么办？此时已经没有未来社会可以威慑，没有下一代犯罪率需要降低，也没有共同体长期安全可以改善。康德的回答仍然是：正义的债不能因为社会即将消失而消失。这个例子故意把一切未来效用抽走，只留下“这个行为是否应当得到相称回应”。

这个思想实验对贝卡利亚很有杀伤力。因为如果刑罚只由未来效果证明，那么在一个已经没有未来的共同体里，刑罚似乎失去理由。康德借此告诉我们：刑罚中可能有一种不能完全还原为治理技术的东西。法律不是只问“怎样让人以后更听话”，还要问“对已经发生的严重不法，我们是否欠受害者、欠公共正义一种回应”。

但康德的刀同样会遇到自己的接缝。把“不得把人纯粹当工具”推到刑罚领域，是非常强的限制；可从“罪责必须是处罚根据”走到“谋杀者必须被处死”，

中间仍然需要说明：为什么责任主体的尊严一定要求同态或近似同态的报应？为什么终身自由剥夺不能表达足够的责难？为什么国家执行死亡不会出现同样不可逆的错误风险？这些问题不能用“正义如此”一句话封口。

所以，本章不判谁赢。它只把真正的冲突抬高了一层：贝卡利亚怕国家为了自己想象的秩序而多伤害一个人；康德怕国家为了别人将来的利益而把已经犯罪的人变成工具。一个防“过度强制”，一个防“工具化”。二者都从尊重人出发，却得到相反的死刑结论。

> 带走的话
>
> 贝卡利亚问“这项痛苦还有没有必要”，康德问“这个人究竟应不应得”。死刑之争第一次从轻重之争变成正当性起点之争。

> **小账**
>
> 这一章看见了什么？——康德怕受罚者被当成吓唬别人的道具，刑罚根据须出自本人罪责。
>
> 还要追问什么？——从应得走到必须处死缺一步：为何终身剥夺自由不足以表达责难。

第 34 章

两个人到底各自踩着哪块地板

哲学争论最有意思的地方，往往不是双方说了不同的话，而是双方说话时脚下踩的地板不同。只盯着结论，就会觉得一个说“不要杀”，另一个说“必须杀”；把地板掀开，才会发现他们在回答的其实不是同一道题。

贝卡利亚的第一块地板，是有限授权。国家不是超出社会的神秘主体，它的处罚能力必须从共同生活的形成中得到说明。个人为了让其余自由获得安全，交出的是必要的最少部分。由此得到一条强限制：如果某种刑罚超出维护共同安全所需的最小强制，国家就不能只凭“我是国家”取得多出来的那一部分权力。

第二块地板，是经验中的人。人会感受痛苦，会受概率影响，会对确定性、及时性、社会舆论和制度环境作出反应。刑罚必须进入人的实际行为机制，而不能只停留在抽象道德账本。因此，贝卡利亚不断比较轻重、快慢、确定与不确定、公开与秘密。这里的“人”是会在生活中改变选择的人。

第三块地板，是公共幸福。好的法律不是把国家意志压到每个人头上，而是尽可能减少共同生活中的不幸。刑罚之所以有资格存在，是因为它服务于一种共同安全结构；它越过必要限度，就从保护自由转成新的伤害。

康德的地板则不同。第一块，是外在自由的法权秩序。国家强制之所以可能正当，不是因为大家在经验中做过一场真实签约，而是因为一个公共法权状态是自由能够相互共存的理性条件。社会契约在这里首先是一条规范理念，不是历史上每个人亲手签过的合同。

第二块，是责任主体。一个能够被刑法归责的人，不只是被环境推动、会对痛苦作反应的动物；他也是一个能够把行为算到自己头上、因而承担责任的人。如果刑罚只看行为以后能产生多少社会效用，责任主体的这一地位会被冲淡。康德因此坚持：先有罪责，后有刑罚，不能反过来为了社会需要找一个人来承受。

第三块，是严格正义与比例。正义不应随着公众恐惧、舆论热度或治理方便任意摇摆。对什么样的罪应有什么样的法律回应，必须有一个不依赖短期效果的原则。康德选择了报应意义上的平等作为这把尺。

把两组地板并列，会发现一个很有意思的共同点：两个人其实都不相信统治者的临时善意。贝卡利亚用事先法律、必要性和社会契约限制权力；康德用不可把人仅作手段、罪责先行和正义原则限制权力。二者都想让处罚摆脱“今天谁生气谁决定”的状态。

所以，如果把这场争论写成“功利主义没有原则，报应主义才讲原则”，并不公平。贝卡利亚当然看效果，但他的效果计算一直被自由、必要和法律预先规定围住。他并没有说只要降低犯罪率，什么刑罚都可以。反过来，把康德写成“完全不管效果”也太平。更准确地说，康德不允许效果成为对这个具体人施刑的原始正当根据，但并不意味着一个国家在刑罚制度设计中完全不关心安全。

SDE在这里可以做一次“位置三分”。如果用“显露—差异序列—特征纠缠”的大白话，康德更用力地守住已经定型的规范显露：罪责、责任、平等、应得，必须让刑罚有一个不随治理便利移动的骨架。贝卡利亚则更用力地追问刑罚在现实路径中怎么起作用：强度、确定性、时间、示范、习惯，都会改变结果。两个人都没有完全缺掉另一面，只是侧重不同。

再换到三界。康德的主场明显更靠近理念界：什么是权利、什么是正义、什么原则可以被普遍化。贝卡利亚则把理念界的权利来源不断拉回现实界和自我界检验：人的身体怎样被折磨，恐惧怎样发生，制度怎样影响行动，刑罚怎样被公众看见。可贝卡利亚也绝不是纯现实中心，因为没有“自由让渡”这一理念起点，他对国家权力的限制会失去根。

因此，不能草率说“贝卡利亚只讲效果，康德只讲原则”。更准确的说法是：贝卡利亚把原则送进现实效果里接受必要性复查；康德把现实效果挡在罪责成立之后，防止效果反过来制造可罚性。一个不断问“有没有少伤害一点的路”，一个不断问“有没有把这个人当成别人目的的工具”。

两个人的薄弱处也因此呈镜像。贝卡利亚若把未来效果推得太远，会面临“为了多数人利益可不可以牺牲少数人”的危险；康德若把应得推得太硬，会面临“即使没有任何防卫必要、即使制度有不可纠正的错误风险，也必须完成报应吗”的压力。一个需要解释效用的边界，一个需要解释正义为何必须采取这种具体形式。

这就是为什么死刑问题难以靠一句“生命最宝贵”或“杀人偿命”结束。两句都抓住了人的强烈直觉，却没有把国家权力的结构说清。真正的哲学争论必须继续追问：谁授权、谁应得、谁受益、谁承担错误、有没有替代、决定能不能改。

> **带走的话**
>
> 贝卡利亚把刑罚放进“权力来源—必要效果”的地板上，康德把刑罚放进“责任主体—应得正义”的地板上。两人的共同敌人都是任意，真正分歧是国家应当先向未来负责，还是先向已经发生的罪责负责。

小账

这一章看见了什么？——一边踩有限授权、经验中的人、公共幸福，一边踩法权、责任主体、严格正义。

接缝在哪里？——效果推远会牺牲少数，应得推硬会无视错判风险，两人弱点正成镜像。

第 35 章

死刑争论真正争的不是“善良”

公众讨论死刑时，最常出现两种互相刺痛的话。一边说：“如果你真正同情受害者，就不会替杀人者说话。”另一边说：“如果你尊重生命，就不会支持国家杀人。”两句话都把对方推到道德人格的被告席上，于是讨论很快从制度问题变成“谁更善良”。贝卡利亚和康德的对撞提醒我们，真正困难的问题恰恰不在善良。

第一层争论，是国家与个人的关系。贝卡利亚会问：一个为了保护自由而成立的国家，到底得到多少强制权？国家的力量来自共同生活，而不是来自超越个人的神秘主权，因此越极端的强制越要说明授权。死刑最特殊之处，是它不只是限制自由，而是终止一个人的全部未来，所以“你凭什么”必须比普通刑罚问得更重。

第二层争论，是个人与责任的关系。康德会问：如果一个人是能够对自己的行动负责的主体，那么面对故意杀人这样的极端不法，我们是否可以只把他看成一个需要被改造、被威慑或被管理的风险源？如果完全不谈应得，刑罚会不会反而取消了他的责任地位，把他降低成行为工程的对象？

第三层争论，是正义与效果的关系。贝卡利亚担心一种没有效果也继续制造痛苦的正义会变成仪式化暴力；康德担心一种只问效果的刑罚会在需要时牺牲个人。两人的问题都没有过时。现代刑罚理论仍不断在报应、威慑、隔离、矫正、修复等目标之间调节，没有一把尺能把所有价值无损换算。

第四层，是错误与不可逆的关系。十八世纪的贝卡利亚没有现代法医学、DNA证据和错案研究，但他的程序关切已经让我们看到：证明机制会失败，酷刑会制造假供，秘密程序会放大错误。只要承认制度可能错，死刑就多出一个普通刑罚没有的结构问题——错误发生后，没有真正的回程。

这也是上一批提出“不可逆性”时必须保持克制的原因。SDE不能因为重视回写，就简单宣布“一切不可逆决定都不正当”。医学中的截肢、战争中的防卫、紧急状态下的某些选择，都可能不可逆。真正合理的判断是：越不可逆，越要提高前置证明、替代方案比较和决定程序的要求。不可逆不是自动禁令，而是风险倍增器。

第五层，是国家作为榜样的问题。贝卡利亚在死刑问题上有一种极富张力的直觉：法律一方面告诉人“不要杀”，另一方面国家又以最庄严的形式完成杀人，这种示范到底教会公众什么？这里不是说司法处决和私人谋杀在法律上相同，而是在问制度行为会不会回写社会对暴力的想象。国家不是站在社会外面，它做的事同样会成为社会经验的一部分。

康德会反击：恰恰因为私人杀人与依法执行刑罚不是同一回事，不能把形式相似当成道德同一。国家的强制如果严格根据公共法律和责任原则实施，不是私人复仇。这个反击非常重要，它阻止我们用“国家也杀人”这样一句表面同构替代完整论证。SDE的纪律也要求承认：同构不是反驳，必须说明两种行为在制度功能上到底哪里相同、哪里不同。

第六层，是受害者的位置。把死刑完全写成“国家与罪犯”二人关系，会漏掉受害者及其家属、公共安全与社会正义感。但把受害者痛苦直接等同于刑罚尺度，同样危险。受害者应被看见，却不能被要求承担制定刑罚原则的全部责任；国家必须在个案痛苦与一般规则之间保持区分。

所以，死刑争论真正争的是四个连接：国家权力怎样连到个人自由，犯罪行为怎样连到责任，责任怎样连到刑罚，刑罚又怎样连回社会。每一条连接都有可能被写得过短。只说“契约没有授权”可能低估责任；只说“杀人者应死”可能跳过国家权力与错误风险；只说“死刑没有威慑效果”可能把正义完全交给统计；只说“社会需要表达谴责”又可能不解释为什么表达必须采取死亡形式。

对普通人而言，这种拆法还有一个意义：当我们在重大事件中迅速站队时，最好先分清自己其实在回答哪一道题。你是在谈国家有没有这种权力，还是谈某个人

应不应得；是在谈刑罚是否有效，还是谈即使无效也必须完成的正义；是在谈个案情绪，还是一般制度。如果题都没有分开，争论越激烈，越可能只是各说各话。

前文在这里结束死刑编，不给读者替代性最终答案。因为本书的任务不是替读者决定死刑立场，而是把两种最强理由都推进到各自的边界。贝卡利亚让国家证明为什么必须伤害，康德让国家证明为什么这个人首先应受处罚。现代制度若要处理极端刑罚，至少不能丢掉其中任何一道问。

> 带走的话
>
> 死刑争论不是“谁更善良”，而是四个制度连接能否成立——权力与自由、行为与责任、责任与刑罚、刑罚与社会。贝卡利亚和康德各守住其中一部分，也各留下自己的免检区。

> **小账**
>
> 这一章看见了什么？——死刑争论可拆成六层，核心是权力、责任、刑罚、社会四个连接。
> 还要追问什么？——「国家也杀人」只是表面同构，处决与谋杀在制度功能上的异同须说明。

第 七 编

法律不仅抓犯罪，也制造犯罪

编 序

本编五章。第36章讲禁止越多为何违法越多；第37章讲秘密告发怎样造出人人互疑；第38章讲被叫成罪犯以后身份怎样被回写；第39章讲刑罚怎样反过来改变社会；第40章收拢这些线索，看贝卡利亚离SDE最近的一刻。

本编是判决三的主场。规则密度改变违法数量，秘密程序改变人与人的关系，标签改变一个人的去路，这些都是贝卡利亚亲眼看见的发生。本编也因此正式收回「他看不见主体发生」的过强判断。

读这一编，可以带着日常问题进去：校规一条条加上去，违纪的学生是变坏了，还是网织密了？单位的匿名举报箱，让人更敢说真话，还是更不敢说话？

第 36 章

禁止得越多，犯罪为什么可能越多

前面六编一直把刑法想象成一只受约束的手：国家发现已经发生的伤害，然后在法律范围内作出回应。现在要把镜头反过来。法律并不只是站在犯罪发生以后，它还站在犯罪发生以前。它画边界、贴标签、改变诱因，也改变人们彼此怎么看。于是一个更危险的问题出现：法律会不会一边说自己在减少犯罪，一边制造出新的犯罪？

贝卡利亚在讨论“怎样预防犯罪”时给出过一个极其锋利的提醒：禁止大量琐碎行为，并不等于预防真正有害的行为；相反，它会把原本并无严重公共危害的日常活动重新命名为犯罪。法律越贪心，犯罪目录越长，社会就可能在统计意义上“越来越犯罪化”，但共同生活未必因此更安全。

这句话和今天常说的“过度犯罪化”很接近，但不能把后来的完整理论倒装给贝卡利亚。他当时没有我们今天的监管国家、网络平台和复杂行政刑法，更没有现代犯罪化原则的全部术语。他真正看见的是一个基础事实：刑法不是天然罪名的登记簿。立法者每写下一条禁止，都在改变“什么行为会把一个普通人变成罪犯”的边界。

想象一个学校。为了防止学生出问题，学校先禁止考试作弊，这很容易理解；然后禁止课间使用手机、禁止在走廊奔跑、禁止未经批准组织社团、禁止在宿舍讨论敏感话题、禁止晚十分钟归寝、禁止携带某些日用品。每增加一条规则，管理者都能给出某种安全理由。最后，学校可能拥有一套几乎人人都曾违反的纪律。此时“违规者很多”并不一定说明学生突然变坏，也可能说明规则的网越织越密。

这就是“法律制造犯罪”的第一层含义：它制造的不是伤害本身，而是法律身份。杀人造成的死亡不是法律创造的，盗窃造成的财产剥夺也不是法律凭空制造的；但一个行为是否被国家提升为“犯罪”，却需要经过法律的界定。只要这一步

存在，立法就必须解释为什么要动用刑罚，而不能把“我已经写进法条”反过来当成正当性的证明。

第二层含义，是规则会改变行为路径。走私就是贝卡利亚喜欢分析的例子之一。若一种商品的税负、禁运和边境制度改变，走私的利润和风险也跟着变。此时法律并不是在一片完全不变的现实上追赶犯罪，它自己参与了收益结构的形成。越是忽略这种反馈，越容易把制度制造的诱因全部归结为个人道德失败。

第三层含义，是过多禁止会稀释法律的权威。如果一个社会拥有大量普通人认为无害、却随时可能触犯的刑罚规则，执行就很难保持真正确定。规则一多，选择性执法的空间变大：人人都有可能违规，却只有部分人被挑出来处罚。这样一来，贝卡利亚最珍惜的“确定性”反而被过度立法自己破坏。

这里出现一个漂亮的回路。立法者为了提高安全，增加禁止；禁止增加，潜在违法者增多；执法资源不足，只能选择性执行；选择性执行削弱可预期性；可预期性下降，人们转而依赖关系、运气和讨价还价；最后，原本为了减少任意而增加的规则，反而制造了更多任意。规则不是越多越法治，关键是规则是否有足够公共理由并能够稳定、公平地执行。

第四层含义，是法律会重新定义道德感。一个行为如果长期被法律标成最严重的罪，人们可能逐渐把法律评价内化为道德评价；反过来，如果法律把真正严重的伤害只当成轻微违规，社会也可能降低对它的敏感。贝卡利亚在公共耻辱问题上已经看见法律与社会意见之间这种互相作用：法律如果滥用耻辱，反而会让耻辱本身失去力量。

SDE在这里要特别防止一个误区：说“犯罪是法律制造的”并不等于说“真实伤害不存在”。发生学反对的是把犯罪类别当成完全先在、与制度无关的自然物，不是否认死亡、伤害、财产损失和恐惧这些事实。一次袭击发生了就是发生了，不能因为法律分类具有历史性，就把受害者的现实损害说成语言游戏。

因此，本章更准确的判断是：现实中的伤害与法律上的犯罪不是同一层东西。前者可能先于某一部法典存在，后者则需要共同体决定何时动用最强制性的制度回

应。好的刑法必须同时尊重两面：不能把一切罪名说成任意建构，也不能把现行罪名假装成自然界早就刻好的目录。

这一步对本书核心母题很重要。贝卡利亚已经明确看到“制度会参与形成行为条件”，说明他绝不是一个把经验主体完全冻结的人。他知道人会受到规则密度、处罚概率、社会意见和教育环境影响。于是“主体先在”的批判必须继续缩窄：真正需要问的，不是他是否知道人会被环境塑造，而是他是否把那个用来给国家授权的规范自由主体也放进这种发生之中。

> 带走的话
>
> 贝卡利亚看到，法律不是犯罪的照相机，也是犯罪边界的刻刀。禁止得越多，不一定意味着社会更安全，也可能只是把更多普通生活切进了“可罚”的格子。

小账

他看见了什么？——每写一条禁止都改写罪犯边界，规则过密逼出选择性执法，反毁确定性。

接缝在哪里？——法律制造的是犯罪身份而非伤害本身，两层不分，受害者的损害会被说没。

第 37 章

告密制度怎样制造一个人人互疑的社会

秘密告发是一种很奇怪的制度。它通常以“更容易发现犯罪”为理由出现：如果举报人不用露面，人们就更敢提供线索；如果国家可以秘密收集信息，危险分子就更难逃脱。只看侦查效率，这套制度似乎很有诱惑力。贝卡利亚却把镜头从“抓到了多少人”移到“这种制度把社会变成什么样”。（原典锚点：剑桥版第15章“秘密告发”。）

他观察到，秘密告发会让人变得隐蔽、保留，甚至彼此把邻居当作潜在告密者。一旦任何一句私下谈话都可能在不知道来源、无法面对指控者的情况下进入国家机器，人就不再只调整违法行为，而会调整整套交往方式：少说、猜疑、隐藏、表演忠诚。制度从法庭一路写回日常生活。

这比“秘密举报不公平”更深。它不是只说一个案件里的被告难以质证，而是说程序结构会生产一种人格和社会关系。一个人若长期生活在“别人可能随时匿名把我交给权力”的环境中，他对朋友、邻居、同事的信任方式会变化。表面上，国家只是多了一个信息渠道；实际上，整个社会的E——那片人与人纠缠的土——被改写了。

用公司例子看更直观。企业为了反腐设置匿名举报箱，本来有合理目的。如果制度同时具备证据核验、反报复、被举报人申辩、恶意举报责任等程序，匿名只是线索入口之一；但如果匿名信息可以直接触发处分，举报人永远不需说明依据，被举报者也不知道指控内容，组织很快会出现另一种行为：同事之间不再愿意承担风险，争议被转化为匿名攻击，管理层收到的信息越来越多，却越来越难知道什么可信。

因此，不能把贝卡利亚对秘密告发的批判机械移植成“现代匿名举报都不好”。今天的吹哨人保护、反腐举报、家庭暴力求助之所以有时需要保护身份，正是因为公开身份可能造成报复。真正要继承的不是一个绝对形式禁令，而是他的制

度问题意识：信息取得方式会不会破坏被指控者的程序保障？会不会让匿名从保护弱者变成无需负责的权力？会不会让社会信任的代价大到超过侦查收益？

这里也能看见贝卡利亚与前面“法官三段论”的内在一致。他一直试图把刑罚从不可见的个人意志中抽出来：法律应公开，处罚应预先确定，程序应让理由能够被检查。秘密告发则把一个决定性信息源重新藏进黑箱，使被告面对的不是一条看得见的公共规则，而是一个不知道从哪里来的指控。

更值得注意的是“恐惧对象”的变化。贝卡利亚后来谈预防犯罪时有一句重要思想：应当让人害怕法律，而不是害怕人。如果一个社会真正稳定，公民知道哪些行为违反规则，知道程序怎样启动，也知道权力受到什么限制；他不需要每天猜“某个官员是不是不喜欢我”“同事会不会偷偷告我”“这句话会不会被截成另一种意思”。当人开始害怕具体的人而不是公共规则，法治就被私人权力重新占据。

秘密制度还会制造一种信息悖论。管理者原本想得到更多真实信息，于是降低举报门槛；门槛降低后，信息数量激增，真假混杂；为了筛选，管理者又需要更大调查权；调查越广，更多人感到不安全；不安全又促使人隐藏和策略性表达；最后，国家得到更多“数据”，却可能离真实更远。数量不是信息质量，信息质量也不是正义。

从SDE看，这里是全书目前最强的一处“回写”证据。制度S一旦形成，并没有停在纸面，而是回写关系E；新的关系又改变行为序列D；行为变化以后，制度再看到一种新的社会，进一步调整规则。秘密告发不是一个静态条文，而是一台不断生产自己环境的机器。

这对“国家发生，主体先在”的母题构成直接压力。因为贝卡利亚已经清楚描述：人在制度中会变得谨慎、虚伪、保留，甚至把感受藏起来。换句话说，经验中的“自我”绝不是先在不变的。他至少已经看到，自我会被程序和权力关系塑形。

所以我们的批判必须再收窄一步。不能说“贝卡利亚没看见人会发生”。他看见了，而且看得相当尖锐。现在真正可能成立的批判只剩下：他在解释国家权力的

规范来源时，是否仍需要预设一个已经具有自由资格、能够让渡权利的主体；这个规范主体与他笔下会被制度塑造的经验主体之间，有没有被接起来？

这就是要求的“敌意近邻”：不要帮自己的判词找证据，要找最能推翻它的地方。秘密告发这一章，就是对“主体先在”最强的一次反证。如果后面教育、公共意见和预防章节继续证明他有更完整的主体发生论，我们必须继续改判。

> **带走的话**
>
> 贝卡利亚没有只批判秘密告发“程序不公”，他已经看见程序会塑造人格和社会关系。制度不是背景；制度会进入人的自我，制造一个互相猜疑或彼此信任的社会。

> **小账**
>
> 他看见了什么？——秘密告发让人少说、猜疑、表演忠诚，程序一路写回日常交往与人格。
>
> 还要追问什么？——今天的匿名举报能护弱者，关键看匿名信息是线索入口还是直接当证据。

第 38 章

一个人被叫成“罪犯”以后发生了什么

刑法最容易给人一种错觉：判决只是在确认一个已经存在的事实。甲实施犯罪，法院查明，判决把“罪犯”这个正确标签贴上去，事情就结束了。但只要把时间继续往后拉，标签本身会开始工作。一个人被公开称为罪犯以后，工作、家庭、信用、朋友、居住、投票、教育机会乃至他如何理解自己，都可能发生变化。判决不是终点，它会成为下一段生活的起点。

◆ 与标签理论对话：本书不能把“身份回写”当作第一次发现

“一个人被叫成罪犯以后会发生什么”并不是SDE才发现的问题。社会学中的标签理论早已把注意力放在社会反应如何参与偏差身份的形成：Edwin Lemert区分初级偏差与次级偏差，Howard S. Becker在《局外人》中进一步强调规则制定、贴标签与偏差身份之间的关系。因此，本书若只说“标签会回写身份”，并没有不可还原的创新。

本书保留“回写”概念，只因为它试图把标签之后的第二、第三轮制度效应一起纳入：判决形成身份标签，身份标签进入就业、教育、平台和警务数据，新的制度数据再反过来影响下一轮风险判断与处罚。如果这条多轮路径没有被具体证明，所谓“回写”就应降级为标签理论的通俗改写，而不是另立新说。

贝卡利亚当然没有现代“污名理论”的术语，可他对公共耻辱、放逐、没收财产的讨论已经触到这条线。他把公共耻辱理解为社会公开撤回信任的一种处罚，同时又警惕这种处罚不能滥用：如果太多行为、太多人都被宣布可耻，耻辱本身会贬值，法律与社会道德之间的联系也会断裂。（原典锚点：剑桥版第23章“耻辱”、第25章“放逐与没收”。）

这里最值得注意的不是他是否赞成某一种名誉刑，而是他承认处罚可以直接作用于“别人怎么看你”。也就是说，刑罚不仅处理身体和财产，还会重写社会关系。一个人的公民身份、名誉和被信任程度可以成为处罚对象。

放逐更明显。贝卡利亚说，一个被永久排除出共同体的人，在政治身体意义上仿佛“公民死了，人还活着”。这句话虽然属于十八世纪语境，却极其适合用来理解法律身份的第二次生命：生物学意义上的人仍在，可共同体承认他的方式已经变化。刑罚能够让一个人从“成员”变成“被排除者”。

现代社会很少用古典意义上的放逐，但相似结构并未消失。一个刑满释放者如果因为犯罪记录长期无法就业、租房、取得职业资格，法律上的刑期已经结束，社会上的处罚却可能继续。这里不能简单把一切后续限制都说成不正当，因为某些岗位确实需要风险审查；真正的问题是，这些限制是否和具体风险相关，是否有期限，是否允许更新判断，还是只要一次标签就永久决定一个人。

这正好让“回写”变得具体。第一次发生是行为：他做了某件事。第二次发生是法律判断：行为被归类为某种犯罪。第三次发生是身份回写：社会开始把“犯过某罪的人”当成解释他全部行为的总标签。第四次又会影响未来路径：求职受阻、关系断裂、合法机会减少，可能反过来改变他的选择空间。

如果我们只在第二步停下，就会以为刑罚是一个句号；把后两步看见，刑罚其实更像一个冒号。制度说完“有罪”之后，社会还会继续说很多话。法律越公开、标签越持久，这个冒号后面的内容越长。

贝卡利亚对名誉刑的限制给了一个重要提示：处罚要和社会真实评价结构相连。如果法律把本来并不被共同体视为严重可耻的行为强行套上极端耻辱，结果可能不是社会道德被法律改变，而是法律的评价本身失去信用。反过来，如果社会对严重侵害完全麻木，法律也不能只迎合现有舆论。两者之间不是简单服从，而是持续回写。

这里尤其适合区分“行为责任”与“本质身份”。说“你在某个时间实施了某项犯罪”，与说“你就是罪犯”看似只少了几个字，结构却完全不同。前一句把责

任绑在具体事实和时间上，后一句容易把整个个人压缩成一个身份。前者允许人承担、修复、变化；后者容易把未来冻结成过去的延长。

康德在上一编提醒我们必须认真对待责任主体；贝卡利亚在这一编则帮助我们看到，认真对待责任主体不等于让责任吞掉整个人。一个人应为行为承担法律后果，不等于共同体获得了永久定义他的权力。责任可以很重，但仍应有边界。

普通生活中也有类似结构。一个孩子撒过一次谎，父母如果说“这次你撒谎了”，是在处理行为；如果说“你就是骗子”，身份就开始回写。一次迟到被定义成“你这个人没有责任心”，一次考试失败被定义成“你不是学习的料”，都是把局部事件升级为本质标签。刑法只是把这种日常机制放大到国家级别。

因此，现代刑罚真正困难的一部分，不是把罪责变轻，而是让责任与身份保持区别。处罚结束后，一个人是否有制度化的返回路径？犯罪记录是否有分级、封存或消除机制？资格限制是否和具体风险有真实联系？这些问题都在问同一件事：法律的回写到底要持续多久。

这也为后面“教育为什么进入刑法书”埋下伏笔。如果人可以变化，那么制度不能只会贴标签；如果人完全不会变化，教育、矫正、重新融入都失去意义。贝卡利亚最终把预防犯罪推进到知识、制度与教育，本身就说明他没有把人理解成永恒固定的“罪犯类型”。

> 带走的话
>
> 判决确认的是一次责任，社会却容易把它扩成一个身份。贝卡利亚已经触到刑罚的第二生命：它不仅结束一个案件，还会重新安排一个人在共同体中的位置。

小账

他看见了什么？——公共耻辱与放逐会重写社会关系，被放逐者「公民死了，人还活着」。

接缝在哪里？——身份回写已被标签理论先占，多轮制度回写若不能具体证明，只算通俗改写。

第 39 章

刑罚如何反过来改变社会

到这一章，我们终于可以把前面散落的线索收成一个更大的判断：刑罚不是国家对社会做完之后就消失的动作。它会被看见、被记住、被模仿、被害怕、被绕开，也会改变人们对法律、权力、暴力与彼此的理解。换句话说，刑罚会回写社会。

◆ 与福柯划界：刑罚作为“表演”和“符号”不是本书的新发现

福柯《规训与惩罚》已经把公开处决、惩罚景观、身体与权力技术放进一条现代刑罚转型史，并特别讨论改革者如何设想一种更规则化、更符号化的处罚体系。第1章和本章把刑罚写成“演出”或“社会会读完再写回来的信”，必须承认这一解释史先占，不能把“刑罚会向社会传递意义”说成本书原创。

本书真正想多走的一步，是把“社会效果”与“处罚权限”持续分账：即使某种公开、可见、可理解的处罚确实产生了预防或象征效果，国家仍要另行证明自己取得了使用该手段的权限、手段没有越过必要与比例边界、制度为错误保留了撤回路径。这个“效果不能替权限签字”的区分，才是本书与纯粹传播效果分析之间要守住的界线。

贝卡利亚在许多地方已经看见这种回写。严刑如果经常不执行，会教会人们把法律当成虚张声势；温和但稳定的处罚，会建立更清楚的预期；秘密告发会制造猜疑；过多禁止会制造新的违法者；公共耻辱滥用会稀释耻辱；法官或行政官若比法律本身更令人恐惧，公民学会的不是守法，而是看人行事。

最后一条尤其重要。一个社会究竟是“怕法律”，还是“怕掌权的人”，外表都可能安静，内在却完全不同。怕法律，是因为规则公开、后果可预期，不论谁来执行都大体一样；怕某个人，则是因为结果取决于他的情绪、关系和权势。前者可能形成稳定合作，后者更容易形成迎合、躲避和投机。

这可以解释为什么同样是“处罚”，长期效果会完全不同。家长每次孩子犯错都按明确规则处理，孩子逐渐学会区分行为与后果；家长如果同样错误有时暴怒、有时纵容，孩子学会的不是规则，而是观察父母脸色。公司也是一样：制度稳定时，人会规划行为；领导个人喜怒决定奖惩时，人会规划关系。

刑罚还会回写社会对暴力的想象。国家如果把残酷公开展示成权威，公众可能同时学到两件相反的东西：一方面知道这种行为被禁止，另一方面又不断看到强者以暴力完成秩序。贝卡利亚反对残酷刑罚时并不只算犯人承受多少痛苦，也在担心这种表演对观看者造成什么长期教育。

但这里必须避免一个简单推论：国家使用强制，所以社会一定更暴力。合法强制也可能恰恰减少私人报复，让冲突从血仇转入公共程序。关键不是“有没有强制”，而是强制以什么形式出现：公开还是秘密，可预期还是任意，必要还是过度，可复查还是封顶。形式不同，回写方向也不同。

贝卡利亚的现代性正在这里显出来。他不是只把国家当作一只事后抓人的手，而开始把法律理解成塑造行为环境的装置。好的制度不是等所有人变坏以后再重罚，而是通过清楚规则、稳定执行、知识传播、公共奖励和教育，改变犯罪发生的条件。

如果用SDE三方程的大白话来写，这条线非常清楚。先有某种制度显露S，例如秘密告发、严厉死刑或温和而确定的处罚；它改变人的行为路径D，人开始隐藏、冒险、服从或重新计算；这些路径积累以后，又改变社会关系的土E，例如信任下降、规则信用上升、官员权威膨胀或公共意见转向；新的E再回头推动后文法律。

这意味着“刑罚效果”不能只看犯罪率一个数字。一个制度即使短期压低某类案件，也可能付出程序信任、社会合作和错误纠正能力的代价；另一个制度即使短期没有巨大震慑，却可能让人更相信规则。真正成熟的效果评价必须把这些长期回写纳入，而不是只数抓了多少人、判了多少年。

同样，刑罚也会改变国家自己。执法机关在某种制度下工作久了，会形成组织习惯。若法律不断给予宽泛裁量，组织会习惯依赖裁量；若证据标准严格，组织会发展更强的调查能力；若社会默认“抓到就算有罪”，侦查机关更容易把证明责任当成形式。国家不是固定机器，它也会被自己使用的工具重新塑造。

这一步对本书特别关键，因为它把“回写”从我们强加给贝卡利亚的外部术语，变成他文本中可以找到的内部线索。秘密告发、教育、确定性、公共意见、过度立法，彼此看似分散，实际上都在说明同一件事：制度会生产它后文要面对的人。

如果这个判断成立，那么贝卡利亚与发生学之间的距离就比最初想象得近。真正差的一步，不是“他完全没看见发生”，而是“他有没有把这些局部发生统一成对主体、权力与法律相互生成的完整理论”。这比一句“他缺回写”公平得多，也更难。

> 带走的话
>
> 刑罚不是一个落在社会上的点，而是一封会被社会读完再写回来的信。贝卡利亚已经多次看见这种回写，只是还没有把分散的回路收成一套统一的发生理论。

小账

他看见了什么？——严刑不执行、秘密告发、耻辱滥用都在回写社会，人学会怕人而非怕法。

接缝在哪里？——刑罚是符号已有福柯先占，本书要守的只是：效果不能替权限签字。

第 40 章

这是贝卡利亚最接近SDE的一刻

如果现在回头看前四十章，最初的母题已经被逼得发生变化。本书起笔时曾把贝卡利亚定位成一个把国家刑罚权“发生化”、却把自由个人留在发生之前的人。这个判断仍然可能成立，但它已经不能再用粗线条表达。因为从酷刑、秘密告发、处罚确定性、过度禁止、公共耻辱到教育，贝卡利亚反复证明他知道人会被制度塑造。

◆ SDE三方程在这里第一次完整落地

为了避免把SDE写成“显露—路径—背景”的线性三步，本章把三方程完整放回案件事实：S＝F(D,E)，案件中的可裁判事实显露，不只取决于证据路径D，也取决于讯问规则、身体处境、角色关系与制度激励构成的特征纠缠E；D＝G(S,E)，下一步查什么，又受当前已经显露的事实假说和制度条件牵引；E＝H(S,D)，新的判决、证据规则和办案经验还会回写制度环境，使后来的案件在不同条件下展开。

123原理在这里也可以被观察：第一步，现有证据路径与制度条件产生张力，例如证据不足却有强烈破案压力；第二步，张力逼出新的事实显露，可能是新证据，也可能是被强迫定型的口供；第三步，新显露回写路径和环境，决定下一轮继续搜什么、相信谁。一个健康程序的关键，是让第三步包含反向检验，而不是只允许确认性回写。

六路径在这里也有实际含义。拆解一个已经形成的案件判断，最自然的是S→D→E：先看判决显露成什么，再追查它沿什么证据路径形成，最后检查路径依赖的制度与关系纠缠；面对仍在调查中的程序风险，则可能从D或E起手。六路径不是六个答案，而是提醒研究者：起点不同，会暴露不同的盲区，不能把“在E中经D成S”误读成唯一分析顺序。

再做一次去术语测试：如果把SDE全部删掉，“现实事实不等于案件事实”“证据取得方式会影响可靠性”“程序要防止确认偏误”依然可以由证据法与认识论说清，这些不是SDE独有。SDE若有剩余价值，只在于把当前显露、后续路径和制度纠缠写成互相生成的三项，并把“结论回写后如何改变下一轮证据”列为同一模型内的必要变量。若这一变量不能带来新的可检验问题，就应老实降级。

所以，第40章先做一次正式收回：不再说“贝卡利亚把人看成不受环境影响的固定原子”。这句话已经被原典反证。经验中的人，在他那里会受痛苦、希望、概率、时间、习惯、意见、知识、制度和教育改变。一个会被环境推动的人，显然不是静止不动的。

但另一半问题还没有消失。贝卡利亚在说明刑罚权来源时，仍然需要一个规范叙事：独立的人为了让其余自由获得安全，牺牲一部分自由，形成主权。这里的“人”承担的不是经验心理模型，而是权利来源模型。他必须在国家之前已经拥有某种可以保留、可以让渡、可以被侵犯的自由资格，否则“国家只得到最小授权”的论证就失去抓手。

于是现在真正值得研究的不是“人到底先不先在”，而是贝卡利亚有两个“人”有没有接起来：一个是规范的人——拥有自由、能够授权、国家必须尊重；另一个是发生中的人——被制度、教育、意见和处罚改变。前者给国家画边界，后者解释刑罚怎样起作用。两个人常常并排出现，却没有被彻底写成同一个动态主体。

这可能就是全书更精确的接缝。不是“主体先在”四个字直接盖章，而是：规范主体与发生主体之间的接口没有被充分展开。　当贝卡利亚要限制国家时，他需要一个不被国家制造的自由资格；当他要解释犯罪与预防时，他又需要一个会被社会制造和改变的经验自我。两者如何统一，是他留下给后人的难题。

这个难题其实非常现代。今天的人权理论也会遇到类似问题：我们说人的尊严和基本权利不能由国家授予，因为如果国家可以授予，也可以随时收回；但社会科学又不断告诉我们，人的能力、偏好、身份和选择条件深受家庭、教育、经济和制

度影响。怎样既承认主体发生，又不把人的基本地位交给环境决定，是现代规范理论的核心张力之一。

因此，SDE如果只说“一切都是发生的”，反而可能输给贝卡利亚。因为如果人的权利资格也完全由现实发生决定，那么一个把某些人长期塑造成低地位的社会，是否就能说“他们的低权利也是这样发生出来的”？显然不能。发生事实不能自动推出规范正当。这里恰恰要求SDE说明：发生解释与权利底线是什么关系。

这一反击非常重要。它让本书不能站在贝卡利亚头顶宣布胜利。贝卡利亚可能在告诉我们：为了让权力受到限制，有些规范地位必须被暂时固定，不能每次都交给现实力量重新生成。自由主体也许不是一个经验事实，而是一条制度必须“当作如此”对待人的规范承诺。若是这样，“主体先在”未必是错误，可能是一种保护性结构。

但贝卡利亚也不能因此获得免检。规范固定为什么选“自由”，而不是别的价值？自由与公共幸福冲突时谁优先？一个因贫困、教育、成瘾、精神状态、社会压迫而严重改变选择能力的人，刑法仍以同样方式理解他的自由和责任吗？规范主体如果完全不接受经验主体的回写，也会变成另一种看不见差异的抽象人。

所以，第40章把原来的总判断改写成一个更难、也更有资格支撑全书的公式：贝卡利亚的伟大，在于让国家权力接受发生解释；他的未竟之处，在于没有完全说明规范自由与经验发生如何在同一个人身上互相校正。

这句话同时保住了他的贡献和我们的批判。没有前半句，贝卡利亚会被写成一个旧时代的功利主义者；没有后半句，SDE又只能给他鼓掌。真正的解构应当让两边都付出代价：他的理论必须回答主体接口问题，SDE也必须回答“发生是否会吞掉权利底线”的问题。

现在再看“自由中心还是幸福中心”，比分也更复杂了。刑罚权来源与边界仍然明显由自由主导；但预防犯罪、法律目的和教育章节越来越把公共幸福推到中心位置。自由可能不是全书唯一中心，而更像一号闸门：任何刑罚先不能越过人的最低自由地位；进入闸门以后，制度再以减少总体不幸、提高安全为优化方向。

因此，后文必须把“预防犯罪—科学与知识—公共官员—教育—公共幸福”整组读完，不能现在就宣布自由中心获胜。尤其贝卡利亚自己把良好立法的目标与最大幸福、最小不幸连接起来，这可能构成对“自由中心”最强的一次内部挑战。如果幸福才是最终目标，自由也许只是实现幸福的条件；如果自由是不可被幸福交换的边界，那么中心判断才真正成立。

这就是为什么本书不急着关门。所谓“看见—收回”，真正的价值不在于每本书一定抓到哲学家的致命漏洞，而在于允许原典把我们的漏洞也抓出来。从前文撤下“固定量刑尺”，到本编收窄“主体先在”，母题不是越来越响，而是越来越难被推翻。

贝卡利亚最接近SDE的一刻，恰恰不是他谈抽象契约的时候，而是他发现制度会反过来塑造人：秘密告发制造猜疑，过度禁止制造犯罪，严厉与不确定共同制造侥幸，教育改变行为条件，法律的示范改变公共意见。这些线索已经组成了一张发生网络。

可他最后仍然需要某些东西站稳，否则改革没有支点：自由、法律、公共安全、人的责任资格。SDE下一步不能只是把这些支点全部拆掉，而要回答一个更困难的问题——怎样让支点既能站住，又不假装自己从来没有来历；怎样让权利能够约束发生，又允许发生不断修正我们对权利实现方式的理解。

◆ 理论深化　从“死刑对撞”到“主体接口”

一、康德不是贝卡利亚的反面，而是另一种限权

把康德写成“残酷报应论者”会让对撞失真。他对刑罚的限制来自责任与人格：不能先为了社会利益需要一个处罚，再去寻找承担它的人。贝卡利亚限制国家多伤害，康德限制国家把人纯粹工具化。二者都在给国家上锁，只是锁装在不同位置。

二、“国家发生，主体先在”第一次被实质改判

本编证据已经不足以支持“贝卡利亚完全把人当固定原子”的宽泛版本。秘密告发、过度禁止、公共耻辱和制度示范都证明，他能看到经验主体受制度塑形。今后只能保留更窄的命题：作为国家权力来源的规范自由主体，与作为行为机制的发生主体之间，接口没有被充分理论化。

三、回写不再只是SDE加上去的词

前文谈回写还有较强的外部分析色彩；本编则可以在贝卡利亚文本自身找到连续证据：制度改变意见，意见改变行为；规则密度改变违法数量；秘密程序改变信任；惩罚方式改变法律信用。这让“回写”成为对原典线索的统一，而不是随意套模。

四、自由中心与幸福中心进入真正决赛

到第40章，自由仍主导“权力是否有资格进入”这一入口，幸福则越来越主导“进入以后怎样优化制度”。如果后文预防与教育章节显示贝卡利亚把自由最终都还原为最大幸福的工具，那么“自由中心”需要改判；如果自由始终构成不可交换的边界，则自由中心才有资格成为正式母题。

五、给SDE新增一个反身问题：发生不能直接推出正当

如果某种不平等、偏见或压迫确实是历史发生出来的，发生学只能解释它怎么来，不能因此说它就是正当。贝卡利亚的规范自由主体迫使SDE承认：发生解释与规范判断不是一回事。未来“给SDE自己写死法”时，这将成为必须正面处理的一条。

六、死刑争论里其实有三种“比例”

我们经常把“罪刑相称”说成一个词，实际上至少有三种不同的比例。第一种是责任比例：行为越严重、主观责任越强，法律回应越重。康德最关心这一层，因为他怕刑罚脱离行为人的罪责，变成纯粹管理工具。第二种是手段比例：即使一个人确实应受处罚，也要比较不同刑罚是否超过实现正当目的所必要的程度。贝卡利亚的锋芒主要在这里。第三种是制度比例：某种刑罚给整个社会带来的错误风险、程序成本、示范效应和长期回写，是否与它所追求的目标相称。

三种比例一旦分开，很多争论会清楚。有人说“故意杀人极其严重，所以死刑相称”，主要在说责任比例；另一人说“终身监禁已经足以防止再次伤害，所以再处死不必要”，主要在说手段比例；还有人指出“任何司法系统都可能出现错判，死刑一旦执行无法纠正”，则在说制度比例。三个人并不一定直接回答了对方。真正完整的刑罚论，必须说明三种比例怎样排列优先顺序。

贝卡利亚的优势，是很早就把第二、第三种比例带进刑法思考：刑罚不应超过必要效果，制度的残酷和不确定会产生新的社会后果。康德的优势，是把第一种比例守得非常硬：不能因为一种处罚对社会更方便，就跳过罪责与应得。二者结合以后，我们才能看见一个更成熟的问题——刑罚既不能超出责任，也不能因为责任成立就自动免除必要性审查。

这也给普通制度治理一个重要启示。一个员工确实犯错，不等于公司可以采取任何有效措施；一个学生确实作弊，不等于学校所有重罚都天然正当；一个平台账号确实违规，不等于永久封禁自动和责任相称。先问“他应承担多少”，再问“哪种手段达到目的”，最后问“这种制度长期会制造什么”。三道比例一起用，处罚才不容易从责任滑向权力习惯。

如果本书后面要提出自己的重建，我建议把这三种比例保留下来：责任比例守住个人，手段比例限制国家，制度比例看见回写。任何一项缺失，刑罚都可能失衡。只讲责任，会忽略制度错误与替代方案；只讲必要，会把人变成治理对象；只

讲制度效果，又可能让具体受害与罪责在宏观数据中消失。贝卡利亚与康德的真正相遇，不是必须选一个，而是逼我们承认刑罚同时生活在三个尺度里。

普通人的三个现场　贝卡利亚这一部分怎样回到生活

场景一：匿名投诉要不要允许

不能用一句“匿名不公平”或“保护举报人最重要”解决。更好的结构是：允许匿名提供线索，但处分必须建立在独立核实的证据上；保护举报者免受报复，同时给被举报者最低限度的知情、申辩和复查。保护线索来源不等于让指控自动成为事实。

场景二：一次永久黑名单

永久取消资格是一种强回写。越接近永久，就越应说明风险是否真的长期存在、有没有重新评估节点、能不能通过后续行为恢复资格。没有退出机制的标签，很容易把一次责任变成终身身份。

场景三：一项处罚结束以后，事情真的结束了吗

看看它留下什么：当事人是否还能重新进入正常生活，同事是否更信任制度，旁观者学会的是规则还是恐惧，执行者是否形成更专业的习惯。处罚的真正效果常常在判决以后才开始显露。

◆ 本编理论加厚　六个不能跳过的反问

一、报应是不是复仇换了一件法袍

听到“应得”两个字，普通人很容易把康德式报应理解成制度化复仇：你让我痛，所以国家也让你痛。这样理解太简单。私人复仇的核心是受害者或其亲属凭自己的愤怒追回伤害，报应论则试图把回应从私人情绪中抽出来，交给一般、公开、对所有人适用的法律。二者都向过去看，却不等于一回事。真正需要审查的是：国

家所谓“相称回应”有没有独立于愤怒的公共尺度，还是只是把社会的报复冲动翻译成法律语言。

康德最强的地方，正是在这里拒绝把刑罚建立在“大家看到以后会更安全”上。他提醒我们，一个人不是可以被随便放进社会机器里的齿轮。可报应论若要进一步说“某种罪只有某种刑罚才能满足正义”，就必须提供新的论证。责任成立，只能推出“可以处罚”或“应当有回应”，未必自动推出回应的具体形态。贝卡利亚的必要性追问因此仍然有效：如果有两种方式都能严肃表达责任，为什么必须选更不可逆的一种？

这一点也说明，报应与预防未必只能二选一。制度可以把罪责作为进入刑罚的门槛，防止为了威慑而处罚无辜或过度处罚；进入门槛后，再比较各种合法刑罚的预防、修复与社会成本。真正困难的是两者谁拥有最后否决权。若预防效果极差，报应是否仍要求执行；若预防效果极好，报应是否允许超出应得。把两个问题分开，争论才不会混成一句口号。

二、效用会不会把少数人的自由换成多数人的安心

贝卡利亚把刑罚拉向未来，最大的理论风险也随之出现：如果制度只看总体结果，会不会出现“多牺牲一个人，让一万人更安心”的计算？这正是康德的警报。任何以公共幸福为理由的制度，都必须回答一个边界问题——有没有某些东西不能因为总效果更好就拿去交换。

贝卡利亚并非毫无防线。他从自由让渡、法律预先规定、必要性和罪刑比例出发，已经给效用加了几道闸门。国家不能为了任何想象中的好处随意处罚；只有法律允许、与犯罪相称、对公共安全必要的强制才有资格进入。因此，把他写成“多数人幸福足够大就什么都能做”的粗糙功利主义并不准确。

但问题仍在：当公共安全与个体自由真正冲突时，哪一项是不能被计算掉的？例如，为了显著提高破案率，可否大幅降低证明标准；为了降低再犯风险，可否把刑罚结束后的高风险者无限期控制；为了防止极小概率灾难，可否长期限制大量无辜者？这些问题都不是十八世纪原典能直接回答，却能够检验贝卡利亚结构里自由

究竟是边界，还是幸福计算中的一个重要变量。后文“自由中心对幸福中心”的决赛，必须靠这类冲突题，而不是靠数哪个词出现得更多。

三、说“法律制造犯罪”，会不会把真正的伤害说没了

“法律不仅抓犯罪，也制造犯罪”是一句很有冲击力的话，也最容易被误读。若理解成“没有法律就没有任何犯罪事实”，受害者的真实痛苦会被概念游戏抹去。一个人遭受暴力、财产被夺、身体受伤，这些发生并不需要先等法典盖章。SDE在这里必须守住事实：发生过的伤害具有现实性，法律不能把它说成不存在。

法律制造的主要是第二层东西：哪些伤害进入刑罚，怎样命名，严重性如何排序，谁获得“犯罪人”身份，国家可以使用什么强制。正因为第二层是制度形成的，它才必须不断接受公共理由审查。把两层分开，既能避免法律实证主义的“写进法条就天然正当”，也能避免极端建构主义的“所有犯罪都只是话语”。

贝卡利亚本人的“社会损害”正好提供中间桥梁。他不满足于“法律说是罪所以是罪”，而要求犯罪化与共同生活受到的损害相连；但他也知道损害排序不能达到几何学精确。这种不完全确定反而是一种诚实：刑法需要稳定边界，却永远不能把自己的分类误认成自然界唯一可能的分类。

四、匿名举报与秘密告发，能不能简单画等号

这一问题如果不处理，现代读者很容易拿贝卡利亚去反对一切匿名举报。那会把历史判断变成制度误用。贝卡利亚批判的核心，是秘密信息在缺乏公开对质和责任约束时直接进入刑罚机器，使被告面对无法识别、无法检验的权力来源，并让整个社会进入相互猜疑。现代匿名举报则可能承担保护弱者、反报复和发现隐蔽违法的功能。

所以关键不是“匿名”两个字，而是匿名信息在程序中处于什么位置。若它只是启动核查的线索，之后必须靠独立证据完成证明，匿名可以和程序正义共存；若它本身就能承担定罪、处分或公开污名的证明功能，贝卡利亚的警告重新出现。制度必须把“保护举报者”与“降低证明责任”严格分开。

这一细分本身就是发生学写法：同一个工具放进不同的程序土壤，会长出不同结果。匿名不是天生善也不是天生恶，真正决定它性质的是证据链、申辩权、复核机制、反报复安排以及恶意举报的成本。只按一个词做价值判断，恰好会退回本书一直批判的“发现式分类”。

五、刑满以后还要不要让社会记住

第38章说标签会回写身份，但“去污名”也不能被理解成要求社会彻底失忆。有些过去行为与未来岗位存在真实风险联系：曾严重侵害特定弱势群体的人进入高度接触同类群体的岗位，制度需要审查；金融领域重大诚信犯罪与某些受托资格也可能存在关联。问题不在于能不能记，而在于记忆是否有比例、期限和具体关联。

一个成熟制度至少应区分三件事：第一，罪责本身是否已经通过刑罚完成；第二，某项后续限制是否针对一个仍然存在、能够说明的具体风险；第三，这个风险是否允许随时间和后续行为重新评估。如果三者不分，刑罚就会从有限责任变成无期限身份。

这里也可以看出“回写”为什么不能只被理解成善意的反馈。回写可能使制度更好，也可能把一个人越写越窄。一次犯罪记录进入就业、信用、社交和家庭，多个系统相互读取，最后形成一个几乎无法退出的闭环。法律若不设计退出路径，社会就可能用“预防风险”的名义制造长期排除，而长期排除又反过来增加重新融入的困难。

六、规范自由主体与发生主体，到底能不能合成一个人

这是本编最重要的理论问题。若把“人是发生出来的”理解成经验描述，很容易成立：语言、家庭、教育、经济条件、制度奖惩都参与形成我们的选择方式。但规范法治又必须说：即使一个人生活在最不利环境里，他也不能因此被国家当成低等人；即使多数人长期歧视某群体，那种“发生出来的共识”也不能自动决定其权利。

因此，一个理论若只剩发生，会失去对既成现实说“不”的位置；一个理论若只剩先在规范，又容易看不见真实的人怎样在不同条件中形成。贝卡利亚的两个主体其实可能不是缺陷，而是现代法治必须同时保留的两种语言：用规范主体守住不可被统治者随意取消的地位，用经验主体理解怎样让真实的人能够行动、承担责任并重新进入社会。

真正缺少的是两种语言之间的接口。例如，贫困和教育不足能够解释某些行为概率，却不能因此取消责任；严重精神障碍可能改变责任能力，却不能降低人格地位；社会环境能够塑造偏好，却不能让国家据此宣布某些人“不配自由”。接口理论要回答：哪些发生条件影响责任程度，哪些不能影响基本权利；哪些差异要求制度提供补偿，哪些差异仍由个人承担。

如果这条线在后续能够成立，那么全书最后的SDE建构就不会是“把先在主体拆掉”，而是更谨慎的重建：把自由资格作为规范底线保留，同时把主体实际行使自由、形成选择、承担责任的路径发生化。这样既不让现实决定权利，也不让抽象权利假装所有人处在同一片土里。

◆ SDE对账

第一，本编新增的“看见”至少有六处：贝卡利亚看见刑罚正当性不能只有未来效果；通过与康德对撞，我们更清楚地看到他自己的效用论边界；他看见过度禁止会制造新的犯罪类别；看见秘密告发会制造猜疑人格；看见公共耻辱会随滥用而贬值；看见制度执行会回写公众对法律和权力的感受。

第二，“收回”候选发生重要变化。“先放人，再限权”仍可保留为通俗提示，但已不能直接等同于“主体完全先在”。更精确的候选动作应改为“分两个人”：规范论证里先放一个拥有自由资格的人，经验分析里又出现一个被制度塑造的人，而二者之间缺少完整接口。这可能比原来的动作更具体，也更难被原典推翻。

第三，死刑编建立了一个双向保底。对贝卡利亚，不能只用“死刑无效”做全部论证，因为康德提醒责任与应得不能被效果吞掉；对康德，也不能只用“罪责应

得”封顶，因为国家权力来源、替代方案、错误风险和不可逆性仍然需要解释。两家最强处都保留，才有资格继续SDE解构。

第四，本编第一次明确区分“发生解释”与“规范正当”。一项制度、身份、偏见可以有清楚的发生史，但发生过不等于应当继续存在；反过来，一条规范为了保护人而暂时固定，也不等于它拥有超历史真理。SDE若不能处理这一区分，就会把“怎么来的”误写成“该不该”的答案。

第五，当前全书中心诊断暂不封顶。自由仍是刑罚权的入口变量，幸福正在成为制度目标变量，创造则在“法律如何制造犯罪类别与社会关系”中显影。后文必须让预防、知识、官员、教育与公共幸福全部进入九格，再决定自由中心是否成立。

> 带走的话
>
> 本书正式收回“贝卡利亚看不见主体发生”的过强判断。更准确的接缝是：他同时拥有一个会被制度塑造的经验主体和一个用来限制国家的规范自由主体，却没有充分写出二者如何互相回写。

> **小账**
>
> 这一章看见了什么？——他笔下有两个人：替国家画边界的规范自由人，被制度塑造的经验人。
>
> 接缝在哪里？——二者怎样在同一人身上互相校正没写出；发生事实也推不出规范正当。

第 八 编

刑法最好的时候，是它不用出现

编 序

本编五章。第41章讲预防为何比惩罚更重要；第42章讲教育为何进入一本刑法书；第43章讲公共幸福究竟指什么；第44章让自由中心与幸福中心正面对决；第45章检查幸福若是根，前面的判词会不会倒。

本编是判决二的主场。刑罚不再是事后发泄，而是公共幸福工程里受必要性和比例约束的一件工具；最好的刑法，是让犯罪少发生、刑罚不必出现。判决四所用的锁钥问题，也在本编第一次校准。

读这一编，可以带着日常问题进去：为了让孩子少犯错，是天天查手机，还是讲清规则再让他自己判断，哪一种才算预防？说「为了大家好」的人，又凭什么替大家定义好？

第 41 章

为什么预防比惩罚更重要

如果一个城市每年都要抓很多纵火犯，最容易想到的办法是把刑罚加重：判得更久、罚得更重、公开得更响。贝卡利亚在《论犯罪与刑罚》快要结束时，却把问题整个翻了过来。他说，与其把主要聪明用在犯罪已经发生以后怎样处罚，不如先问怎样让犯罪少发生。这个转身看起来平常，实际上改变了刑法的时间方向。传统刑罚的目光盯着过去：谁做了什么、该还多少；预防的目光则转向未来：什么条件会让类似行为再次发生，我们能不能在不扩大任意权力的前提下提前改变这些条件。（原典锚点：剑桥版第41章“如何预防犯罪”。）

这不是一句‘预防为主’的口号。贝卡利亚后面列出的办法很具体：法律要清楚、简单、可知；法律不能替某个等级服务而应保护每个公民；人们应该害怕的是法律本身，而不是某个具体权势人物；知识应当扩散；官员的利益结构不能诱导他们背离法律；社会不应只奖励告密和抓捕，也要奖励公开的善行；最后，教育要使人更早形成判断与习惯。这些措施放在一起，已经不像一份传统的刑罚清单，而像一张犯罪发生条件图。

用SDE的大白话说，刑罚处理的是已经形成并可辨认的“显露”。盗窃已经发生、伤害已经发生、判决已经作出，制度开始给既成行为分类、定责、回应。预防则把镜头往前推：一件事沿什么差异序列长出来，处在什么特征纠缠里，哪些刺激、规则、机会、知识和关系不断给它加力。只盯着显露，制度会变成消防队；开始看差异序列和特征纠缠，制度才可能问为什么总在同一个地方起火。

但预防最危险的地方也从这里出现。处罚至少面对一个已经发生的行为，预防却面对一个还没有发生的未来。如果国家只要说‘为了防止犯罪’，就可以无限监控、提前限制、预测谁可能危险，那么预防会比惩罚更容易扩张。一个人因为做了坏事被处罚，与一个人因为‘可能做坏事’而被提前控制，不是同一件事。后者稍不谨慎，就会把自由从行为边界改造成风险评分。

所以，贝卡利亚式预防不能和现代意义上的‘事前控制一切’画等号。他自己给出的主要办法不是把每个人都当潜在罪犯盯住，而是改善公开制度：让法律更清楚，让特权更少，让知识更充分，让行政者守法，让奖惩结构更一致，让教育更有效。换句话说，他首先想改变的是公共环境，而不是给每个个人装上更长的锁链。

这一点非常重要。真正的预防不是把刑罚提前，而是让刑罚尽量不必出现。家长若说‘为了防止孩子犯错，所以每天检查手机、限制朋友、替他做全部决定’，这是把处罚逻辑提前；学校若通过清楚规则、稳定反馈、真实教育，让学生逐渐知道为什么某些行为会伤害别人，这才更接近改变发生条件。两者都叫预防，实际却是两条完全不同的路。

贝卡利亚的温和刑罚论也因此得到新的解释。他不是单纯认为‘轻一点更仁慈’，而是认为残酷刑罚往往是一种低水平治理的补丁：前端制度没有让规则可知、执行稳定、社会关系可信，后端只好靠剧烈痛苦制造震动。越依赖这种震动，越容易形成一种奇怪循环——法律需要不断加重，才能维持同样的恐惧；恐惧越大，人们越把法律理解成敌人；法律越像敌人，守法就越依赖外部强制。

从这个角度看，‘预防比惩罚好’不是少盖几座监狱这么简单，而是在改变国家治理的发动机。处罚是对已经出现的结果施力，预防则试图重新安排结果发生之前的路径。问题也随之升级：谁有权设计这些路径？什么叫好的环境？国家为了让人‘变好’可以走多远？贝卡利亚马上会把我们带到知识、官员、奖励和教育，也正是在那里，幸福中心对自由中心的攻击真正开始。

> **带走的话**
>
> 贝卡利亚最深的刑罚改革不是把刑罚削轻，而是把国家的聪明从‘怎样更有效地惩罚’往前推成‘怎样让犯罪更少发生’。但预防只有在不把未来风险直接当成现实罪责时，才仍然属于自由的朋友。

> **小账**
>
> 他看见了什么？——他的预防清单是一张犯罪发生条件图：法律清楚、反特权、知识、官员、奖励、教育。
>
> 接缝在哪里？——预防面对尚未发生的未来，若把风险当罪责，它会比惩罚更容易扩张。

第 42 章

教育为什么会进入一本刑法书

一本讨论犯罪和刑罚的书，最后突然谈教育，初看很像作者跑题。可贝卡利亚恰恰把教育称为最可靠、同时也最困难的预防犯罪办法之一。他甚至承认这个问题太大，已经超出自己这本小书的范围，而且与政府性质紧密相连。正因为他说自己写不下去，这几页才格外值得看：它们暴露了刑罚理论背后真正的人之模型。（原典锚点：剑桥版第45章“教育”。）

如果人只是一个固定的利益计算器，教育其实没有那么重要。国家只需要把犯罪收益和处罚成本摆清楚，人自己算账就行。可贝卡利亚谈教育时，显然知道人的感受、判断和习惯并不是出生时全部完成。他强调，教育不能只靠命令换取表面的服从，而要通过经验、感受和后果，使人逐渐形成稳定的趋向。这里已经有一个很强的发生性洞察：做选择的那个人，本身也在被生活塑造。

这对我们前文的核心判词构成直接反证。我们曾说，贝卡利亚为了限制国家，先放进一个已经拥有自由的个人，然后由这些个人交出最少量自由形成公共权力。作为规范论证，这个结构仍然成立；但教育章节告诉我们，他并没有把经验中的人完全看成现成成品。一个孩子怎样理解规则、怎样感受他人的痛苦、怎样把短期冲动和长期后果连起来，都要经历学习。‘会选择的人’不是自动从自然状态中跳出来的。

于是必须区分两层自由。第一层是资格自由：无论一个人受过什么教育、处在什么阶层、聪明还是迟钝，他都不应因此失去作为法律主体的基本地位。这一层自由主要用来限制国家。第二层是能力自由：一个人能不能看见选项、理解后果、控制冲动、形成长期计划、在压力中维持判断。这一层自由显然会受到家庭、教育、贫困、知识和制度经验的影响。贝卡利亚没有用今天的术语做这种区分，但他的文本已经把两层同时摆在桌上。

教育因此不是刑法之外的慈善事业，而是刑法为何可能少用的一项基础设施。如果社会只在一个人犯罪以后才第一次认真和他讲规则，说明规则进入得太晚。真正有效的预防，不是从幼儿园开始背刑法条文，而是让人在很多小场景里逐渐形成一种经验：规则有理由，别人有边界，自己的行为会带来后果，权威也受规则约束，犯错以后有修复而不是只有羞辱。这样的教育不是把孩子训练成怕罚的人，而是在形成能够和规则共同生活的人。

这里还有一个很容易被忽略的细节。贝卡利亚反对只靠命令获得短暂服从。命令最方便，因为成年人只需要说‘不许’，孩子马上停下来；但这类服从往往只在监督者在场时有效。真正形成的判断则不同：即使老师、父母、警察不在旁边，一个人也能把行为与他人、后果和长期生活连接起来。前者靠外部强制，后者把规则的一部分变成自我组织。

这正好帮助我们重新理解刑罚的确定性。确定性并不等于到处有人盯着。更成熟的确定性，是社会成员知道规则通常会被一致执行，违法不是靠关系逃脱，守法也不是因为某个权威的脸色。教育如果把这种经验从小写进人的判断方式，刑罚就不需要不断提高音量。相反，一个孩子从小看到的是‘有关系就没事、没关系才守规矩’，再好的法条也会在现实中被另一套课程抵消。

但教育同样有危险。只要国家宣布‘为了减少犯罪，我们要教育出正确的人’，教育就可能变成思想整形。哪些价值可以教？哪些生活方式只是不同而不是犯罪风险？国家能不能为了所谓公共幸福，把不喜欢的性格、文化或表达都说成需要矫正？贝卡利亚把教育与政府性质联系起来，其实已经让这个危险露了头：教育从来不只是传知识，也会决定什么样的人被称为好公民。

因此，教育进入刑法的正确方式，不是让刑法吞掉教育，而是让刑法承认自己的边界。越能通过平等教育、基本保障、清楚规则和可信制度减少犯罪条件，越没有理由把所有社会问题都交给刑罚。一个社会若一边削弱教育和公共支持，一边不断扩大刑罚，等于先把土弄坏，再惩罚从坏土里长出的结果。贝卡利亚没有把这句话写出来，但他的预防路线已经给出方向。

带走的话

教育章节迫使我们第一次正式收回‘经验主体完全先在’的说法。贝卡利亚知道，能够理解规则、感受后果并作选择的人也要被培养出来；真正未完成的不是主体有没有发生，而是规范上的自由资格怎样与经验中的形成条件接上。

小账

他看见了什么？——会选择的人要被培养，教育靠经验与后果形成判断，不靠命令换表面服从。

还要追问什么？——国家若以减少犯罪为名教育出正确的人，教育便可能滑成思想整形。

第 43 章

公共幸福究竟是什么

现在必须正面处理一个一直在全书边缘发光、却还没有被我们放到中央的词：幸福。贝卡利亚在全书开头批评旧法律常常只是少数人的激情工具或偶然需要的产物，并把良好立法指向一个著名尺度——让更广泛的人共享更大的幸福。后来的功利主义会把‘最大幸福’变成极有影响力的道德公式，但如果我们直接把成熟的边沁倒装回1764年的贝卡利亚，就会遮住他自己的结构。

贝卡利亚说的幸福首先不是快乐总量排行榜。它在《论犯罪与刑罚》中反复以几种更具体的样子出现：人能安全生活，不必害怕任意权力；法律对不同阶层采取共同尺度；处罚只到必要为止；社会少一些无用痛苦；人的劳动、交往和知识不被不必要的禁令压住；政府能用稳定规则而不是秘密恐惧维持秩序。把这些内容合起来，公共幸福更像一种共同生活的质量，而不是把每个人的笑脸相加。

这一点首先能解释为什么贝卡利亚如此重视安全。对他而言，人们进入社会并接受共同规则，不是为了把自己交给更强的人，而是为了摆脱彼此随时侵害的脆弱状态。安全因此不是幸福之外的附属品，而是公共幸福最基础的地板。没有安全，财产、行动、关系和计划都无法稳定；可如果安全被解释成国家无限强制，幸福又会被自己吞掉。所以他才不断用‘最少必要’给安全政治上锁。

第二，公共幸福具有分配含义。贝卡利亚不是只问国家总体看起来强不强，而是警惕利益和特权集中到少数人手中。法律如果让一个阶层享受豁免、让另一个阶层承担主要痛苦，即使社会总财富增加，也很难符合他想象的良法。这里的‘更多人共享’很重要：公共幸福不是统治者宣布‘国家整体更好’就成立，它必须落到普通成员能够实际享有的安全和自由。

第三，公共幸福具有反浪费含义。刑罚制造痛苦，因此每一份痛苦都应说明用途。如果轻刑已经足以防止犯罪，再增加的痛苦就不是社会必要成本，而是权力的浪费，甚至暴政。这个思路与现代政策评估很接近：不是问一个措施能不能产生一

点好处，而是问同样目标能不能用更低的人身代价实现。也正因为如此，贝卡利亚才会反复讨论确定性、及时性和温和——他想提高的不是痛苦总量，而是制度效果与强制成本之间的比例。

第四，幸福并不只是刑罚之后的结果，它开始反过来决定刑罚以前应该怎样安排制度。知识扩散能减少迷信和任意，官员结构能降低滥权，奖励可以让社会不只会处罚，教育可以形成更稳定的自我约束。这些内容说明，贝卡利亚真正关心的是一整套公共生活怎样减少犯罪和恐惧。刑罚只是其中最尖锐的一件工具，而不是全部治理。

可‘公共幸福’也有最危险的一面：谁来替公众定义幸福？如果答案变成一个自认为冷静、理性、懂得整体利益的立法者，那么旧时代的君主任意可能换一件衣服回来。过去统治者说‘因为我是君主，所以我知道该怎样罚你’；新的立法者若说‘因为我代表公共幸福，所以我知道什么对你最好’，权力仍可能越过个人。贝卡利亚的理论必须证明，公共幸福不是一张新的无限授权书。

这就是自由重新进入的地方。每个人只让出最少量自由、处罚必须由法律预先规定、社会也受契约约束、等级不能获得刑罚豁免——这些命题共同发挥一种功能：它们让公共幸福不能随意吞掉具体个人。可以说，幸福告诉制度往哪里走，自由告诉制度有些路不能抄近道。二者并不是平行的两句漂亮话，而是目标和边界的关系。

如果这样理解，之前‘自由中心’的判断就必须接受一次真正的危机。一个中心不仅要在重要句子里出现，还要能解释整套制度为什么朝某个方向运动。自由可以解释为什么限制刑罚、为什么罪刑法定、为什么反对死刑，却不太容易单独解释为什么要扩散知识、奖励美德、改革教育、设计官员激励。公共幸福反而可以把这些分散章节串起来：它们都在问怎样让共同生活少靠恐惧、多靠可持续的制度条件。

但这还不是最终改判。因为一个目标能够解释更多章节，不等于它在规范上拥有最高地位。真正的测试必须是冲突题：如果牺牲一个无辜者能让多数人更安心，

贝卡利亚是否允许？如果秘密监控显著降低犯罪，他是否接受？如果酷刑能偶尔提高破案率，公共幸福是否足以正当化？他在许多地方的回答其实趋向否定。这说明自由和程序并不是随时可以被效用兑换的普通成本。

所以，本章暂时给‘公共幸福’一个更准确的定义：它不是多数人的即时满意，也不是国家统计表上的总收益，而是一种分布到成员身上的、以安全和可预期自由为重要内容的共同生活状态。它确实可能是贝卡利亚的方向盘，但方向盘前面还有限权闸门。下一章要做的，就是把两个候选中心放到同一张桌上，谁也不准靠口号取胜。

> **带走的话**
>
> 贝卡利亚的幸福不是‘大家高兴就行’，而是减少无用痛苦、扩大共同安全、削弱特权与任意的公共生活质量。它比自由覆盖的章节更广，却必须接受自由和法定程序的持续约束。

小账

这一章看见了什么？——公共幸福是共同生活质量：安全是地板，另有分配与反浪费两层含义。

还要追问什么？——谁替公众定义幸福？自称代表公共幸福的立法者，也可能是新的任意。

第 44 章

贝卡利亚到底是自由中心，还是幸福中心

现在进入本编最重要的一场审判。前面几万字一直把‘自由中心主义’作为第一候选：个人先有自由，只交出维持共同安全所必需的最小部分，国家因此只能使用有限刑罚。这个结构非常强，几乎一眼就能看见。但要求中心必须通过‘同出一根’检验：不能因为一句话最醒目，就把它宣布成全书根基；要看三大核心能不能从这里一起长出来，也要看对手中心是否解释得更好。

◆ 解释史校正：契约与功利并存，本来就是老问题

本书此前把“自由中心还是幸福中心”写得像第一次发现两条线并存，这需要纠正。Bellamy的剑桥版导论、Hart关于“边沁与贝卡利亚”的研究，以及Harcourt对现代刑法基础的重读，都已经注意到贝卡利亚一方面用自由让渡和契约限制刑罚权，另一方面又用公共幸福、预防和效用解释制度目标。两条线如何相容，本来就是解释史中的核心问题。

因此，“自由限权型公共幸福中心”只能作为本书对已有张力的一种压缩表达，它本身不是创新点。本书真正尝试增加的判据是：判断中心不能看哪个词出现次数多，而要看冲突时谁拥有否决权、例外必须跨越多高门槛。若公共幸福要求一种手段，而自由、法定性和必要性能够把它否决，那么“中心”就不能简单等于“最终目的”。反过来，如果某个自由限制只有在极高公共危险门槛下才可突破，门槛高度本身就是理论结构的证据。

先替自由中心做最强辩护。第一，刑罚权来源直接从自由让渡推导。国家不是天然有权，正因为每个人只贡献最小部分自由，超过必要限度的强制才成为滥用。第二，罪刑法定保护的是行动可预期性：一个人能够在行为之前知道边界，才能安排自己的生活。第三，酷刑、秘密告发、任意解释之所以可怕，都因为它们让个人

命运重新落入不可预期的人身权力。第四，死刑论证更直接问个人是否真的交出了生命。把这些放在一起，自由显然不是装饰。

再替幸福中心做最强辩护。第一，全书序言就把良法放到更广泛公共幸福的尺度上（原典锚点：剑桥版“引言”）。第二，刑罚目的不是偿还某种抽象道德债，而是防止未来损害（原典锚点：剑桥版第12章“刑罚的目的”）；有效性因此是核心。第三，刑罚确定性、及时性、温和性都以减少犯罪、降低无用痛苦为理由。第四，后半部的科学、官员、奖励、教育更难从单纯自由推出，却可以自然放进改善公共生活、减少犯罪发生的目标。第五，贝卡利亚并不把自由描述成绝对不可触碰；社会本身就是以让渡一部分自由为条件成立。

这场对决真正困难的地方在于：自由和幸福并不处在同一逻辑层。幸福更像‘为什么要建立这套制度’的目标，自由更像‘为了这个目标最多可以动到哪里’的边界。假如把地图和护栏问谁更重要，很容易得到假问题。没有地图，车不知道往哪开；没有护栏，车可能为了更快到达而直接冲下悬崖。贝卡利亚的独特性，恰恰可能不在于只选其中一个，而在于把二者安装成前后两级。

但本书的方法要求不能用“都重要”直接结束分析。那就要问：当两个变量冲突时，谁能否决谁？如果公共幸福增加可以正当化无限制刑罚，那么幸福是根，自由只是一个可交换成本；如果某些强制即使可能提高总体安全也仍然不允许，那么自由拥有更高地位。贝卡利亚在酷刑、死刑、法官任意方面确实表现出强烈边界意识，但他的反对常常同时使用‘无用’‘不必要’‘效果更差’等结果论证，并不总是建立绝对禁令。

因此，文本不支持把他轻易读成现代权利绝对主义者。他更像在说：一个真正为公共幸福服务的制度，没有理由使用超过必要限度的强制；如果一种强制不能证明必要和有效，它既伤害自由，也损害公共幸福。自由与幸福在大部分场景中不是敌人，而是同向的。只有极端案例才迫使二者分家，而贝卡利亚没有为所有极端案例建立一套完整优先序。

这使我们得到一个比‘自由中心’更接近原典的候选结构：公共幸福定向，自由最小让渡设限。　如果必须在两个视角里选一个中心，那么公共幸福目前解释范围更大；但它不是裸奔的幸福中心，而是被最小自由让渡、法律预定、比例和必要性层层限住的幸福中心。可以暂称‘自由限权型公共幸福中心’。这个名字虽然不如‘自由中心’简洁，却更能解释为什么贝卡利亚既像早期功利主义者，又始终是一位限制国家刑罚权的思想家。

用一个日常例子。学校的目标是让全体学生安全学习，这是公共效果；但学校不能因此随意搜查每个人的私人空间、公开羞辱可疑学生、用一个人的痛苦吓住其他人。目标决定制度为什么存在，边界决定制度不能怎样存在。只说‘保护自由’，可能解释不了为什么学校还要组织教育、辅导和预防；只说‘为了全校好’，又可能让任何控制都找到借口。真正可工作的制度，需要把两层一起写进结构。

SDE在这里也要警惕自己的中心主义工具。寻找中心很有用，因为它迫使我们找到承重梁；但真实思想有时就是复合结构。若为了让诊断漂亮，硬把幸福和自由压成一个谁吞掉谁的单中心，我们自己就在制造新的‘发现’。中心判断必须能输：如果一个二级结构比单中心解释力更强，就应该允许‘中心＋闸门’成为结果，而不是为了方法整齐牺牲对象。

所以本章作出第一轮改判：单纯的自由中心主义暂不成立。　更强的候选是‘公共幸福中心，最小自由让渡构成硬闸门’。这个判断还要继续接受下一章的失败测试：如果幸福真是根，前面那些自由判词会不会全部被降格成工具？如果会，我们就必须面对贝卡利亚可能比想象中更功利；如果不会，就要说明为什么幸福中心仍不能越过自由边界。

◆ 与更近的规范理论划界

“谁能否决谁”并不是无人占据的理论空地。Dworkin把权利称为对共同体总体目标计算的“王牌”；Nozick的边际约束同样强调，个人不能仅因总体结果

更好就被当作实现他人利益的手段。若只说“自由能够否决效用”，本书没有越过这些成熟的权利理论。

“例外门槛多高”也有更近的邻居。Michael S. Moore 所代表的阈值义务论允许某些义务在通常情形下保持强约束，但当灾难性后果越过极高门槛时，后果才可能压过原有禁令。贝卡利亚在死刑问题上承认非常态例外，确实可以被读出类似的阈值结构；因此，本书不能把“高门槛例外”本身算作发明。

本书还能保留的一格，是把这些原本用于回答“应当怎样做”的规范工具，反过来用于思想史诊断：面对同一作者内部的价值冲突，不统计哪个词出现得多，而观察冲突时哪一边有否决权、另一边要付出多高的例外理由成本。若这一诊断在其他思想家身上不能产生可复核、可反驳的中心定位，它也应降级为一种修辞。

◆ 校准：先在答案已知的对象上试这把尺

一把新尺子，先要在答案已知的对象上量一量。如果它连公认的结果都量不出来，拿它去量贝卡利亚，结论再漂亮也不可信。为此，本书先把诊断工具写成三个问题。第一问是否决：两种价值冲突时，哪一边让路。第二问是门槛：让路的一方要越过这道边界，需要付出多大的理由成本。第三问是位置：如果存在例外，例外开在哪一类情形，由谁判定。前两问来自前述规范理论，第三问是在检验贝卡利亚晚年材料时逼出来的（见第56章）。

第一个校准对象是边沁。在他那里，一切善恶都可以换算进同一本效用账；他列出的不应处罚的情形——无根据、无效、代价过高、不必要——全都是效用内部的判断，而不是效用之外的否决。三问的读数是：没有任何价值能否决效用，门槛为零，也就谈不上例外的位置。这把尺读出的是“效用中心”，与通行解释一致。

第二个校准对象是康德。《法权论》说，刑罚只能因为一个人犯了罪而施加于他，绝不能仅仅作为促进他人或社会利益的手段；刑罚法则是定言命令，谁要在幸福论的曲径里为罪犯寻找减免的理由，那就是不幸。他甚至设想，一个即将解散的岛上社会，也必须先处决狱中最后一个杀人犯。三问的读数是：应得对效用有绝对

否决，门槛极高。作者初读时把第三问记为“没有例外位置”，这是漏读，见本章下一节的盲编码复核。这把尺读出的是“应得中心”，与通行解释一致。

第三个校准对象是罗尔斯。《正义论》让第一原则（平等的基本自由）词典式地优先于第二原则：基本自由只能为了自由本身而受限制，不能为了更大的经济或社会利益被交换。但他也承认，在社会条件尚不足以让这些自由得到有效行使时，可以暂时采用较一般的正义观。三问的读数是：自由对效用有否决，门槛极高，例外开在“社会整体条件”这一格，而且是事先、普遍地规定的，不针对某一个人。这把尺读出的是“自由中心、高门槛、例外附着于社会条件”，与通行解释一致，也与本丛书罗尔斯卷的判断相合。

三个已知对象都读对了，才轮到贝卡利亚。读数是：在酷刑与死刑两处，授权对效用有否决（第20、30章）；门槛很高，只有国家存亡级的危机才能越过（第30、32章）；例外开在“足以颠覆国家”的人与罪这一格，而判定谁足以颠覆国家的，是国家自己（第56章）。与罗尔斯相比，差别不在门槛高低，而在位置：罗尔斯的例外附着于事先规定的社会条件，贝卡利亚的例外附着于对某一个人的个案判断。后者更危险，因为它把“必要”的宣布权又交还给了受威胁的一方。

这次校准有一个明显的限制：三个对象都是名家，答案广为人知，操作者难免受先见影响。更可靠的做法，是请不知道结论的读者按同样三问独立编码，再比较读数是否一致。在此之前，诊断工具的状态记为“通过作者校准，待盲测复核”（F8）。

◆ 盲编码复核：换一个不知道结论的人来量

按上一节的要求，本版把三问交给一位不知道本书结论的独立编码者，并把第三问拆开写成“例外挂在哪里、由谁判定”，请其凭原典对六位思想家独立编码：除边沁、康德、罗尔斯和贝卡利亚之外，另加两个本书从未校准过的对象——密尔和诺齐克。需要说明，这位编码者是一个大模型实例，只拿到问题定义与名单，没有读过本书；它能排除作者的先见，却不能替代人类读者的盲测，后者仍记为待做。

思想家	谁否决谁	门槛	例外挂在哪、由谁判定	与作者读数
边沁	效用没有对手，权利只是计算项	零	无需例外；“不应处罚的四种情形”本身就是效用计算	一致
康德	应得否决效用	极高	有例外：共犯多到全部处死会使国家失去臣民时，主权者可凭紧急命令改判流放（Ak. 6:334）；判定者是主权者自己	不一致：作者漏读此例外
罗尔斯	自由否决效用	极高	挂在事先规定的一般社会条件上，由原则事先确定	一致
密尔	自由否决家长式干预	中等	挂在行为类别上（涉及他人的伤害），另附社会发展条件	新对象
诺齐克	边际约束否决效用	极高	只在注释中为“灾难性的道德恐怖”留口，未说挂在哪、由谁判定	新对象
贝卡利亚	自由让渡设定的必要性否决超额效用	很高	死刑例外：无政府时期（一般条件），以及身陷囹圄仍握有颠覆国家之力者（个案）；判定者是国家	一致

结果有三层。第一，边沁、罗尔斯、贝卡利亚三项与作者读数一致；编码者还独立得出一个判断：罗尔斯与贝卡利亚只能靠第三问分开，因为前两问里两人都是“自由否决、门槛极高”。这正是第56章“门槛不只有高度，还有位置”的要点，现在它不再只是作者一个人的读数。

第二，康德一项不一致，而且错在作者。《法权论》谈到共犯的处置时确实留了一个例外：当共犯多到全部处死会使国家失去臣民，主权者可以用一道紧急命令改判流放。这道门的钥匙也在主权者手里。但它朝哪边开？朝宽宥开——例外让国家少杀，而不是多杀。

第三，这一处不一致逼出了诊断工具的最后一个问题：门朝哪边开。同样是被锁住的一方自己握着钥匙，门若朝宽宥开，最坏的结果是国家少罚了一些人；门若朝更重的强制开，最坏的结果是国家借例外扩大自己的强制。两者的危险完全不对称。于是三问最终写成一组锁钥问题：谁上锁（否决），锁多硬（门槛），钥匙挂

在哪、谁握着（位置），门朝哪边开（方向）。第56章用这组问题给出本书的核心判断。

F8因此改记为“盲编码部分一致”：一致的三项维持原判，不一致的一项按编码者改正，并据此补入“方向”。诊断工具没有降级为修辞，但它也不再是作者一个人的尺。

> **带走的话**
>
> 解释史早已注意到贝卡利亚的契约与功利双线。本书不再把两线并存视为新发现；当前更值得检验的，是用一组锁钥问题——谁上锁、锁多硬、钥匙在哪、谁握钥匙、门朝哪边开——判断哪一条线在冲突中拥有更高结构地位；这组问题已在三例上校准，并经不知结论的编码者盲编码复核，康德一项据此改正。

> **小账**
>
> 这一章看见了什么？——判中心看冲突时谁能否决、门槛多高、例外挂在哪、门朝哪边开。
> 接缝在哪里？——锁钥问题只经一个大模型实例盲编码，康德一项已改，人类盲测仍待做。

第 45 章

如果幸福才是根，前面的判词会不会倒

一次真正的改判不能只换标签，还要回头检查前面几十章会不会跟着塌。假如公共幸福是中心，自由只是边界，那么前文关于刑罚权、罪刑法定、酷刑、死刑、程序和身份回写的判断，是否都要重新解释？这一章专门做压力测试，不保护任何旧判词。

先看刑罚权来源。我们之前说，国家必须拿出处罚权的出生证明，因为个人只交出最少量自由。如果幸福是根，这句话可以被重新翻译：个人之所以愿意进入共同体并接受有限强制，是因为共同生活带来的安全与稳定优于彼此任意侵害。最小让渡并非把自由宣布为宇宙第一原则，而是说为了获得共同利益，不需要交出的自由就不应被拿走。这样读，社会契约仍然成立，只是自由从最终目的变成防止手段过度的稀缺资源。

再看罪刑法定。我们原来强调它保障个人预期，使人不受事后发明的罪名支配。幸福中心同样能够解释：规则公开、明确、稳定，会降低恐惧和任意，减少寻租，提高社会成员对未来的可规划性，从而改善共同生活。也就是说，罪刑法定既是自由制度，也是公共幸福制度。单靠这一条还无法决定谁是根。

再看酷刑。若酷刑极其有效，幸福中心会不会允许？这是最尖的测试。贝卡利亚反酷刑的一部分理由确实是可靠性：疼痛测出的常常是耐受力而不是真相，无辜者可能招供，强者可能坚持。他还用无罪状态和程序正当性限制刑讯。这说明反酷刑并不只建立在‘效率低’上。如果我们构造一个思想实验，假设某种酷刑能百分之百得到真相，贝卡利亚的文本是否仍能从权利和无罪原则上彻底拒绝？这里存在争议，至少不能替他写出现代绝对权利论。幸福中心在这个极端案例里显露出真正风险。

再看死刑。贝卡利亚既问国家是否得到夺取生命的权利，也问死刑是否必要和有效。双重论证本身就说明他不满足于单一效用。若社会可以通过长期剥夺自由达

到同样预防，死亡增加的不可逆痛苦没有必要；若个人并未把生命作为可让渡份额交给共同体，权力来源本身也有缺口。幸福解释第二把钥匙，自由解释第一把钥匙。死刑章节因此成为‘中心＋闸门’结构最强的物证。

再看预防。幸福中心在这里明显占优。为什么要扩散知识、改善教育、调整官员结构、奖励善行？如果唯一目标只是保存每个人已有自由，理论可以选择较消极的国家：不越界即可。贝卡利亚却不断讨论怎样主动构造更少犯罪、更少恐惧的社会条件。这个积极面说明他不仅想把国家手缩短，也想让制度更聪明。公共幸福比自由更能解释这种主动设计。

那么自由是否因此沦为一个方便工具？不完全是。贝卡利亚的最小让渡逻辑使自由不是普通成本。若某项公共目标根本不需要牺牲这部分自由，拿走它就失去正当性；权力必须持续证明‘非拿不可’。这和单纯计算总收益不同。即使总体收益有增加，如果有更轻的替代方案，较重强制仍然越界。自由在这里不是永不触碰的绝对墙，却是一种必须由必要性逐步打开的门。

从SDE看，这比简单‘自由中心’更有意思。幸福像D1意义目标：制度为什么往这个方向走；自由像优化约束：即使目标正当，也不能任意选择路径；而法律、比例、程序、教育则是不同层次的路径组织。这样一来，贝卡利亚的刑罚论不是一条单变量直线，而是一种目标—路径—约束结构。我们先前把自由当第一性，可能是因为它在刑罚权章节最亮；把全书后半部纳入后，目标层的幸福才显出来。

但新的风险也随之更清楚：谁来定义公共幸福，谁就可能控制整套优化。贝卡利亚在序言中赞赏一种能够冷静观察人性、把众多人的行为放在同一视野里加以衡量的立法者形象。这个形象在反对旧法的激情与偶然时非常有力，却也可能成为新的高位：一个人仿佛站在社会之外，知道什么对多数人最好。这里很可能出现全书新的‘收回’候选——不是‘先放一个自由人’，而是‘先放一个冷静的立法者’。

这一候选比此前更接近贝卡利亚后半部的困难。教育谁来设计？奖励什么美德？哪些知识值得扩散？什么风险需要预防？什么刑罚最有用？如果每个答案最后

都依赖一个能从整体看社会的理性立法者，那么贝卡利亚虽然把君主的任意赶下台，却可能把‘公共幸福的观察者’请上另一个高台。旧权力说‘我因为身份而知道’，新权力说‘我因为理性而知道’。二者当然不同，但都需要继续追问观察者自己的土。

这一章因此没有把前面的判词推倒，反而把它们重新排序：自由不再是唯一中心，却成为幸福政治不可随意跳过的闸门；必要性不是纯粹效用，而是连接目标与自由成本的审查；程序不是效率附属物，而是防止‘公共幸福’被少数人冒名代表的制度装置。原来的很多判断不但不倒，反而得到更完整的位置。

> **带走的话**
>
> 幸福中心并没有摧毁前面的自由判词，而是把它们从‘最终目的’改写成‘反暴政闸门’。真正开始危险的，不是幸福本身，而是有人宣称自己能够站在社会之外替所有人计算幸福。

> **小账**
>
> 这一章看见了什么？——改判后旧判词不倒：死刑双钥匙里，自由管来源，幸福管必要。
>
> 接缝在哪里？——若酷刑能百分之百得真相，文本能否彻底拒绝尚有争议，幸福中心在此露险。

第 九 编

中心主义审判：把三大核心重新放上桌

编 序

本编三章。第46章把权力来源、刑罚效果、犯罪预防三大核心重新放上桌；第47章做同出一根检验，看三者能否都从公共幸福长出来；第48章用九格总对账，看他到底把哪一格放在中央。

本编是判决四的主场。三组检验指向同一结构：公共幸福定方向，最小自由让渡设硬闸，创造多处显影却未统合。自由限权型公共幸福中心由此正式登记，追问也随之转向谁来判断公共幸福。

读这一编，可以带着日常问题进去：学校要让全体学生安全学习，能不能因此搜查每个人的书包？九成人觉得更安心，少数人却天天被盘查，这还算公共幸福吗？

第 46 章

把贝卡利亚三大核心重新放上桌

写到这里，我们不能只靠逐章点评向前走。一本专著如果只有八十个聪明判断，却没有一个能把它们收住的承重结构，最后仍然只是高质量读书笔记。现在要把前文和本编的材料全部压回三个核心问题，看它们究竟是不是一棵树上长出的三根大枝。

第一核心是权力从哪里来。这条线从全书最早的社会契约、最小自由让渡出发，一路经过罪刑法定、法官受限、程序、酷刑、死刑。它反复追问：谁有资格对另一个人施加强制？什么程序使强制可被接受？什么强度已经越过授权？在这一核心里，贝卡利亚最锋利的贡献是把‘国家当然能罚人’变成必须证明的结论。刑罚权第一次被要求拿出生证。

第二核心是怎样让刑罚真正减少犯罪。这条线包括刑罚目的、确定性、及时性、温和、比例以及死刑效用。它把刑罚从象征性报复拉回人的行为机制：人怎样感受风险，为什么极端严厉不一定更有效，延迟怎样切断行为与后果的联系，逃脱希望怎样削弱威慑。这里的贝卡利亚不是法权哲学家，而像一个早期制度设计师，要求每一份痛苦都拿出功能说明书。

第三核心是怎样改变犯罪发生的环境。这条线直到后半部才完全展开：清楚法律、反特权、知识扩散、官员结构、奖励、教育、减少无用禁令。它比第二核心再向前一步。第二核心仍在问‘怎样处罚更有效’，第三核心问‘怎样让处罚少到不必发生’。制度不只是站在终点等人犯错，而开始参与塑造人走到终点以前的路。

三个核心如果分别看，很容易被归入三种理论：第一像权利与法治，第二像功利与威慑，第三像社会政策与教育。真正的问题是，贝卡利亚为什么能把它们写在同一本薄书里而不觉得互相冲突？如果找不到共同根，我们就只能承认他是一个把启蒙时代多种思想拼在一起的改革者；如果能找到共同根，‘贝卡利亚哲学’才真正出现。

自由能不能做这个根？它对第一核心解释极强：国家权力必须有限，因为个人并未交出全部自由。对第二核心也能解释一部分：既然强制侵犯自由，刑罚只能达到必要的最小程度，所以必须寻找轻而有效的方案。但到了第三核心，自由的解释力明显下降。为什么奖励善行？为什么扩散知识？为什么教育要形成稳定感受？这些并不是单纯‘少限制自由’就能推出的。

公共幸福能不能做这个根？对第一核心，它说社会之所以形成有限刑罚权，是为了共同安全与更好的共同生活；无限权力反而制造恐惧，破坏公共幸福。对第二核心，它直接要求刑罚以更低痛苦取得更可靠的预防效果。对第三核心，它更自然：知识、官员、奖励和教育都是改善公共生活、减少犯罪生成条件的办法。就解释覆盖面而言，公共幸福明显胜出。

但还需要第三个候选：安全。也许贝卡利亚真正中心既不是抽象自由，也不是广义幸福，而是个人安全。人们让渡自由为了安全；犯罪按危害社会安全排序；刑罚为了阻止新侵害；程序保障免受国家任意；预防为了让社会更稳定。安全几乎也能串起全部内容。可如果安全是唯一中心，知识、教育、平等和自由就容易被缩成治安工具，贝卡利亚关于幸福和反特权的更广阔关怀会被压窄。安全更像公共幸福的底座，而不是全部。

还有一个候选叫理性。贝卡利亚显然反对激情立法、传统迷信和司法任意，希望用清楚、可比较、可公开说明的理由重新安排刑罚。但‘理性’仍然更像方法，不像目标。理性可以替暴政服务，也可以替自由服务；真正决定它往哪里用的，仍然是公共安全、自由与幸福的价值结构。把贝卡利亚简单叫‘理性主义中心’，会把他的政治目的藏起来。

所以，三大核心重新上桌以后，暂定根已经比较清楚：贝卡利亚是在做一项公共幸福工程。他不是把幸福理解为任意总量，而是把公共安全、最低限度强制、规则可预期和较广泛的利益分布放进同一个政治目标。自由之所以重要，是因为一个靠任意强制换来的‘幸福’在他看来不能稳定，也不符合社会契约；预防之所以重要，是因为最好的公共生活不应靠持续制造痛苦维持。

这也解释了为什么他的书最后不是以一项刑罚技术结束，而是走向教育。一个理论如果最后停在‘怎样判’，说明它仍以案件为中心；贝卡利亚最后问‘怎样让人少走到案件里’，说明他的视野已经从司法机关扩大到社会本身。刑法只是公共幸福工程的一道窄门。

> **带走的话**
>
> 贝卡利亚三大核心可以重新压成三个问题——权力怎样正当、刑罚怎样有效、犯罪怎样少发生。自由最擅长解释第一问，幸福能同时解释三问；因此公共幸福中心的证据进一步增强。

> **小账**
>
> 这一章看见了什么？——三大核心压成三问：权力怎样正当、刑罚怎样有效、犯罪怎样少发生。
>
> 还要追问什么？——安全与理性也能串起全书，前者更像幸福的底座，后者只是方法。

第 47 章

它们真的都从公共幸福长出来吗

现在进行‘同出一根’正式检验。做法不能是把三个核心都贴上‘为了幸福’四个字，那只是换标签。真正的同根，要求从同一个起点能够连续推出关键判断，而且中间不能偷偷借用另一套未说明的原则。我们用三个核心逐一演算，再让自由中心做一次反向竞争。

第一组：权力来源。假设根是公共幸福。人们单独处在不稳定环境中，彼此的自由容易冲突，因此需要共同规则保障安全；共同强制只有在能改善这种共同生活时才有理由；既然强制本身制造痛苦并压缩自由，就只能取达到目标所需要的最小部分；于是超过必要的刑罚不是‘多一点保险’，而是公共制度自己增加无益损害。由此可以推出社会契约、最小让渡、必要性和比例。这个推导基本连续。

再假设根是自由。个人自由需要保护，因此建立国家；国家只能拿走最少自由；所以刑罚要有限。也能推出来。但这里有一个缺口：为什么个人需要建立共同体，而不是最大限度保持私人自由？必须加入安全、共同利益或更好生活，才能说明自由为什么愿意让渡自身的一部分。也就是说，自由在起点很强，却需要借公共结果解释为什么进入制度。

第二组：刑罚机制。假设根是公共幸福。刑罚只是达到较安全社会的一种工具，所以应比较确定性、及时性和严厉性，寻找较低痛苦、较好效果的组合；如果更重并不更有效，多出来的痛苦没有公共理由；如果某刑罚反而诱发更残酷行为，它就违背自己的目的。贝卡利亚关于温和、确定与及时的论证几乎顺着这一根自然长出。

再假设根是自由。由于刑罚剥夺自由，应尽量轻；但为什么‘确定性’和‘及时性’特别重要？单纯自由并不能回答。为了说明这些机制，仍要引入威慑效果和犯罪预防。自由能给出上限，却不能单独给出最佳路径。因此第二核心明显由幸福／效用目标解释得更完整。

第三组：犯罪预防与教育。假设根是公共幸福。既然目标是减少社会损害和无用痛苦，那么最理想的状态不是高效处罚很多人，而是较少犯罪、较少处罚；因此要让法律清楚、减少特权、传播知识、设计守法官员、奖励有益行为、改善教育。每一步都在改变共同生活的发生条件。这里几乎不需要额外借力。

再假设根是自由。我们当然可以说知识使人更自由、教育培养自由、特权伤害自由，但这些论证需要给‘自由’装进越来越多内容，最后自由会变成一个什么都能解释的词。一个理论概念如果必须不断扩义才能覆盖文本，往往说明它不是最好的根。

因此，同根检验的结果已经足够清楚：公共幸福比自由拥有更强的全书生成力。　这并不意味着贝卡利亚主张可以为了多数人幸福牺牲任何个人。恰恰相反，他的独特结构是把个人自由、法律预定和必要性放进公共幸福工程内部，成为限制手段选择的硬条件。于是我们可以第一次比较正式地命名：贝卡利亚在刑罚问题上呈现的是一种自由限权型公共幸福中心主义。

这个判断还需要和成熟功利主义保持距离。贝卡利亚没有建立一台可以把所有人的快乐、痛苦统一计量的计算机，也明确知道政治判断无法达到几何学精确。他更常做的是方向性比较：无用痛苦应减少，任意应减少，逃脱空间应减少，安全和可预期性应增加，制度利益应更广泛分布。把这些直接等同于后来系统化的效用最大化，会把发生史倒过来。

而且，他的公共幸福不是纯粹结果论，因为程序本身具有独立价值。法律必须预先存在，法官不能随意增加刑罚，被告不能因为社会愤怒就被酷刑逼供。即使某些程序限制短期看起来降低效率，它们仍然保护共同体不滑回私人权力。最合理的理解不是‘结果压倒程序’，而是程序本身也是公共幸福长期结构的一部分。

这一步也把此前‘自由中心 vs 幸福中心’的二选一改造成一个更精确层级：幸福居于目标层，自由居于正当强制的约束层，创造则开始出现在制度如何制造犯罪类别、信任关系和主体能力的发生层。三者并不是一锅粥，却已经有明确主次。

对于SDE来说，这种结构比把一个视角抬成绝对本体更有价值，因为它暴露了贝卡利亚真正偏重的地方，也保留另外两维对中心的制约。

现在再看‘收回’。如果公共幸福是中心，那么原来的‘先放一个自由人’虽然仍然存在，却未必是最深接缝。更深的接缝可能发生在另一处：为了让公共幸福成为可执行的政治目标，贝卡利亚需要一个能够识别共同利益、比较政策效果、摆脱私人激情的立法者。这个立法者在文本里并非神，但常常像站在高处的冷静观察者。其他人的激情、利益和无知都被发生化了，观察者自己的观察位置却没有同样充分地被追问。

这给我们一个更具体、更能画出来的动作名候选：‘把立法者请上看台。’看台上的人俯视整场比赛，能看见各方行为，能比较规则效果，能宣布什么更有利于大家；可他自己也有视角、信息、利益和时代限制。贝卡利亚用这个位置替代旧法的激情，是巨大进步；但如果看台永远不接受反看，它就可能成为新的免检区。后面的总解构必须专门处理这一点。

> 带走的话
>
> 同出一根检验支持“公共幸福定方向、自由设硬闸”的解释，但这一双线结构有充分的解释史先占。本书的候选增量只落在判定中心的方法：看谁拥有否决权、例外要跨越多高门槛，而不是数概念出现频率。

> **小账**
>
> 这一章看见了什么？——自由须不断扩义才盖住教育与知识，公共幸福却能连续推出三核心。
>
> 接缝在哪里？——他人的激情都被发生化，冷静立法者自己的观察位置却没被同样追问。

第 48 章

九格总对账：他到底把哪一格放在了中央

中心已经改判，但还不能只靠文字直觉。按照的工艺，现在把贝卡利亚放进九格：横向看理念界、现实界、自我界，纵向看创造、自由、幸福。九格的目的不是给他打九个分，而是检查哪些位置写得最厚、哪些位置只是偶尔闪现、哪些位置成为其他位置的派生物。只有这张表走完，‘公共幸福中心’才算真正过闸。

先看第一行：创造 × 理念界。 贝卡利亚非常清楚法律并非只是把自然界已有的犯罪名称抄下来。什么行为进入刑罚、怎样分级、什么程序成立，都要经过立法。尤其当他批评无用法律会制造新的违法、削弱有用法律权威时，他已经看到制度会生产自己面对的对象。可他主要关心的是怎样把这种创造做得更合理，没有继续追问法律分类本身怎样在历史冲突中形成。创造被看见，却没有成为中心。

第二格：创造 × 现实界。 刑罚会真实改变身体、时间、财产和社会机会。监禁不是纸面句子，死刑更不是抽象符号。贝卡利亚对酷刑、死刑和过度严厉的敏感，使他不断把制度概念拉回身体后果。他还看到高额奖励、官员利益、制度执行方式会改变实际行为。但这些现实变化主要被评价为是否有助于公共目标，创造本身仍是被观察的过程，而非最终价值。

第三格：创造 × 自我界。 这是本书前文直到秘密告发、污名和教育才逐渐显出来的一格。制度会制造猜疑，标签会改变身份，教育会形成习惯与感受，奖励会塑造荣誉方向。也就是说，‘人’不是完全在制度之外等着被法律处理。这个格子对我们非常关键，因为它直接削弱了‘经验主体先在’的旧判词。但贝卡利亚没有把这些零散洞察合成一套主体发生理论，所以创造轴总体仍然偏弱。

第二行进入自由 × 理念界。 这是贝卡利亚最强的一格之一：罪刑法定、立法与司法分工、法律明确、一般规则、社会契约、最小自由让渡。自由在这里首先表现为规则边界：国家不能看见一个具体人以后才临时决定什么是罪，也不能让法官

以热情为名扩张刑罚。理念世界里的自由，是‘事先知道国家能做什么、不能做什么’。这几乎构成现代刑事法治的骨架。

第五格：自由 × 现实界。 酷刑、监禁、死刑、财产刑把自由从抽象权利变成身体问题。一个人究竟被关多久，能不能被折磨，国家有没有资格夺取生命，这些不是理念上的自由口号，而是现实里的门、锁、绳索、时间和身体。贝卡利亚在这一格同样很强。尤其死刑论证让‘自由让渡到哪里为止’变成最尖锐的边界问题。

第六格：自由 × 自我界。 这一格比前两格复杂。贝卡利亚一方面预设能够选择、趋利避害的行动者，另一方面又承认教育、激情、知识、处境会塑造人的选择。他没有系统回答‘一个人要具备什么能力才算真正自由’，也没有建立现代责任能力理论。但正是在这里，规范主体和经验主体第一次撞到一起。自由不是完全缺席，而是接口不足。

第三行是决定中心的幸福轴。先看幸福 × 理念界。 全书最开头就把良法指向更广泛的公共幸福，后面所有制度安排都在问怎样用一般规则减少任意、特权和无用痛苦。幸福在理念界不是私人快乐，而是立法应该服务的公共方向。法律本身的优劣，最终要接受这一方向检验。

第八格：幸福 × 现实界。 这是刑罚技术的大本营。确定性为什么比极端严厉更重要，及时性为什么能加强行为与后果的联系，温和为什么可能更有效，死刑为什么要证明必要，所有这些都在比较真实社会效果与真实痛苦成本。贝卡利亚不是为了得到一个更漂亮的理论，而是为了让制度少制造无用伤害、同时更可靠地保护安全。幸福在这里直接决定路径优化。

第九格：幸福 × 自我界。 预防、知识、奖励和教育最终都落到人的生活感受上。一个人是否只怕具体权势人物，还是信赖一般规则；是否把告密和背叛看成通往利益的正常手段；是否在成长过程中形成对他人边界的感受；是否相信守法者和特权者面对同一尺度——这些经验共同构成社会是否宜居。公共幸福在这里不是统计结果，而是被每个人实际经历的安全、信任与可预期性。

九格走完以后，一个结构非常清楚：贝卡利亚并不缺自由，反而自由轴极强；他也已经多次看见创造，尤其是制度对犯罪类别、信任和人的习惯的回写。但真正把全书不同部分串成方向的，是幸福轴。刑罚权为什么需要限制？因为无必要的强制增加痛苦、破坏共同安全；为什么讲预防？因为不犯罪比惩罚犯罪更符合公共利益；为什么讲教育？因为社会要形成能够较少依赖强制的成员。幸福不是每章都写在标题上，却像重力一样一直在拉。

因此，本编可以正式更新中心诊断：贝卡利亚不是自由中心，而是自由限权型公共幸福中心。　自由不是被幸福吞掉，而是作为公共幸福工程的第一道硬闸；创造则被他多次看见，却没有获得同等理论地位。用SDE的话说，他的强项不是三维完全平衡，而是幸福目标非常清楚、自由约束非常有力、创造维度零散显影。

而新的接缝也因此更加集中。既然公共幸福是中心，谁能判断公共幸福就变得异常重要。贝卡利亚批判旧法来自少数人的激情和偶然，需要一个更冷静的立法视角来比较社会效果。问题是：这个观察者站在哪里？他如何知道‘更广泛的人’真正需要什么？被治理者能不能反过来改写目标？少数人的痛苦什么时候能阻止多数人的收益？知识不足时怎样承认不确定？这比‘自由人是否先在’更可能成为本书最终要钉住的裂缝。

所以，本章把新的收回候选正式登记为：‘把立法者请上看台。’　贝卡利亚把君主从神授高台上拉下来，却仍然需要一个理性立法者在看台上观察全局、安排最有利于公共幸福的规则。这个动作不能现在就判死，因为看台也可能只是方法角色，而不是形而上学先在。后文必须回到原典继续抓接缝：看这个立法者有没有被制度约束、有没有被公众经验回写、有没有承认自己可能看错。

九格还给SDE自己一个警告。我们很容易因为九格漂亮，就说‘贝卡利亚缺创造，所以SDE更完整’。这不够。多一个格子不等于理论更真。SDE必须证明：补入创造和回写以后，能够更好解释具体法律问题，能够产生会输的预测，能够说明在哪些条件下贝卡利亚反而比SDE更有效。否则九格只是一个把别人重新分类的柜子，自己却不接受分类。

这一部分到这里，真正完成的不是多写八章，而是一次中心改判。从前文的‘国家发生，主体先在’，到前文的‘规范主体与经验主体接口不足’，再到本编的‘自由限权型公共幸福中心’，我们已经连续两次让原典反过来修正SDE的第一印象。这恰恰是这套方法应该有的样子：判词不是为了显得锋利而焊死，而是要允许对象用自己的文字把我们的刀顶回来。

◆ 扩展论证　八个压力测试：把中心判断放回生活

压力测试一：为了降低犯罪率，能不能把所有人都监控起来。假设一套系统真的能够显著减少盗窃和暴力，只要每个人二十四小时都被识别、定位、记录。从公共幸福的表面计算看，犯罪下降、安全提高，似乎值得；从贝卡利亚的最小让渡结构看，却不能只比较结果。还要问：实现同一目标有没有更轻手段，监控权由谁使用，数据能否被挪作他用，普通人有没有拒绝和申诉空间。只要更轻的路径足够有效，无限监控就越过必要性。这个例子说明，‘幸福中心’若没有自由闸门，会立刻膨胀成治理万能论。

压力测试二：为了震慑全班，能不能重罚一个学生。老师发现一次作弊，决定公开取消这个学生一年评优资格，以便让其他人再也不敢作弊。结果也许真的有效，但贝卡利亚式结构仍会问：这一后果和行为是否相称，是否事先明确，是否存在教育和修复手段，处罚是不是把一个人当成制造别人恐惧的工具。公共效果不能把具体人的边界全部冲走。这里与康德的警告重新相遇：贝卡利亚更看重未来效果，却不能因此把受罚者只当作社会工程材料。

压力测试三：如果奖励比惩罚更有效，制度是不是应该只奖励。贝卡利亚专门提出奖励美德，说明他已经不满足于‘做错就打’的单向机制。但奖励同样会改变行为结构。单位如果所有人都围着奖金做事，未被奖励的必要工作可能被忽视；学校若只奖励分数，学生会学会追逐可计量结果。奖励不是天然温和，它也在塑造目标。真正的问题依然是：制度想让什么发生，为什么这个目标值得，被奖励的行为会不会反过来改变共同体。

压力测试四：知识越多，犯罪一定越少吗。贝卡利亚对科学与知识传播抱有启蒙时代的强烈信心：愚昧、偏见和秘密权力退潮，公开理性更有可能出现。但今天我们知道，知识也能提高犯罪技术，算法也能被用于诈骗，专业能力并不自动生成善。于是他的判断应当被收窄：知识不是善的保证，它首先减少的是某些依赖无知和神秘的统治形式；知识要服务公共幸福，还需要制度责任、价值教育和可追责的使用边界。

压力测试五：教育能不能把所有问题都变成个人问题。社会面对犯罪时很容易说‘加强教育’，仿佛只要把人教好，贫困、歧视、机会结构和制度激励都可以不谈。可贝卡利亚把教育放在一整套预防结构的最后，而不是拿教育替代所有制度改革。清楚法律、减少特权、改善官员、奖励良好行为都在前面。真正的教育不能替坏制度洗白。一个社会如果规则本身不公平，却要求孩子通过教育学会顺从，它培养的不是自由主体，而是适应不公的主体。

压力测试六：多数人觉得安全，少数人长期害怕，算不算公共幸福。假设一项政策让九成居民感觉更安心，却使某个少数群体持续遭到额外盘查、怀疑和身份标记。单看多数满意度，政策可能成功；但贝卡利亚反对特权、强调共同法律和个人安全，使这种结果无法轻易被叫作公共幸福。真正的公共必须能解释为什么同一套制度没有把安全建立在某一群人的持续不安全上。‘共享’二字因此比‘总量’更重要。

压力测试七：如果立法者比公众懂得更多，是否可以替公众决定。公共卫生、金融风险、人工智能等领域都需要专业知识。一个完全拒绝专家判断的制度会变得低效甚至危险；可专家若凭专业身份直接取得最后决定权，又可能重现‘看台上的人’。贝卡利亚留下的难题不是专家要不要存在，而是专家知识怎样进入公开法律、怎样接受反驳、怎样说明不确定、怎样被事实纠正。理性不是免检身份，而是一种必须留下论证路径的工作方式。

压力测试八：如果AI能比人类法官更稳定，贝卡利亚会支持吗。仅从减少法官情绪和同案差异看，算法似乎实现了他对稳定规则的梦想；但算法的训练数据、特征选择和目标函数同样可能把旧偏见藏进机器。如果系统以‘降低再犯率’为唯一

目标，它可能把某些群体长期判为高风险。于是两百多年前的中心对决重新出现：公共幸福的目标是谁定义，自由闸门怎样进入模型，模型是否允许被告理解并挑战。AI不是贝卡利亚问题的终结，而是把问题从法官的内心搬进了系统设计。

这八个压力测试共同说明，本编的中心改判不是把‘自由’换成‘幸福’那么简单。公共幸福只有在它能被分解为公开目标、最小强制、平等分配、可纠错程序时才具有正当力量；一旦它被压缩成一个总分，任何人都可能以提高总分为理由要求别人让路。贝卡利亚的真正现代性，也许正是他同时给出了两种语言：用幸福要求制度解释效果，用自由要求制度解释代价。

◆ 理论深化　四个必须留给后文的问题

一、公共幸福能不能由多数决直接代表。不能。多数决定的是一种程序结果，不自动等于公共幸福；如果多数可以用选票取消少数的基本程序和人身安全，贝卡利亚的限权结构就会被抽空。后文必须区分‘多数偏好’与‘公共理由’，否则幸福中心会滑成多数中心。

二、理性立法者是不是贝卡利亚新的免检区。这个问题需要原典继续核。序言里的冷静观察者、后半部对知识和官员的讨论，确实都强化了一个从总体看社会的视角。但他同时强调一般法律、公开规则、平等约束，并非简单把立法者变成无责任专家。后文要抓的，是立法者有没有制度性的被纠错路径。

三、教育与自由到底是什么关系。如果教育只是让人更服从现存规则，它会成为温和版本的强制；如果教育让人能够理解规则、质疑规则并参与形成规则，它才真正增加自由能力。贝卡利亚的教育短章提供方向，却没有完整回答这一分岔。这里非常适合连接后来的公民教育与AI时代个性化教育问题。

四、幸福中心是不是会把刑法扩张成整个社会工程。贝卡利亚自己从刑罚一路走到科学、官员、奖励和教育，视野越走越宽。好处是看到犯罪发生条件，危险是刑法问题可能吞并社会问题。后面必须守住一条线：预防犯罪不等于把所有不理想生活都犯罪风险化。刑法越懂发生，越应该知道什么时候退出。

◆ 时代转接　三幅今天的画面

第一幅画面，是城市里越来越多的摄像头、门禁、实名系统和风险预警。它们都可以用‘预防’说明自己：不是等事故发生再处理，而是尽量提前识别危险。贝卡利亚会理解这种愿望，因为他自己就把刑罚从事后痛苦推向事前制度；但他也会逼问一件事：为了预防，究竟需要拿走多少自由？如果一项技术只因为存在，就被自动认为必要，那么‘必要性’已经失去限制作用。现代治理真正需要继承的不是他的某项具体答案，而是双重证明结构——既证明目标有公共理由，也证明所用手段没有超过实现目标所需的限度。

第二幅画面，是平台社会里的评分。司机有评分，商家有评分，借贷有评分，内容有评分，未来司法也可能越来越依赖风险分。评分最大的诱惑，是把过去很模糊的判断变成统一数字。可本编九格已经提醒我们，数字不是中立的自然物：选什么指标、怎样赋权、用多长时间的数据、允许什么纠错，本身都是制度判断。贝卡利亚反对法官凭情绪任意，本意是让权力更可预期；如果今天把不可见的算法权重当成新的‘冷静观察者’，只是把任意从法官脸上搬进服务器。

第三幅画面，是人们越来越习惯把复杂社会问题交给刑法表达态度。出了新风险，就要求入罪；发生严重事件，就要求加重；看到令人愤怒的行为，就希望刑法立即给出最强回应。贝卡利亚最值得今天重新听见的，不是某一条刑罚建议，而是他的顺序：先问真正的社会损害，再问是否需要刑罚，再问有没有更轻的制度办法，最后才问刑罚多重。这个顺序把愤怒变成论证，把‘应该严惩’拆成一连串必须回答的问题。越是信息传播迅速、情绪可以瞬间聚集的时代，这个顺序越重要。

三幅画面放在一起，可以看出贝卡利亚为什么仍然不是博物馆里的刑法史人物。十八世纪面对的是酷刑、秘密审判和身份特权，二十一世纪面对的是算法、数据、风险治理和舆论压力；工具完全不同，但结构问题惊人相似：国家或平台能不能因为‘为了大家好’就多拿一点权力，谁来证明那一点确实必要，普通人是否知道规则怎样运作，错误能不能被纠正。公共幸福若要成为今天仍可使用的概念，就必须始终和这些程序性问题绑在一起。

◆ 原典对读札记　五处容易把贝卡利亚讲坏的地方

第一处，不能把‘最大幸福’直接翻译成‘多数人想怎样就怎样’。贝卡利亚反复关心的不是一次投票里的多数偏好，而是法律怎样让共同体成员较稳定地享有安全、可预期和较少无用痛苦。他批评特权，要求一般法律，要求社会同样受契约约束。这些内容都说明，公共幸福并非简单人数优势。若把它讲成‘多数压少数’，会把他最重要的限权装置全部删掉。

第二处，不能把‘预防犯罪’讲成现代风险治理的无限提前。贝卡利亚当然希望犯罪少发生，但他主要依靠公开法律、知识、制度一致性、官员守法、奖励和教育来预防，而不是把可能性直接当罪责。今天的风险评分、预测性警务和大规模数据监控可以和他的预防思想对话，却不能冒充就是他的方案。历史上的开门不能被后来技术装修成另一间房。

第三处，不能把‘教育’讲成软性的服从训练。他明确警惕只靠命令得到短暂、表面的服从，更看重感受、经验与必要后果的形成。这个细节说明他并不满足于一个被外力推着走的人。教育真正值得保留的方向，是形成判断能力和稳定行为，而不是把权威命令藏进孩子心里。若把教育写成国家塑造顺民，反而会错过他为何把教育放进预防而不是处罚。

第四处，不能把‘理性立法者’神化。贝卡利亚需要冷静、公开、可比较的立法理性，是为了反对旧法被激情、秘密和少数利益支配；这个历史方向必须保留。但理性不是一个人的身份，也不是一句‘我是专家’。真正值得继承的是论证可公开、规则可一般化、政策要比较效果和代价。谁若把自己宣布为天然理性的观察者，反而落入我们后文要追问的看台问题。

第五处，不能把‘自由限权型公共幸福中心’写成一个新教科书标签。它只是本书经过六批原典攻击后形成的工作判断，必须继续接受后面章节的推翻。如果第49章以后发现贝卡利亚把某些个人权利明确放在任何公共效用之前，中心还要向自由回摆；如果发现立法者在他的制度中已经被充分回写和纠错，‘看台’也要撤。判词的价值不在于响，而在于它随时可能输。

第一，本编完成了一次实质改判。此前第一候选‘自由中心主义’降级，正式改为‘自由限权型公共幸福中心主义’：公共幸福居于意义目标层，自由最小让渡、罪刑法定、必要性和比例构成强制路径的硬闸门。

第二，本编进一步收回‘经验主体先在’。教育章节显示贝卡利亚已经明确看到人的感受、习惯和判断会被培养；因此真正问题不是他完全看不到主体发生，而是没有完整说明规范自由资格与经验形成条件之间怎样转换。

第三，新的‘收回’候选出现：‘把立法者请上看台。’旧法的问题在贝卡利亚看来，是激情、特权与偶然支配法律；新秩序需要更冷静地观察社会效果。可观察者自身的视角、信息与纠错如何发生，尚未充分显形。这一候选将在后续第49—56章接受原典接缝检验。

第四，九格对账显示：幸福轴最厚，自由轴次之且具有否决功能，创造轴多处显影但没有统合。尤其‘法律制造犯罪类别’‘秘密制度制造猜疑’‘教育塑造选择者’三处，已经证明贝卡利亚并非只会处理定型结果。

第五，本编给SDE自己留下推翻条件：如果后续原典表明贝卡利亚对立法者的知识局限、公众参与、制度纠错和目标变化已经有系统安排，那么‘把立法者请上看台’必须降级；如果公共幸福无法解释其最关键的权利限制，而自由具有不可被效用交换的绝对地位，则本编中心改判也必须撤回。

> **带走的话**
>
> 九格对账确认，贝卡利亚的公共幸福轴最能统领全书，自由轴承担强限制，创造轴多处显影但尚未统合。新的总收回候选由‘先放一个自由人’升级为‘把立法者请上看台’：真正需要追问的，开始变成谁有资格替共同体判断什么叫公共幸福。

小账

这一章看见了什么？——九格里幸福轴最厚，自由轴次之且能否决，创造轴多处显影未统合。

还要追问什么？——多一格不等于更真，SDE须证明补入创造后能给出会输的预测。

第 十 编

看台与楼梯：谁替共同体判断公共幸福

编 序

本编八章。第49章分清哪里是中心、哪里只是结果；第50章作出第一场正式判决；第51章问社会契约以前站着谁；第52章问那个自由的人有没有童年。

第53章看语言、家庭、贫困、身份怎样参加犯罪选择；第54章讲自由也是长出来的；第55章把「先放人，再限权」降为规范策略；第56章用晚年材料终审，得出锁钥同手。

本编是判决五的主场。看台上的立法者要替众人判断公共幸福，就必须公开理由、接受反驳；楼梯不足、钥匙握在国家手里，说明纠错通道正是最该追问的地方。判决四的锁钥同手也在本编定形。

读这一编，可以带着日常问题进去：业主公约写着「紧急情况除外」，什么算紧急，由谁说了算？一项规定明显过时，普通人的经验有没有一条路走回去改它？

第 49 章

哪里是中心，哪里只是结果

一个思想体系最容易被讲坏的方式，是把所有重要词都并列起来。自由很重要，幸福很重要，安全很重要，法律很重要，预防很重要，教育也很重要——如果到这里就停，当然不会错，却什么也没有解释。真正的中心诊断必须回答：这些词之间有没有上下关系？哪些是我们最终想得到的，哪些是为了达到目标不得不经过的路，哪些是为了防止路径失控而设的边界，哪些只是制度运行以后出现的结果？

把贝卡利亚放回这一张关系图里，首先要分清“公共幸福”和“自由”到底各自承担什么功能。序言里的那位冷静观察者，被设想为把众人的行动导向“最大多数人的最大幸福”；刑罚权章节又说，每个人只是因为必要才交出一小部分自由，而且只会交出最低限度。两个判断看上去都像根。可如果继续问：为什么只能交最少自由？答案并不只是因为自由神圣，而是因为多拿走的自由不再对共同安全有必要，反而增加无用痛苦，损害共同生活。也就是说，自由的“少交”本身已经在接受公共目标与必要性的检验。

但反过来同样成立：如果公共幸福可以随时要求个人牺牲，那么“最大幸福”就会变成一张没有边界的授权书。贝卡利亚显然不愿意走到这里。罪刑法定、禁止法官临时加码、反酷刑、反对无必要的死刑、要求比例和确定性，都在告诉国家：即使你声称为了大家好，也不是想拿多少自由就拿多少。于是自由不是幸福的普通手段，它更像目标函数旁边的一组硬约束。

可以用一个修路的例子理解。城市希望减少拥堵，这是目标；可以拓宽道路、调整公交、错峰、收费，这是路径；但不能为了交通效率随便拆掉任何人的房子，更不能不补偿、不说明程序。产权和程序权在这里不是交通效率的结果，而是优化过程不得随意穿越的边界。贝卡利亚的自由，在许多章节里正承担类似角色：公共幸福告诉制度“往哪里走”，自由告诉它“哪些捷径不能走”。

必要性又是什么？它不是最终目标，而是一道比较闸门。国家若要限制自由，必须说明为什么这个限制对共同安全不可缺少，为什么更轻的手段不够。比例是什么？它是让不同强制保持差异秩序的组织原则。确定性和及时性是什么？它们是提高预防效果、减少靠极端严厉制造震动的路径条件。预防和教育则更明显：它们是希望在刑罚出现以前改变犯罪发生条件的上游路径。

这样一拆，贝卡利亚的体系突然有了层级：最上面是公共幸福或公共福祉的方向；中间是自由、平等、必要、比例等限制与结构；再往下是罪刑法定、程序、确定性、及时性、预防、教育等制度技术；最底下才是犯罪率下降、恐惧减少、社会更可预期等经验结果。把这些层混在一起，就会误以为贝卡利亚只是罗列改革建议；分开以后，才看见一个非常现代的治理模型正在形成。

这张图同时也使“幸福中心”接受第一次反击。假如公共幸福只是所有制度运行良好的统称，而不是一个真正能指导选择的标准，那它就不能算中心。所谓中心，必须在发生冲突时能够裁决。例如：提高刑罚确定性可以降低犯罪，却需要更大规模监控；减少死刑可能增加长期监禁；强调法律明确可能牺牲个案灵活；保障少数权利可能降低多数人的即时安全感。遇到这些冲突，‘公共幸福’若只能事后赞美成功，不能事前说明为什么走这条路，它就只是结果标签。

贝卡利亚的文本并没有把所有冲突都解开，但他至少给出一个方向：公共幸福不是任何快乐总量，而是和一般法律、共同安全、减少无用痛苦、反特权、可预期秩序连在一起的制度幸福。它不是单纯把每个人的快乐分数相加，而更接近“让共同体成员在同一套公开规则下尽可能少受任意支配”。这也是为什么本书把它称为‘自由限权型公共幸福’，而不是简单叫‘最大幸福主义’。

如果这个判断成立，那么自由就不是被降级了。恰恰相反，自由从“唯一目标”变成了公共幸福的结构性条件：没有自由闸门，幸福中心会膨胀成多数人的工程；没有幸福方向，自由又可能只剩互不碰触的消极边界，无法解释为什么还要预防犯罪、改善教育、奖励有益行为。两者不是二选一，而是功能不同。

第49章因此给出第一张清账表：公共幸福是方向，自由是硬约束，必要与比例是审查结构，法律明确与司法受限是制度装置，确定性与及时性是行为路径，预防与教育是上游工程，犯罪减少与社会信任是可能结果。这个层级不是最终真理，却比把所有概念排成一排更能解释贝卡利亚为什么从死刑一路写到教育。

> 带走的话
>
> 一个理论的中心，不是出现次数最多的词，而是在冲突中决定方向的那一层。到目前为止，公共幸福更像贝卡利亚的方向，自由更像不能被轻易穿越的边界；两者合在一起，才构成他真正的制度发动机。

小账

这一章看见了什么？——幸福定方向，自由作硬约束，必要比例作审查，犯罪下降只是结果。

还要追问什么？——公共幸福若只能事后赞美成功、不能事前裁决冲突，就只是结果标签。

第 50 章

第一场正式判决

走到第50章，本书必须第一次停下来判。所谓“判”，不是给贝卡利亚盖棺，而是把前文已经积累的证据暂时压成一组可以被后文推翻的承重判断。没有这一步，后面越写越多，所有发现都会变成散点；有了这一步，后面的每一章才知道自己究竟是在支持、修正还是推翻什么。

第一判：贝卡利亚不是从“国家拥有刑罚权”开始，而是从“国家必须证明刑罚权从哪里来”开始。这个转向是整本《论犯罪与刑罚》的第一扇门。刑罚权不再天然附着于王冠、传统或法官身份，而要追溯到共同体成员为了安全所作的有限让渡。国家的强制因此第一次被迫拿出出生证明。

第二判：贝卡利亚真正改变的不是刑罚的温度，而是刑罚的举证责任。过去容易问被告：你为什么不该受罚？他的结构则不断把问题推回国家：为什么这是犯罪？为什么由刑罚处理？为什么要这么重？为什么必须这么快？为什么一定要用死刑？为什么不能用更轻的手段？这种“国家先说明”比任何单项改革都更深。

第三判：贝卡利亚不是一个只会把人当利益计算器的简单功利主义者。他当然大量使用快乐、痛苦、利益、概率和效果语言，但他同时看见教育塑造判断、秘密告发塑造关系、奴役改变人格、法律过多制造犯罪、制度确定性改变选择。人并非一块静止的算盘；他的文本里已经出现了环境回写主体的多处证据。

第四判：贝卡利亚的契约限权与公共幸福双线不是本书的新发现。当前更有解释力的表述是“公共幸福定方向，自由设硬闸”；本书试图增加的，是用“谁能否决谁、例外门槛多高、例外开在哪里”来检验这一结构（第44章已先作校准）。

第五判，也是最重要的暂判：贝卡利亚把国家权力发生化了，却没有把支撑这套发生的两个观察位置完全发生化。前端，是能够交出自由的个人；后端，是能够从总体上判断必要、比例和公共幸福的立法视角。前者像社会契约的原料，后者像

制度设计的看台。两者都不是绝对先在，但它们的形成条件都没有得到与国家权力同等强度的追问。

这里必须立刻加入两个保底。第一个保底给自由个人。规范法理论经常需要先承认一个人的基本资格，才能防止国家用“你还没有形成真正自由能力”为理由剥夺其权利。一个婴儿、重病者、认知能力有限者，不会因为无法完整进行契约计算就失去人的地位。因此我们不能把“规范上先承认主体”简单判成发现学错误。它可能是保护人的制度选择。

第二个保底给理性立法者。任何立法都需要某种总体视角：如果只允许每个人从自己眼前利益出发，公共规则很难形成。所谓‘看台’未必意味着某个天才站在社会外面，而可能只是一种要求——立法时尽量离开局部激情，比较不同人的利益和长期效果。真正的问题不是能不能有看台，而是看台能否被拉回场内：它的信息从哪里来，它有没有盲区，它能否被公众经验修正，它判断错误时谁能让它改。

这两项保底一加，SDE的批判必须收窄。我们不能再说“贝卡利亚根本不懂主体发生”，因为教育和社会环境章节已经反证；也不能说“他崇拜全知立法者”，因为他反对激情、特权和秘密，本意正是限制任意。更准确的说法是：他已经看到个人与制度都会变化，却没有把‘变化如何进入正当性结构’写成完整机制。

于是第一场正式判决不是宣判有罪，而是划定后文审理范围。接下来要查三件事：第一，社会契约以前的那个‘人’究竟承担的是经验描述还是规范地位；第二，贝卡利亚关于教育、奴役、贫困、荣誉和法律的论述，能否让主体发生真正进入刑罚理论；第三，那个‘明智观察者’是否只是修辞，还是全书事实上依赖的决策位置。

如果这三项检验最终证明贝卡利亚已经给出了足够的回写机制，本书必须撤销“两个观察位置未发生化”的判断。反过来，如果它们持续作为不被追问的入口和出口，新的收回动作就会比‘主体先在’更准确：不是简单先放一个人，而是把整个刑罚制度夹在两个暂时免检的位置之间。

带走的话

贝卡利亚的伟大，在于把国家从“我当然可以罚你”逼到“我必须解释为什么罚你”；他的未完成之处，可能在于解释国家的同时，还没有同样彻底解释‘谁在交出自由’以及‘谁在替所有人判断公共幸福’。

小账

这一章看见了什么？——他改的不是刑罚温度，而是举证责任：每一步都要国家先说明。

接缝在哪里？——交出自由的个人与判断公共幸福的立法视角，都没受到同等追问。

第 51 章

社会契约以前站着谁

贝卡利亚写刑罚权时，有一幅非常清楚的画面：人们为了摆脱彼此侵犯带来的不安全，被必要性迫使，各自把一小部分自由放进共同的“存款”里；这些份额合在一起，形成主权，主权者则成为这笔共同存款的保管者。更关键的是，每个人只愿意交出最低限度——仅仅多到足以使别人也愿意共同保护他。国家权力因此不是整块赠送，而是逐份积累。

这幅画面极有力量，因为它立刻改变了国家和个人的方向。不是个人先属于国家，然后国家仁慈地还给他一些自由；而是个人先拥有自由，国家只能从这些自由中取得必要份额。正因为来源有限，权力的边界才有限。贝卡利亚反死刑、反酷刑、反过度处罚的许多论证，都靠这一顺序得到支撑。

但SDE此时必须问一个看似冒犯常识的问题：社会契约以前站着的这个人，究竟是什么样的人？他已经知道什么叫“自己的自由”吗？已经能比较安全与自由的交换吗？已经有一个稳定的自我边界，知道哪部分属于自己、哪部分可以让渡吗？如果这些能力都已经完成，那么国家的发生被解释了，主体的发生却被留在了镜头之外。

先别把这个追问误解成“没有社会就没有人”。当然不是。生物意义上的人早已存在，贝卡利亚也不需要写一部儿童发展心理学。问题在于：社会契约论中那个能够成为权利单位、利益单位和选择单位的主体，并不是一具身体自动拥有的全部内容。一个人知道自己有权、能预见后果、能把未来安全和眼前自由拿来比较、能理解规则对所有人有效——这些都需要语言、经验、他人和制度。

举一个孩子的例子。三岁的孩子也会说“我的”，也会抗拒别人拿走玩具，但这不等于他已经掌握现代法律意义上的自由、权利和让渡。他要经过无数互动才逐渐知道：我的东西有边界，别人的也有；我今天答应的事明天仍然算数；规则不是

只约束别人；当我不满意时可以提出理由而不是只靠哭闹。这些能力并非某一天从头脑里被发现，而是在关系中一点点长出来。

贝卡利亚的社会契约不是一份真实历史合同，这一点无需挑刺。启蒙时代的契约语言首先是一种规范装置：它问，如果国家权力要被自由平等的人正当化，那么他们最多会同意什么？这类假设的价值就在于限制权力，而不在于证明远古时代真的召开过签字大会。若把它当史实再反驳‘没有证据’，只是打错靶。

真正值得追问的是，即使作为规范装置，为什么必须把个人想成一个已经完成的利益单位？这里有一个制度理由：只有先把每个人立成不能被吞掉的单位，公共幸福才不会直接把人化成材料。贝卡利亚说每个人都会把自己当成各种组合的中心，这句话并不浪漫，甚至带着怀疑人性的味道，却恰好使国家不能假装个人会无偿为了抽象公共利益牺牲自己。

所以“主体先在”具有两面性。一面是理论简化：它把形成自由能力的家庭、教育、阶层、语言和制度条件放在括号外；另一面是规范防线：它拒绝国家说‘你的自我尚未成熟，所以你的自由可以先拿走一点’。如果SDE只强调第一面，就会误伤现代权利结构。发生学不能用“你是被社会塑造的”推出“社会因此可以拥有你”。

这恰好是本卷需要守住的边界：发生解释与规范归属不能混为一谈。自由能力可以是发生的，基本权利却可以被制度选择为不依赖能力水平而平等承认。就像一个人的语言能力是后天长出来的，但他并不因为不会说话就不配被保护；责任能力可以有程度差异，但生命和身体也不能因此任由别人处置。

于是社会契约以前站着谁，答案要拆成两层。规范上，站着一个必须被当作自由平等成员的人；经验上，站着一个正在关系、教育、贫富、身份和历史中形成其判断能力的人。贝卡利亚把第一层写得极强，第二层零散看见，却没有把两层接成一套完整理论。我们的批判应当落在“接口”而不是“有没有主体”。

带走的话

贝卡利亚在社会契约以前放置自由个人，不只是理论偷懒，也是给国家设边界的规范技术。真正未完成的，不是承认个人，而是没有充分说明：一个被无条件承认为权利主体的人，其实际选择能力如何在社会中发生，又怎样反过来参与决定自己愿意接受的公共规则。

小账

这一章看见了什么？——契约前的自由个人既是简化，也挡住国家以「你尚未成熟」拿走自由。

接缝在哪里？——规范上被承认的人与经验中正在长出能力的人，没有接成一套理论。

第 52 章

那个自由的人有没有童年

如果第51章还停留在抽象主体，这一章就把他变回一个会出生、会长大、会害怕、会模仿、会被奖励和羞辱的人。问“那个自由的人有没有童年”，不是为了用心理学取消法理学，而是为了检查：贝卡利亚在自己的文本里，是否已经留下主体发生的证据。答案是肯定的，而且最直接的证据就在他谈教育的地方。

贝卡利亚把完善教育称为预防犯罪最可靠同时也最困难的方法之一。他并不满足于命令式服从，而是注意到经验、感受、后果与习惯如何形成判断。这说明，在经验层面，他完全知道人不会带着完成版的理性从出生走进社会。一个人如何看待规则、痛苦、别人、荣誉和未来，是可以被培养、被扭曲、被制度长期塑造的。

童年把刑罚哲学里一个常被隐藏的问题暴露出来：威慑理论假定人能够把行为与未来后果连起来。可这种能力本身从哪里来？一个人若从小经历的是规则反复变化、惩罚看关系、承诺不兑现、暴力可以解决问题，那么成年以后再告诉他“法律具有稳定确定性”，这句话不会自动进入他的行为系统。制度的确定性不仅是法典属性，也是一种被生活经验反复训练出来的信念。

这解释了为什么贝卡利亚把教育、法律清晰和官员守法放在同一组预防方案里。孩子不是在课堂上单独学会法治的。他看到父母是否守约，老师是否双重标准，管理者是否对强者例外，警察是否按规则办事，法院是否给出理由。每一次经验都在回答一个问题：规则究竟是共同边界，还是强者临时使用的工具。所谓法治教育，最深处不是背条文，而是反复经历规则真的这样工作。

从SDE看，这就是“土”参与自由能力发生。家庭、学校、社区、语言和制度不是主体外面的布景，而是在塑造他怎样识别选择、怎样预期后果、怎样理解别人也是主体。自由并不只是选项数量，它还包括能不能看见选项、能不能理解长期后果、能不能抵抗即时冲动、能不能把自己放进共同规则中。

但这里马上出现危险：既然自由能力需要培养，国家是不是就可以说“为了让你真正自由，我先替你决定”？这是所有教育型政治最危险的拐点。只要统治者声称自己知道什么叫成熟人格，就可能把不同生活方式都归入“尚未被正确教育”。因此，主体发生理论必须和贝卡利亚的限权原则绑在一起：可以创造形成自由能力的条件，却不能把“你还不够自由”当作取消权利的许可证。

一个家庭例子很直观。父母希望孩子学会自主，最简单的办法是替他把所有错误选项拿走：几点睡、交什么朋友、学什么专业、怎么花钱全部决定。短期看，孩子犯错更少；长期看，他没有练习判断。另一种做法是给出年龄相称的边界，在边界内允许选择并承担可修复的后果。后者风险更高，却更可能长出真正的自由能力。国家与公民的关系当然不能直接等同父母与孩子，但这个例子提醒我们：‘为了保护你’和‘让你有能力保护自己’是两种治理逻辑。

贝卡利亚的教育章节之所以重要，正因为它让“自由人”不再完全是纸上假设。他已经意识到，一个社会若充满奴役、恐惧和不确定，人会变得更短视、更残酷、更依赖眼前刺激；自由社会则可能鼓励更长期的兴趣和公共关切。换句话说，自由既是制度前提，也是制度结果。人需要自由来正当化制度，制度又会反过来塑造能否实际使用自由的人。

这形成一个循环：自由主体参与形成法律，法律保护自由，受保护的环境又培养更能承担自由的人。传统契约论喜欢从第一步开始，仿佛自由主体已经站在那里；发生视角则更关心这个循环没有真正的绝对起点。一个人出生在已有语言、家庭与法律之中，长大以后又参与修改这些东西。自由不是从真空出发，而是在一代代回写里获得具体形式。

这是否推翻贝卡利亚？不。恰恰相反，它使他的思想更强。如果把社会契约只理解为一次历史性的自由让渡，它很容易被事实驳倒；如果把它理解为一个不断重做的正当性要求——现有法律必须能够向今天活着的人说明理由，并允许他们通过制度参与修改——那么“自由人的契约”就从一次性起源神话变成持续程序。贝卡利亚在反对祖先意志永远捆绑后人时，已经接近这一方向：法律权威来自活着的共同体，而不是死去祖先留下的不可更改誓言。

因此，童年不是对贝卡利亚的外部补丁，而是把他自己已经看见的教育、习惯和活着的社会接起来。自由人的资格可以被先承认，自由人的能力却必须发生；法律可以保护主体，但保护本身也会参与塑造主体。这比简单说‘主体先在’更准确，也更难。

> **带走的话**
>
> 那个自由的人当然有童年。贝卡利亚已经看见教育和制度会塑造实际选择者，因此他的真正空白不是“人会不会发生”，而是没有把自由的规范资格与自由能力的经验发生写成一个相互回写的循环。

> **小账**
>
> 他看见了什么？——威慑所需的预期能力要在童年里练，规则稳不稳是孩子反复经历出来的。
>
> 还要追问什么？——自由能力需要培养，却不能变成国家以「替你决定」取消权利的许可证。

第 53 章

语言、家庭、贫困、身份有没有参加犯罪选择

一旦承认自由能力有发生过程，刑法就会遇到一个很麻烦的问题：如果人的选择受到语言、家庭、贫困、身份和制度长期塑造，我们还能不能说“这是你的责任”？很多讨论在这里迅速滑向两个极端。第一个极端说，环境影响一切，所以个人没有责任；第二个极端说，只要承认环境影响，就等于给犯罪找借口，所以最好不要谈。两边都把问题做简单了。

贝卡利亚的长处，是他既不把人写成完全脱离利益和环境的纯意志，也不因此放弃处罚。他关心诱惑、机会、刑罚概率、法律确定性、荣誉、奴役、教育，这些都说明行为不是从真空里蹦出来的；但他仍然把法律当作面向行动者的公共规则。这种结构提示我们：解释一个行为怎样发生，与判断一个人是否应承担责任，是相连但不同的两步。

先看语言。一个人怎样命名行为，会改变他看到的选项。把收受利益叫“人情往来”，把欺负弱者叫“开玩笑”，把群体暴力叫“替天行道”，并不会自动改变法律性质，却会改变行为人在行动前怎样理解自己。语言提供分类，分类影响正当化，正当化又可能降低内在阻力。刑法如果只在结果出现后宣布罪名，就容易漏掉行为在群体语言里怎样被一步步正常化。

再看家庭。家庭不是决定犯罪的机器，但它是最早的规则实验室。孩子会在这里第一次经历权威是否有边界、冲突是否只能靠暴力、承诺是否算数、弱者是否可以被羞辱、错误以后有没有修复机会。不同家庭当然不会机械生产不同犯罪，但这些经验会进入一个人的冲突处理方式。说“家庭参与了选择能力形成”与说“家庭替成年人负责”完全不是一回事。

贫困更容易被误用。一方面，经济压力会改变机会结构和风险权衡：一个人眼前资源极端匮乏，未来折扣率可能更高，某些短期收益的吸引力会变大；另一方面，绝大多数贫困者并不犯罪，富裕者也会实施复杂犯罪。所以贫困不是免责公

式，更不是犯罪本质。它是发生条件之一，需要和制度机会、社会支持、执法选择、教育、个人经历一起看。

身份同样如此。一个人不断被当作“危险的人”“坏学生”“有前科的人”，这种外部分类会改变别人给他的机会，也会改变他理解自己的方式。贝卡利亚在讨论被指控者的羞辱、秘密告发和制度名誉时，已经敏感到法律标签具有回写能力。现代社会更能看到，刑事记录、平台标签、风险等级一旦固定，可能把一次行为延伸成长期身份。

从这里回看贝卡利亚的责任观，会发现一个重要优点：他反对把‘灵魂有多坏’直接当刑罚尺度。因为人格判断越宽，环境与偏见越容易偷渡进来。法官说某人‘天性危险’，可能其实是在读取他的阶层、口音、外表或生活方式。贝卡利亚宁愿把法律拉回行为和公共损害，这在今天仍然是一道防止身份刑法的堤坝。

可单纯退回行为也不够。行为之所以可责，仍需考虑认知、控制、意图、胁迫和能力。真正成熟的责任理论必须同时做两件事：第一，不把人简化为环境的产物，保留行动者能够回应理由的地位；第二，不把行动者想成没有历史的孤岛，承认其选择空间和能力受到现实条件影响。两者缺一，刑法都会失真。

可以把责任理解成一扇门。门外是发生解释：这件事怎样一步步长出来；门内是规范判断：在这些条件下，行为人是否仍有足够的理解和控制能力，是否有可行的替代路径，国家因此可以要求他承担多少责任。发生解释不是为了替人开脱，而是为了把门的位置放准。不了解发生条件，就可能把根本没有真实选择空间的人和拥有大量选择却仍故意伤害他人的人放进同一格。

这也会反过来限制预防政策。假如一类犯罪与贫困、教育缺失、药物依赖或群体文化高度相关，国家当然可以改善这些条件；但不能因此把处于这些条件中的人直接当成潜在罪犯。风险因素不是罪责。一个统计关联只说明某种发生概率变化，不等于某个具体人已经做了该被处罚的事。贝卡利亚要求刑罚面对已发生的法律事实，这条边界在数据时代反而更重要。

所以，语言、家庭、贫困、身份都“参加”了选择，却都不能单独“代替”选择。主体不是一个没有土的点，也不是被土完全推动的木偶。更准确的图像是一条洪流：个人的欲望、能力、规则、关系和机会不断纠缠，最终在某个具体时刻形成行动；行动发生以后，又会回写身份、关系和制度。刑法只能截取其中一个结算点，但不应误以为结算点就是全部发生。

> **带走的话**
>
> 承认犯罪选择有土，不等于取消责任；真正的进步，是把责任从‘一个坏灵魂突然作恶’改写成‘一个能够回应理由的人，在具体发生条件中作出可归责的行动’，并让刑罚只对这个结算点承担它能够证明的那部分强制。

小账

这一章看见了什么？——语言、家庭、贫困、身份参加选择却不代替选择，解释是为放准责任之门。

接缝在哪里？——风险因素不是罪责，统计关联不能把身处不利条件的人预先当成罪犯。

第 54 章

如果一个人的自由也是长出来的

现在来到最容易被误解的一章。说“自由也是长出来的”，听上去仿佛是在否认自由：既然我的选择受家庭、语言、教育、制度和身体影响，那我还算自由吗？如果每一个决定都能追出一串来历，自由是不是只剩幻觉？这其实把“有来历”和“没有自由”错误地画成了同义词。

发生视角恰恰可以给自由另一个定义：自由不是没有条件，而是在条件中形成新的路径。一个完全没有限制的世界并不会自动产生自由，反而可能没有任何可识别的选择；真正的自由发生在约束、目标、能力和环境之间，一个人能够看到不止一条路，理解它们的后果，并参与决定走哪一条。自由不是从外部原因中逃出去，而是在原因网络里获得可回应、可修正、可创造的行动空间。

这与贝卡利亚有一个意外的连接。他从未把自由写成随心所欲。社会之所以需要法律，就是因为每个人都把自己放在利益组合的中心，如果人人都想把共同约束只施加给别人，社会无法稳定。自由因此一开始就与共同规则纠缠：我保留自己的大部分自由，是因为我也接受别人保留同样的自由，并同意一小部分限制来防止相互侵犯。

所以，自由的发生并不会自动摧毁社会契约，反而可以把契约解释得更现实。不是一群已经完成的人在某个瞬间签下合同，而是一代代人在已有制度中成长，学习规则、争论规则、修改规则；制度又通过教育、执法和公共生活塑造下一代人。每一代人都继承一个已经成形的秩序，同时又有机会把自己的经验写回去。

在这幅图里，‘最小自由让渡’也获得新的意义。它不再只是一个远古起点，而成为每次国家扩张强制时都必须重新通过的检验：这项限制是否真是共同生活必要的？是否有更轻替代？受影响的人有没有表达和申诉？实施后的结果是否反过来要求修法？如果回答是否定的，旧有授权就不能因为曾经存在而永久有效。

这与贝卡利亚反对祖先意志永久绑架后人形成呼应。他强调法律来自活着的社会，而不是一份由祖先签署、后代只能服从的遗嘱。只要把这一点再向前推，自由的契约就天然带有回写结构：今天的人不能把明天的人永远锁死，因为明天的社会也有自己的经验、风险和判断。法律需要稳定，却不能把稳定写成终结。

这会改变我们看罪刑法定的方式。罪刑法定当然要求行为发生前已有规则，否则自由无法规划；但规则事先存在并不等于规则永远正确。真正的法治应该同时拥有两种时间：对个案而言，规则必须在前，不能事后加罪；对制度而言，社会经验又必须在后面不断回写规则。一个案件不能靠临时改法处理，但一串案件可以成为修法的理由。

于是自由发生与法律确定性并不矛盾。个案层面的确定性保护人免受临时权力，制度层面的可修正性保护社会免受旧规则僵化。贝卡利亚非常强地写出了前者，对后者有所暗示，却没有系统展开。SDE可以补的，不是把法律永远流动化，而是把‘什么时候必须稳定、什么时候允许回写’分层。

同样，责任也要分层。一个人在具体案件中如果具备足够理解和控制能力，就应对行为承担责任；这不妨碍我们同时承认他的能力形成受环境影响，并在刑罚之外改造环境。两条线可以同时成立：个体责任回答“这次行为谁来承担”；发生治理回答“怎样减少同类行为再次长出来”。前者不能替代后者，后者也不能取消前者。

这正好重新解释贝卡利亚为什么把预防放在处罚之前。若自由是一个会成长、会受制度塑造的能力，那么最好的刑法并不是等一个人做完坏事再用痛苦纠正，而是让社会条件更有利于形成可预期、可合作、能理解他人边界的行动者。教育、知识、稳定规则、反特权于是都不是刑法的旁枝，而是自由能力的发生条件。

但SDE仍需自我限制。我们不能因为说‘自由会发生’，就认为可以测量谁“真正自由”并据此分配基本权利。发生描述越精细，权力越可能借它分类人。算法说某人自控能力低、未来风险高、环境恶劣，如果据此提前加重限制，就把发生理论变成新的身份刑法。贝卡利亚最小让渡原则恰好可以反过来约束这种危险。

带走的话

自由有来历，不等于自由不存在。更准确地说，法律应当无条件承认人的基本自由资格，同时承认实际自由能力在关系和制度中发生；国家可以改善发生条件，却不能以“替你形成自由”为名先取消你的自由。

小账

这一章看见了什么？——自由是在条件中长出新路径，最小让渡成了每次扩张强制都要重过的检验。

还要追问什么？——个案要规则在前，制度要经验回写，何时稳定、何时回写须分层说明。

第 55 章

“先放人，再限权”还能不能成立

前文我们曾把贝卡利亚的收回动作概括成“先放人，再限权”：为了限制国家，他先在理论上放置一个拥有自由的个人，然后由这些人让渡最少自由形成国家。经过第51—54章的攻击，这个判断已经不能原样保留。因为原典中的教育、奴役、知识、荣誉、制度确定性都表明，贝卡利亚并非看不到经验主体的形成。若继续说他把‘人’完整地放在发生之前，就是过判。

可这个动作也没有完全消失。它应当被收窄成：先承认一个规范主体，再从这里限制权力。　这里的‘先’不是时间顺序，而是正当性顺序。国家要想取得刑罚权，必须先把每个人当作有自身利益和自由边界的成员，而不是先把人定义成国家资源。这个规范先行不是漏洞，而是现代权利结构的重要条件。

那么，真正还值得解构的是什么？不是“为什么先承认人”，而是“这个被承认的人怎样进入公共判断”。如果每个人都有自己的利益、经验和自由边界，公共幸福如何从这些不同位置形成？是把所有偏好相加吗？由立法者替他们计算吗？通过公开讨论吗？通过代表制度吗？贝卡利亚没有给出今天意义上的民主程序理论。他更关心把旧法从激情和特权中救出来，因此把希望寄托在一般法律和理性观察。

这让‘先放人’与‘把立法者请上看台’开始合并。台下是一群被当作自由单位的人，台上是试图把他们的利益组织成公共幸福的立法视角。真正缺失的不是两端，而是中间那条上台与下台的楼梯：个人经验怎样进入立法，立法结果怎样回到个人，少数人的反对怎样被听见，失败政策怎样被数据与生活经验纠正。

如果没有这条楼梯，社会契约就容易只在开场时出现一次。人民把自由存入共同账户，之后由法律替他们说话。贝卡利亚已经强调法律来自‘活着的社会’，这是非常重要的线索，但他没有把这种持续授权设计成完整政治机制。这与他的写作对象有关：他写的主要是犯罪与刑罚，不是一部宪政民主论。我们不能因为一本刑罚论没有完成全部政治哲学，就把它判成失败。

但专著的任务不是替对象遮住边界，而是精确标出边界在哪里。贝卡利亚的模型在‘权力从哪里来’上非常发生，在‘法律怎样持续吸收被治理者经验’上则较薄。它像一座桥，入口和出口都很清楚，中间支撑还不够：个人是来源，公共幸福是目标，可从无数具体生活到总体判断之间，缺少一套细密的回写程序。

这也是为什么‘必要性’会变得危险。必要性本来是限制国家的词，但如果判断必要的主体和目标都不透明，它又能成为扩权的词。一个立法者说‘为了公共安全必须如此’，形式上完全符合贝卡利亚的语言；真正区分限权和扩权的，是他能否公开说明替代方案、证据、不确定性、影响分布，以及能否在新经验出现后修改判断。必要性必须是一种可被反驳的程序，而不是权力的形容词。

由此，本书的收回动作需要第一次正式改名。‘先放人，再限权’不宜再作为总收回，因为它把一种合理的规范策略误写成理论缺陷。更准确的候选应当是：“搭了看台，少了楼梯。” 贝卡利亚需要一个观察公共幸福的立法视角，这是制度理性必须；问题是从众人的生活经验通往这座看台、再从看台回到众人的持续回路，没有被充分展开。

这个名字比‘理性立法者中心主义’更公平。它承认看台有必要：立法不能被每一次局部激情拖着跑，必须有总体视角、长期视角和比较能力；同时又要求看台不是悬空的：它必须接入民众经验、异议、事实反馈和修法程序。没有看台，公共规则碎成局部争夺；没有楼梯，看台变成免检高地。

这也使我们重新理解自由与公共幸福的关系。自由不是台下每个人只守自己的边界，公共幸福也不是台上一个人替大家做总账。真正成熟的制度，应当让个人自由经验进入公共理由，让公共决策接受个人权利限制，再让实施结果回到后文规则形成。这样一来，自由与幸福不再只是两个静态变量，而是一套持续互相校正的回路。

第55章阶段改判：‘先放人，再限权’从总批判降级为规范策略描述；新的总收回候选改为‘搭了看台，少了楼梯’。贝卡利亚成功取消了君主任意，却尚未充分说明公共理性怎样持续从被治理者的经验中长出来、再被这些经验纠正。

带走的话

「先放人，再限权」只是规范策略；贝卡利亚搭了看台，却少了让被治理者经验上下往返的楼梯。

小账

这一章看见了什么？——「先放人」是规范策略而非缺陷，缺的是台下经验通往看台的楼梯。

接缝在哪里？——判断主体与目标若不透明，必要性会从限权的词变成扩权的词。

第 56 章

权力被发生化，主体尚未被完全发生化?

第56章必须给前文一直使用的定位式做一次终审。最早我们写：‘权力被发生化，主体尚未被发生化。’后来教育章节逼我们改成：‘规范主体与经验主体的接口不足。’前文又出现公共幸福中心和‘把立法者请上看台’。现在经过第49—55章，旧定位式已经太粗，需要换成一个更能容纳这些反证的版本。

◆ 推翻条件第一次真正执行：贝卡利亚后来有没有走下“看台”

前文把“搭了看台，少了楼梯”当作候选诊断，意思是贝卡利亚更擅长让理性立法者观察社会，而没有充分制度化被治理者经验怎样持续返回立法。按照本书自己立下的规则，这个判断不能只靠《论犯罪与刑罚》一本文本，必须拿贝卡利亚后来的实践生涯去撞。

撞击的结果是：原判断需要进一步降级。Treccani的贝卡利亚传记显示，他1771年进入米兰的最高经济行政机构，随后长期处理货币、粮食等公共行政问题；1791年又进入民刑司法改革委员会。也就是说，他不是只坐在书斋看台上的理论家，而确实进入行政机器处理制度修正。即使这些职务不能自动证明他已经建立现代意义上的公众反馈程序，也足以推翻“他本人从未走下看台”的强判断。

还要看到时代位置。《论犯罪与刑罚》的“致读者”本就把主要劝说对象指向承担公共福利职责的开明统治者与改革官员。哈布斯堡治下伦巴第的改革很大程度采取自上而下的行政形式。因此，“楼梯不足”首先是1764年文本和当时政治结构的限制，不应轻易写成贝卡利亚个人缺陷。把今天成熟的民主参与装置倒推为他的必修课，会落入本书自己反对的后见之明。

所以新的、较窄判断是：1764年文本把经验—立法反馈更多寄托于观察者和改革官员，而没有发展出一套稳定、公开、制度化的反馈架构；贝卡利亚后来的行政

实践又表明，这不是“拒绝反馈”，而是尚未形成现代制度语言。 “搭了看台，少了楼梯”从人格与总体理论判断，正式降级为文本结构和现代延伸诊断。

◆ 晚年文本复核：1792年的死刑意见没有把例外门槛放低

检验不能只停在传记摘要。1792年，贝卡利亚与 Paolo Risi、Francesco Gallarati Scotti 在刑事改革讨论中形成关于死刑的少数意见（Voto）。这份晚年材料继续反对通常状态下的死刑，同时仍保留一种极端例外：足以颠覆国家的非常态犯罪。这说明1764年死刑章中的高门槛结构并未在晚年被简单取消。

这个结果让F2得到更精细的状态：自由并没有被晚年贝卡利亚降格为可随时交换的公共幸福工具；但它也不是无例外的绝对禁令。真正稳定的是“例外必须付出极高理由成本”。同时，晚年意见更强调死刑的不可逆与替代刑罚，这使本书的“例外门槛”诊断第一次能够在同一作者两个时间点之间进行纵向比较。

它也补强F3：贝卡利亚不只是后来进入行政机关这一传记事实，而是在改革委员会的具体规范讨论中继续参与制度修正。因此“搭了看台，少了楼梯”更不能作为人物总判，只能保留为1764年文本以及现代制度化反馈视角下的延伸诊断。

◆ 门槛不只有高度，还有位置

1792年意见书证明例外门槛没有被放低，但同一份材料还有另一种更冷的读法，不能回避。意大利刑法史学界有学者指出，真正的废死者是1786年在托斯卡纳一举废除一切死刑的利奥波德，而不是贝卡利亚（Padovani， 2014）；也有学者讥讽，一面主张废死，一面为足以颠覆国家的罪犯保留死刑，与“废死”的名号并不相称。批评直接落在本书的“例外门槛”上：门槛再高，门也开着，而且开在政治犯罪这一格。

回到原典。1764年的死刑章说，处死一个公民只在两种理由下才可能被认为必要。第一种，是他即使被剥夺自由，仍拥有足以威胁国家安全的关系与力量；这种情形出现在国家恢复或丧失自由之际，出现在混乱取代法律的无政府时期。第二种，是他的死成为阻止他人犯罪的唯一真正约束；而贝卡利亚随即论证，这一条件

在常态下并不成立。1792年意见书保留的，正是第一种里的“足以颠覆国家”。两处的共同点是：例外不是按行为种类开的，而是按对国家存续的威胁开的。

这就给诊断工具补上了第三个问题：例外开在哪里，由谁判定。对普通犯罪，包括盗窃、伤害乃至普通杀人，贝卡利亚的门封得很死，1764年和1792年都没有松动。对“足以颠覆国家”的人，门开着，而判断一个人是否足以颠覆国家的，恰恰是自认为受到威胁的国家本身。第9章问过“谁有权宣布必要”；这个问题在全书最硬的一道闸门前又回来了，而且回到了它最危险的位置。

公平地说，贝卡利亚的例外以实际拥有的关系与力量为条件，而不是以思想或言论为条件；他也明确说，在法律安定、民众拥护的政府之下，这种必要并不存在。这使他的例外远比同时代许多刑法典狭窄。但结构上的缺口仍在：例外的判定者就是利害关系方，文本没有为这一判定另设审查。

因此F2改记为“部分触发”。“自由限权型公共幸福中心”的判断保留，但要补一句：这道自由硬闸对普通犯罪几乎封死，在国家存亡处留了一道门，门的钥匙在国家手里。这也说明“位置”不是装饰性的第三参数：只看门槛高度，贝卡利亚与罗尔斯相去不远；加上位置，二者的差别才显露出来。

◆ 锁钥同手：这道闸的钥匙在谁手里

先说一件小区里常见的事。业主公约写着：物业不得擅自进入业主家中——这是一把锁；紧接着一行小字：紧急情况除外——这是一把钥匙。问题在于，什么算紧急，由谁说了算？如果由物业自己判断，那么这把锁锁的是物业，钥匙也在物业手里。平日里它也许从不越界，但公约本身已经把“能不能进门”的最后一句话，交给了公约要约束的那一方。

公司规定“不得查看员工邮件，涉及公司安全的除外”，判断是否涉及安全的是公司；学校规定“不得翻查学生书包，存在安全隐患的除外”，判断有没有隐患的是学校。这类条款太常见，以至于我们很少注意它的结构：锁和钥匙在同一只手里。本书把这种结构叫作“锁钥同手”。

把第44章那组问题放回来：谁上锁、锁多硬、钥匙挂在哪、谁握着、门朝哪边开。边沁没有锁，一切都在效用账里；康德的锁极硬，唯一的钥匙在主权者手里，但门朝宽宥开；罗尔斯的钥匙不在任何一方手里，而挂在事先写明的一般社会条件上；诺齐克的锁极硬，钥匙却没有写明归谁。贝卡利亚的锁由每个人交出的最小自由铸成，对普通犯罪几乎焊死；钥匙在国家手里；门朝更重的强制——死刑——开。在这几位里，只有他同时占了三样：一把很硬的锁，钥匙在被锁的一方手里，门朝强制开。

于是本书对贝卡利亚的中心判断可以写成一个更锋利的句子：贝卡利亚的限权体系，不是“自由中心”，也不只是“公共幸福定向、自由设闸”，而是一道锁钥同手的闸——锁由自由铸成，钥匙却留在它要锁住的国家手里。

压成全书的一句判词，就是：**他逼国家为手里那把刀出示出生证明，却把最硬一道闸的钥匙，留在了国家手里**。

这个句子把全书三处散开的接缝收成了一处。第9章问“谁有权宣布必要”，那是钥匙在日常刑罚里的小号版本；本章的“足以颠覆国家”由国家判定，是同一把钥匙在最硬一道闸前的大号版本；第72章的风险分数，是钥匙被交给一个模型，由它替国家说出“必要”。必要的宣布权、例外的判定权、风险的定级权，是同一个问题的三张脸。第78章会说“把临时钥匙焊死在锁里”，那是钥匙被冻住；这里问的是钥匙握在谁手里。同一把钥匙，两种死法。

还有一处更耐人寻味。第64章的扩样编码显示，贝卡利亚一遇到“权力归谁”的问题——法官能不能解释法律、贵族能不能另罚、庇护所能不能在法律之外——就先问“凭什么权利”。这位最会追问“钥匙归谁”的作者，恰恰在自己的死刑例外条款处没有追问。这不是道德指责：在1764年的米兰，除了国家，他也找不到第二个可以握钥匙的人。但结构上的空位就在这里。

◆ 与施米特、罗西特和“自己不做自己的法官”划界

“谁判定例外”不是本书的发现。施米特1922年《政治的神学》开篇就说：主权者就是决定例外状态的人。阿甘本后来把例外状态写成现代治理的常规范式。更

古老的法谚“任何人不得做自己案件的法官”，早就禁止利害关系方自己裁判自己。

还有几位更近的邻居。罗西特1948年的《宪法独裁》给紧急权力立过一条准则：宣布进入紧急状态的人，不应是将要行使紧急权力的人。费雷约翰与帕斯奎诺2004年把紧急权力分成“新罗马模式”与“立法模式”，前者的要点正是把宣布者和行使者分开；阿克曼同年的《紧急宪法》则用逐级提高的议会多数，限制紧急状态的延续。国际法里，关贸总协定第21条允许缔约方采取“其认为”保护基本安全利益所必需的措施，这种自己说了算的安全例外，直到2019年世贸组织“俄罗斯—过境运输案”专家组报告才被确认仍须接受客观审查。至于“方向”，刑法里早有一种不对称：不利于被告的类推被禁止，有利于被告的类推通常允许。欧洲人权法里还有一个现成的制度实例：《欧洲人权公约》第6议定书（1983年）废除和平时期死刑，却允许各国以本国法律为战时留门，第13议定书（2002年）才把这扇门也关上。这道门的钥匙在各国自己手里，门朝死刑开——正是本书所说的结构。本书没有在这组条约上做过检验，这是下一步最直接的试验场：若复合读数说不出单看“谁握钥匙”说不出的东西，它就只是重述。

这些占位者一摆出来，“锁钥同手”就必须说清自己不是什么。它不是“自我裁判”的又一个名字——那一层已被法谚和罗西特占住；它也不是“谁决定例外”的主权理论——那一层属于施米特。它是一个复合读数：钥匙的位置乘以门的方向，用来给一套限权理论定位。单看位置，康德和贝卡利亚同属“自己握钥匙”；单看方向，刑法的有利与不利类推只管解释法律，不管限权理论里的例外条款。两者相乘，才读得出贝卡利亚那一格：限权理论最硬的一道闸上，钥匙在被锁者手里，而门朝更重的强制开。

所以本书在这里能保留的增量只有两件：一是把这个复合读数当作思想史诊断工具，它已由第44章的盲编码部分复核；二是它给出的一个可以输的历史预测，下一节就去检验。若两件都站不住，“锁钥同手”应降为上述占位者的组合重述。

◆ 检验：死刑是从哪扇门回来的

如果一道限权闸的钥匙在被锁的一方手里，那么当闸被重新打开时，最先打开的应当正是这把钥匙管着的那扇门。落到死刑史上，预测是：凡由行政一方主导恢复死刑，死刑应先从“国家安全”这扇门回来，再向普通犯罪扩散；凡恢复的压力主要来自民间——例如私刑泛滥之后由议会或公决恢复——死刑应直接从普通犯罪这扇门回来。失败条件在编码之前写定：在行政主导的恢复中，若普通犯罪先于国家安全罪回来，或两者同时回来、看不出先后，预测即告失败。

为了避免只挑顺手的案例，本版请一位不知道本书假设的编码者（同样是大模型实例），按统一口径检索1786—1970年间“和平时期废除死刑、后又在法律上恢复”的事件，逐条编码“谁主导恢复”和“首批恢复的罪名属于哪一类”，并附来源与把握度。共得36件，其中4件把握低；全表见书末附录。口径排除了战时废除的案例，例如意大利1944年废除、1945年即以临时政府法令为抢劫、勒索、绑架等普通犯罪恢复死刑；若按更宽的口径收入，它是又一个行政主导、普通犯罪先回来的反例。检索也不可能穷尽：36件是编码者找到的全部，不是历史上的全部。汇总如下：

谁主导恢复	首批只有国家安全类	首批只有普通犯罪类	两类同时	合计
行政（君主、摄政、政府紧急令、独裁政府）	9件：托斯卡纳1790、奥地利1795、意大利1926、挪威1941、荷兰1943、苏联1950、阿根廷1956、古巴1959、巴西1969	1件：奥地利1933	7件（其中把握低2件）	17
立法或民意（议会常规立法、公决）	1件：丹麦1945（惩处通敌）	15件：美国11个州、瑞士1879、秘鲁1860、新西兰1950等	0	16
混合、难判	1件：葡萄牙1916	1件：锡兰1959	1件：西班牙1934	3

按事先写定的失败条件，结果是部分触发。行政主导的17件里，首批只有国家安全罪的9件与预测一致；奥地利1933年正好相反——陶尔斐斯政府绕开议会，以紧急令先为谋杀、纵火和恶意毁坏财物恢复即决审判的死刑，1934年才扩及暴乱与爆炸物罪；另有7件两类同时恢复，看不出先后。只计把握中等以上的15件，是9件一致、1件相反、5件同时。“先从国家安全这扇门回来”这句话，在三分之一强的行政主导案例里站不住。

奥地利一例还有一个不能拿来自救的细节：它首批的“纵火”“毁坏财物”，实际针对的是纳粹的爆炸袭击。握钥匙的一方可以把政治威胁写成普通罪名——这个现象值得单独研究，但本书不用它来挽救预测。按法律文字编码是事先定下的口径，事后改口径，就是第82章所说的“永不认输”。

不过，数据里有另一个非常整齐的不对称，而且更贴近“锁钥同手”的本义。把握中等以上的案例中，行政主导的15件，首批罪名里含有国家安全类的有14件；立法或民意主导的15件，只有1件（丹麦1945年惩处通敌，属于战后清算）。钥匙在行政手里时，国家安全这扇门几乎总在首批打开的门里；钥匙不在行政手里时，这扇门几乎从不打开。瑞士1848年起就把政治犯罪写进宪法禁令，是后一种的极端：钥匙根本不在行政手里，1879年死刑只能从普通犯罪那扇门回来。

必须说清楚：这个“包含”版本是看了数据之后才形成的，不能算通过了检验，只能算一个新假设。本书把它记为F11′：在1970年以后的恢复事件——一个本书尚未查看的样本——里，若行政主导的恢复首批不含国家安全类罪名的比例超过三分之一，或立法主导的恢复首批含国家安全类罪名的比例超过三分之一，F11′即告失败。

还有四点限制。第一，编码者是大模型实例，人类编码者的复核仍待做；“谁主导”在议会制下尤其难判，新西兰、锡兰两例就是如此。第二，托斯卡纳1786年的法典并没有贝卡利亚式的例外条款，死刑照样先从政治门回来。这说明钥匙不是贝卡利亚的文本造出来的：只要紧急状态由行政判定，钥匙在实践中一直在那里。贝卡利亚的条款没有造出这扇门，它做的是另一件事——把这扇门写进一套反死刑理论，让它显得合理；叶卡捷琳娜二世1767年的《圣谕》吸收贝卡利亚的反死刑论

证，同时照搬“仍有能力煽起公共动乱者”的例外，就是例证。第三，“国家安全”在不同政体中的边界差别很大，本次按法律文字编码，口径偏窄。第四，行政主导的恢复大多发生在政变、战时或独裁之下，F11′测到的也许是政体与战争，而不是钥匙的位置；扩样时必须按政体与战时分层。

所以F11记为“部分触发”：“先国家门、后普通门”的顺序预测失败了三分之一强；“锁钥同手”的历史预测力降为待检的F11′。它作为诊断工具仍然保留——那一部分由第44章的盲编码支撑，不靠这张表。

用三方程重写这个问题：制度显露S由政策路径D与社会纠缠E共同形成；政策实施后的新S必须回写D和E，才可能修正原设计。123原理的第三步正是这里的关键——不是“立法者看见社会就够了”，而是执行结果能否改变下一轮立法路径与参与条件。去掉SDE术语，我们仍可把它叫政策评估、行政反馈和制度学习；SDE只在把三者放进同一互生模型并要求回写不缺席时保留一小块增量。

先保留它正确的一半：贝卡利亚确实把国家刑罚权发生化了。他没有把刑罚权当作王权天然属性，而是把它追溯到必要性、自由让渡、共同安全与一般法律。主权不是凭空的实体，而像由每个人贡献的一小部分自由组成的共同存款。仅这一点，就足以使刑罚权从‘本来就在’变成‘必须说明怎么来’。

再看旧式错误的一半：说‘主体没有发生化’不准确。贝卡利亚明确谈教育，谈奴役与自由怎样塑造人的性情，谈法律确定性怎样改变行为，谈秘密告发怎样制造猜疑，谈知识怎样改变社会，谈奖赏怎样引导行动。这些都说明他知道经验中的人会被环境改变。主体不是完全静止。

那么缺什么？缺的是主体发生与权力正当化之间的制度接口。一个人怎样形成判断，与国家有没有权处罚他，不能直接互推；但二者又不能永远断开。若人的能力和偏好会被制度塑造，那么立法者就不能把当前偏好简单当作自然数据；若国家又有能力通过教育和制度改变人，那么它更必须说明这种塑造的边界。发生事实越进入治理，正当性程序反而要越严格。

同时，立法者的位置也必须放进定位式。贝卡利亚需要从局部利益中抽身的观察视角，否则无法谈公共幸福；可任何观察者都站在某个时代、知识和制度位置上。十八世纪的‘冷静人性观察者’今天看来并不是从社会外面看社会，而是依靠当时的统计、经济学、道德心理和启蒙信念。看台本身也有土。

因此新的定位式可以写成：贝卡利亚把刑罚权从先在权力改写成“自由成员—必要性—公共幸福”的制度发生，却把两处关键回写留得不足：一处是经验主体怎样持续进入规范授权，一处是立法判断怎样持续接受被治理者与实施结果的反向修正。　这比‘主体先在’更精确，也更能解释为什么他的许多原则直到今天仍然有效，而某些制度想象又显得过于整齐。

用SDE三方程的大白话说，贝卡利亚看见了显露不是天生的：刑罚制度可以重造；他也看见了路：确定性、及时性、比例和预防会改变行为序列；他甚至多次看见土：教育、奴役、荣誉、法律传统和社会关系都会推着人走。真正薄的地方，是回写被分散在各章，没有成为统一机制。法律怎样被结果改，立法者怎样被经验改，公共幸福怎样被少数人的受苦改，仍主要靠‘明智立法’而不是明确制度化。

降级之后，‘搭了看台，少了楼梯’仍可保留为一幅1764年文本的结构图，但不再是对贝卡利亚的总收回动作。它既能解释贝卡利亚的强项，也能解释边界。看台代表他超越局部激情、追求一般法律和公共幸福的努力；楼梯不足则解释为什么法官被要求尽量机械、立法者却承担过多判断，为什么社会成员作为权力来源很强，作为持续参与规则重写者却较弱。

这幅图同样不是最后判决。下面三组反证中，第一组已由本章开头的传记与1792年材料部分执行，其余两组仍待检验。第一，若进一步研究贝卡利亚的政治经济学讲义和行政工作，发现他对信息、利益反馈与政策调整有更丰富机制，那么看台并不悬空；第二，若他的‘活着的社会’概念能够支持持续授权而非一次契约，那么楼梯比我们现在看到的更完整；第三，如果他的写作目标本来只限刑罚正当性，不负责政治参与理论，那么要求他补完楼梯可能属于越界审判。

所以本书现在最重要的不是把新判词焊死，而是把后文证据方向写清。真正的SDE解构不应该越写越肯定，而应当越写越知道自己可能错在哪里。一个判断如果没有可推翻条件，它就正在变成自己最想反对的那种‘发现’。

到这里，我们也能看清前文的路线为什么不断变化。前文从刑罚权出生证开始；前文发现必要性和法律解释的接缝；前文从酷刑看到工具会制造事实；前文从威慑和死刑建立权利—必要双钥匙；前文看见法律会回写社会；前文由教育和公共幸福改判中心；前文则把主体与立法者一起拉回发生过程。每一次改判都不是前面白写，而是对象在逼方法缩小自己的免检区。

这也正好回应本系列的总原则：今天的人把昨天的发现解构为发生，又很容易把自己的发生写成新的发现。我们如果现在宣布‘贝卡利亚真正的秘密就是搭了看台少了楼梯’，明天同样会有人问：为什么必须用‘回写’衡量他？为什么SDE自己的三方程是合适的尺？为什么公共参与一定是更好的楼梯？这些问题不能被一句‘SDE只是工具’打发。工具也要接受使用结果和反例的审查。

因此第56章最后不写句号，而写一个逗号。贝卡利亚最重要的遗产不是某一条刑罚轻重公式，而是逼国家持续说明自己。SDE若真要接住他，就不能只逼贝卡利亚说明，还要逼自己的判断说明：你依据什么材料，排除了什么替代解释，在哪些证据出现时愿意撤回。只有这样，‘发生’才不会变成新的免检词。

◆ 理论深化　两个被预先安置的人

经过第49—56章，前面的零散疑点可以压成一幅图：贝卡利亚的刑罚秩序中，国家本身不再是先在者。它从自由份额、必要性和共同安全中产生；可是国家发生的两端，各站着一个相对稳定的人物。入口处，是被规范上承认为自由平等、能够拥有自身利益的个人；出口处，是需要把大量局部利益组织成公共幸福的理性立法者。前者防止共同体把人吞掉，后者防止共同体碎成彼此冲突的即时欲望。两个人物都有制度理由，问题只在两者之间如何持续互通。

如果只保留入口处的个人，不要看台，公共生活容易退成偏好加法：谁声音大、谁人数多、谁当下愤怒，规则就跟着走；如果只保留看台，不要入口处的个

人，公共幸福又会变成替别人做主的宏大工程。贝卡利亚最可贵的地方，是他已经用最小自由让渡把入口守得很紧，又用公共幸福把制度拉向总体效果。最薄的恰好是中间：怎样把分散经验变成公共理由，再把公共决定开放给后文经验。

因此，本编没有把‘两个预先安置的人’判成两块错误的磐石。更准确的说法是：它们是贝卡利亚为解决两个真实难题而搭的脚手架。个人脚手架解决‘国家从哪里得到权力’，立法者脚手架解决‘公共规则怎样超越局部激情’。SDE真正要问的是，当房子已经建起来以后，脚手架有没有被误当成地基；也就是我们是否忘了个人能力有发生、立法理性也有发生。

这一点对现代刑法特别重要。今天的数据治理让‘看台’变得前所未有地高：国家可以通过数据库、统计模型、风险预测看到过去立法者不可能看到的总体规律；同时，数字身份也让‘个人’变得前所未有地可计算。技术似乎既能更了解台下的人，也能让台上的判断更精确。可精确并不自动等于正当。看得更多，只意味着举证责任更重：模型如何形成、谁被遗漏、错误如何纠正、哪些自由即使能换来收益也不能交易，都必须更透明。

所以，贝卡利亚在AI时代最值得继承的不是‘理性算得更准’，而是‘强制必须不断说明理由’。机器越能计算，越不能让计算本身成为理由。所谓公共幸福如果最终只剩一个目标函数，所谓最小自由如果最终只剩参数约束，那么人的异议、经验和新事实就会被压成数据输入。真正的楼梯，就是让这些输入有能力改变目标函数本身，而不是只在既定模型里调权重。

◆ 压力测试　六个问题把“看台—楼梯”推到极限

压力测试一：立法者掌握更好的数据，是否就可以少听公众意见？不能。数据能说明发生频率、分布和相关性，却不能自己决定哪些代价应当由谁承担。犯罪率下降5%是否值得全民位置追踪，不是统计学单独能回答的问题。事实进入规范判断，还需要公开价值理由。

压力测试二：如果公众强烈要求重罚，理性立法者能不能拒绝？可以，而且有时必须拒绝。贝卡利亚的全部改革本来就是在对抗激情化惩罚。看台的必要性恰在

于不让即时愤怒直接变成刑罚。但拒绝不能只说‘你们不理性’，而要公开说明比例、效果、误判和长期后果，让拒绝本身也接受公共检验。

压力测试三：如果少数人的生活经验无法在多数数据中显现怎么办？这正是楼梯存在的理由。平均值可能改善，某一小群人却持续承担过度搜查、错误识别或污名。公共幸福若只看总量，会看不见分布。真正的公共必须允许边缘经验进入制度修正。

压力测试四：如果一个人的选择能力被环境显著削弱，是否自动减责？不能自动。环境事实需要通过具体责任规则进入案件，例如是否影响认识、控制、胁迫或可期待性。否则‘贫困’‘成长环境’会变成新的身份标签。发生解释要细，法律归责要具体。

压力测试五：如果法律长期稳定却已经明显过时，是坚持确定性还是允许法院解释？贝卡利亚会强烈偏向立法修正而不是司法扩张。今天更成熟的答案可以是分层：刑罚扩张必须受到极强文本约束，但漏洞和不合理可以通过解释边界、违宪审查、立法更新等公开机制处理。关键不是零解释，而是让解释可见、可争、可复查。

压力测试六：如果AI能给出比人类更准确的再犯风险，能不能据此提前限制自由？准确只是事实属性，不是正当性通行证。风险预测可以帮助配置支持资源，却不能轻易把可能性当成已发生罪责。贝卡利亚要求刑罚建立在已被法律认定的行为上，这条古老边界在预测技术面前反而成为新的保护。

本编清账：六条

第一，正式完成中心层级清账：公共幸福暂居意义目标层；自由承担硬约束；必要与比例承担审查；罪刑法定、确定性、及时性、预防与教育承担不同路径功能。

第二，‘自由中心主义’不再恢复为主判。第49—50章确认，自由的强度极高，但其主要作用是限制公共幸福工程不能越界；因此继续使用‘自由限权型公共幸福中心’作为工作判断。

第三，‘主体先在’进一步降级。贝卡利亚已经看到经验主体会被教育、制度、奴役、知识和身份塑造。真正问题改写为：规范自由资格与经验自由能力之间缺少完整接口。

第四，‘把立法者请上看台’经本编深化为‘搭了看台，少了楼梯’，又经本章开头的反证降为1764年文本的结构诊断与现代延伸诊断。看台不是错误，而是总体立法理性的必要位置；楼梯不足才是问题，即公众经验、异议、实施结果怎样持续回写立法判断。

第五，本编新的总定位式是：刑罚权被发生化，主体也部分发生化，但‘自由主体—公共立法—实施结果’之间的循环尚未被充分制度化；本章新增的“锁钥同手”则把自由硬闸的例外结构单独拿出来检验（F11）。

第六，给SDE自己留下反证：若后续研究贝卡利亚的政治经济学讲义、行政实践或其他文本，能够证明他已经形成系统的信息反馈、公众参与与政策修正机制，则‘楼梯不足’必须撤回或降级。

◆ 反向攻击　四个最强反驳必须留下

反驳一：为什么一定要把立法者“发生化”？一个刑法理论本来就可以只规定立法者应当如何做，而不必解释立法者的社会来源。这个反驳很强。规范理论经常使用理想角色：理想法官、理性立法者、合理人，并不是因为现实里真有这种纯粹人物，而是为了建立评价标准。如果贝卡利亚的‘明智观察者’只是这种规范角色，那么SDE说他“站在看台上”不能直接构成缺陷。对此，本书只能把指控收窄：问题不是角色理想化本身，而是当这个角色拥有定义公共幸福、必要性和比例的权力时，理论有没有说明它怎样避免把自己的局部知识误当成总体知识。也就是说，发生化不是要求写传记，而是要求给判断来源与纠错留下接口。

反驳二：社会契约里的自由个人本来就是道德资格，不是心理事实，你凭什么问他有没有童年？如果一个理论说‘每个人都应被视为自由平等者’，它并没有声称婴儿出生时已经懂得法律、概率和契约。拿发展心理学攻击规范人格，属于层次错误。这个反驳同样成立一半。第52章已经因此把‘主体先在’改成‘规范资格与经验能力接口不足’。我们仍然追问童年，是因为刑罚不仅分配权利，也根据行为人的认识、选择、可责性和威慑反应设计制度。只要理论同时使用‘人会对确定性作出反应’‘教育可以改善行为’这类经验命题，规范主体与经验主体就不可能永远互不相干。问题不在于取消规范资格，而在于说明经验差异何时可以、何时不可以进入责任和政策。

反驳三：公共幸福既然是多数人的总体目标，为什么还需要这么强的个人自由闸门？如果某项严厉措施确实能让绝大多数人更安全，限制少数人不是合理代价吗？这正是‘幸福中心’最危险的试验。贝卡利亚的文本给我们的回答不是抽象的人权宣言，而是他的整个刑罚权来源结构：国家只能取得必要的最小自由份额，超出这一份额的强制没有正当来源。于是公共幸福并不是先形成一个总量，再从中扣除个人代价；它从一开始就被权力来源限制。少数人的身体和生命不是可以随意进入总账的普通变量。若某种所谓幸福只能靠任意酷刑、秘密惩罚或无法预知的规则维持，它在贝卡利亚自己的结构里就已经失去合法性。

反驳四：你说‘楼梯不足’，是不是用二十一世纪民主理论要求十八世纪刑法改革家？这项警告必须认真接受。贝卡利亚的任务是攻击酷刑、秘密程序、死刑和司法任意，不是写完现代参与式民主。他已经完成了自己的历史任务，我们不能因为他没有预见听证、数据治理、算法审计就判他失败。因此，‘楼梯不足’只能作为一个跨时代使用时的结构边界，而不能写成对个人能力的责备。更公平的表达是：当我们今天继续使用贝卡利亚的公共幸福与理性立法语言时，必须补上他那个时代尚未充分制度化的反馈装置；这不是说他‘应该早就发明’，而是说他的理论被带进今天以后需要完成新的回写。

四个反驳共同把本书的批判压回一个更小、更硬的位置：SDE不能因为能够描述更多发生条件就自封为更高级理论。它只有在补入这些条件后，能更好区分何时

应当追责、何时应当预防、何时国家必须退出、何时公共目标必须改写，才算真正增加了解释力。否则‘土、路、回写’也可能只是另一套漂亮词汇。

所以，本编结束时，我们给自己的判断加三条自限。第一，不把规范理想角色等同于经验实体；第二，不把环境影响等同于责任取消；第三，不把制度反馈不足等同于历史人物失职。SDE的任务不是证明贝卡利亚‘不完整’，因为任何理论都不可能写完一切；任务是确定当我们把他的制度继续用于今天时，哪些位置最需要新的发生机制才能避免旧原则变成新权力。

如果还要把这一轮收束成一个日常比喻，可以想象一座公共剧场。贝卡利亚做的第一件大事，是把原先坐在包厢里凭身份发号施令的主人请下来，告诉所有人：规则必须对每个观众一样，管理者的权力来自大家共同维持演出的需要。接着，他又需要一个看台来观察出口是否拥堵、秩序是否安全、灯光是否有效。SDE现在追问的不是‘为什么有看台’，而是看台上的信息从哪儿来，观众能不能告诉看台自己看不见什么，看台做错决定以后有没有楼梯下来重新改。

这也说明为什么‘搭了看台，少了楼梯’比‘主体先在’更适合作为一幅文本结构图。它既保留贝卡利亚的历史伟大，也指出理论在当代延伸时最容易产生的问题：公共理性一旦脱离具体经验，会把自己当成中立；经验一旦没有进入公共理性的制度通道，又只能通过愤怒、舆论或违法爆发出来。真正稳定的法治，不是把一边消灭，而是让二者持续交换。

这一交换同时给后面的AI编埋下一根最深的钉子。机器最像一座没有身体的看台：它可以看更多案件、更多变量、更多概率，却也更容易让人忘记它的视角是被训练数据、目标函数和制度任务造出来的。贝卡利亚担心法官的消化状态、激情和关系影响判断；今天我们要担心的是数据选择、标签定义和优化目标影响判断。两百多年过去，‘谁在看’的问题没有消失，只是从人的脸上移进了模型。

因此，前文不是把贝卡利亚推远，而是把他推到今天：一个真正忠于贝卡利亚的制度，不会因为拥有更强计算工具就减少解释义务，反而会因为强制能力更强而增加解释义务。国家越能预测、分类和提前干预，就越要证明为什么有权这样做、

为什么这一步必要、错误如何修复、被影响的人怎样进入后文规则形成。这个方向，既是对贝卡利亚的延伸，也是对SDE自己的约束。

> **带走的话**
>
> 贝卡利亚已经把国家刑罚权拉入发生过程，也多次看见主体与制度的相互塑造。经其后续行政实践反证，本书不再把“搭了看台，少了楼梯”作为对贝卡利亚个人或总体理论的定性；更窄的判断是：1764年文本主要依靠理性观察者与改革者吸收社会经验，尚未发展出现代意义上制度化、持续化的反馈架构。自由硬闸本身则是一道锁钥同手的闸：锁对普通犯罪几乎焊死，钥匙在国家手里，门朝死刑开（F11部分触发，F11′待检）。

> **小账**
>
> 他看见了什么？——死刑例外按对国家存续的威胁开，不按行为种类开，1792年仍未放低门槛。
>
> 接缝在哪里？——判定谁足以颠覆国家的正是国家自己，钥匙与锁同手，门朝死刑开。

第 十 一 编

给贝卡利亚找最强的辩护人

编　序

本编五章，专替贝卡利亚找最强的辩护。第57章说他没把人当成可精确计算的对象；第58章说他早已看见激情、概率和制度反馈；第59章说“自由人”是规范起点，不是经验画像；第60章说限制国家需要一个不被国家制造的人；第61章说人权有来历，也能有优先性。

这一编是判决四的主场。自由在这里被确认为规范上的硬约束：它不必是无来历的实体，却能挡住公共目标对个人的吞并。第58章同时为判决三作证：严刑使人麻木、法律过多制造犯罪、秘密告发让人猜疑，都是他亲眼看见的发生。

读这一编，可以带一个家常问题进去：医生看到模型给出73%，能不能就把这个数字当成病人本身？再带一个：父母参与了孩子的成长，孩子是不是就归父母所有？

第 57 章

他并没有说人可以被精确计算

前文一直有一个危险的读法：贝卡利亚谈比例、必要、确定性、概率，又喜欢把立法写成冷静计算，于是我们很容易把他想象成一个十八世纪的“刑罚工程师”——仿佛只要把犯罪损害、抓获概率、处罚痛苦放进一张表，就能算出最优刑罚。这个读法很顺，甚至和后来边沁的功利计算天然相接。但如果回到《论犯罪与刑罚》，它恰恰过于顺了。贝卡利亚从未真的相信人类行动可以像三角形一样被完全测量。

最关键的证据就在他讨论犯罪与刑罚比例的地方。他设想过：如果数学计算能够处理人类行动那些“晦暗而无限的组合”，我们当然可以得到一张从最重到最轻的精确刑罚阶梯。可紧接着，他就退了一步：现实中做不到，明智的立法者只需要标出主要分界，避免最严重的犯罪配上最轻处罚，或者反过来。也就是说，他真正追求的是秩序不颠倒、边界能预期，而不是每一档都拥有绝对刻度。这里早已埋着对“精确刑罚机器”的自我限制。

更有意思的是证据理论。贝卡利亚在讨论犯罪证明时明说，所谓道德确定性，严格讲仍是一种概率；人之所以把它叫作确定，是因为日常行动必须在足够可靠的概率上停止怀疑。这个判断非常重要。它说明他知道司法不可能进入上帝视角，证据不是把真相百分之百搬到法庭，而是在相互独立、相互支撑的材料中，把不确定性压到足以行动的程度。一个承认“确定性里面仍有概率”的人，很难被公平地写成迷信精确的人。

再看他的行为人模型。他确实反复谈痛苦、利益、希望、恐惧，也认为人会对处罚的确定性和及时性作出反应。但“会作出反应”不等于“按统一公式反应”。第七章里，他反对把意图当成犯罪的最终尺度，理由之一恰恰是意图受到当下感受、既有心理状态、观念次序、激情和具体处境影响，同一个人在不同时间也会不

同。换句话说，他不是没看到差异，而是因为差异太多，才不愿意让法官凭“我觉得他心里更坏”去自由加码。

这给本书一次真正的改判。过去我们把“量刑尺”列为收回候选，后来已经撤下；现在还要更彻底一点：贝卡利亚不是要把人精确算完，而是要把权力放进可公开比较的框架。比例、确定性、及时性更像三道公共检查，不是三个万能变量。它们要做的，是把处罚从个人情绪、身份特权和秘密判断中拉出来，而不是把人缩成几个数字。

为什么这一区分重要？因为今天算法治理正好会把贝卡利亚重新推向“计算”的方向。风险评分可以给出再犯概率，数据模型可以估计某种政策和犯罪率之间的关系，量刑工具甚至可以把过去案件压成统计分布。如果我们误以为贝卡利亚的理想就是“算得越准越好”，就会轻易把这些技术说成他的梦想成真。可如果他的核心其实是公开、必要、最少强制和反任意，那么计算反而必须接受这些原则的审查：数据从哪来？误差落在谁身上？能不能质疑？预测是否被偷换成罪责？

一个家常例子很清楚。医生面对一个复杂病人，会使用血压、化验、影像和风险模型，但好医生不会把“模型给出73%”当作病人本身。数字是帮助判断的工具，不是把人耗尽的定义。贝卡利亚对概率和主要分界的理解，更接近这种态度：制度需要可比较的信息，却不能假装复杂人生已经被一个刻度完全装下。

所以，第57章给最强辩护人第一分：说贝卡利亚“把人化成计算器”，证据不足。真正可以继续追问的不是他算得太多，而是他把哪些价值预先确定为计算目标，又由谁决定哪些损失可以相互比较。计算问题从“能不能算”退到“为了什么算、谁来定账”。这一步收窄以后，SDE的批评反而更硬，因为它不再攻击一个贝卡利亚自己已经否认的稻草人。

第57章阶段改判：贝卡利亚不是精确计算主义者。他要的是可公开比较的比例结构与足够可靠的判断，而不是把人类行动化成完美数学对象。任何以“他迷信精确”建立的总批判，至此正式撤回。

带走的话

贝卡利亚不是要把人精确算完，而是要把权力放进可公开比较的框架。

小账

他看见了什么？——精确刑罚阶梯做不到，只需标出主要分界；道德确定性仍是概率。

接缝在哪里？——哪些价值被预先定为计算目标，哪些损失可以相互比较，由谁定账。

第 58 章

他已经承认激情、概率和社会复杂性

如果第57章只是证明“贝卡利亚没有迷信数学”，还不够。更强的辩护是：他不仅知道不确定性存在，而且已经在多个位置把人的激情、社会习惯、法律环境和制度反馈写进理论。也就是说，我们此前用SDE的“土”去攻击他时，必须承认他自己早已看见相当多的土。问题只能是这些看见有没有被整合，而不能假装它们不存在。

先看刑罚严厉性。他说，刑罚越来越残酷时，人心也会逐渐变硬；一个时代用车裂制造恐惧，久而久之，人们会适应这种强刺激，原本骇人的刑罚不再产生同样效果。这里没有一个固定心理人。人的感受阈值会随着周围制度和习惯改变。刑罚不是向一个永恒不变的人施加信号，而是在塑造一个会适应、会麻木、会形成新常态的社会。

再看确定性。他强调较轻但几乎逃不掉的处罚，往往比极重却可能躲过的处罚更有效。为什么？因为行动者不是读取法典上的最高刑就机械停止，而是在经验中形成“这件事究竟会不会落到我头上”的期待。执法松紧、腐败、赦免、既往案例都会进入这个期待。这里的行为不是公式直接作用，而是制度经验经过记忆和希望之后形成判断。希望本身就是贝卡利亚特别看重的人性力量。

证据部分更明显。他把独立证据和相互依赖的证据区分开，提醒证据数量不是越多越好：如果十个证据都建立在同一个脆弱来源上，它们只是把同一个错误重复十遍。这种判断今天看仍很现代。它至少说明，他知道制度里的“信息”有结构，不能只数个数。所谓事实进入法庭，要看证据之间怎么连接。

他谈民族状态和刑罚时，也没有坚持一套跨时代不变的强度。全书结尾甚至说，刑罚严厉程度应与社会状态相适应；社会交往越发展、人的性情越柔和，维持必要关系所需的刑罚强度就应下降。今天我们当然可以批评这种“文明—野蛮”叙事带有十八世纪色彩，但就方法而言，它已经在承认刑罚效果与社会环境共变。

再看预防犯罪。他担心法律过多会反过来制造更多“犯罪”；秘密告发会让人彼此猜疑；教育和知识传播会改变人的判断；官员与奖励制度会改变行为动机。这些都不是一个只会做静态加减法的人会自然提出的问题。贝卡利亚已经看见制度不仅响应行为，还会生产行为的条件。前文把这一点概括为“法律不仅抓犯罪，也制造犯罪”，并非SDE替他加上的思想，而是从他自己的多个章节长出来的。

因此，第58章必须进一步限制“楼梯不足”这一总收回。我们不能说贝卡利亚没有反馈意识。他明明知道严刑会使人麻木，秘密程序会改变社会关系，法律数量会影响违法规模，教育会改变未来行为。更准确的说法只能是：这些反馈主要以聪明立法者需要掌握的经验规律出现，而不是以一套让被治理者、异议者和实施结果制度化进入立法修正的程序出现。不是没有反馈，而是反馈的组织方式偏“观察者式”。

这个差别很细，却决定批判是否公平。一个医生会观察病人吃药后的反应，再调整处方，这当然已经是反馈；但如果病人只能被观察，不能表达自己无法被指标捕捉的痛苦，反馈仍可能不完整。同样，贝卡利亚的立法者会观察刑罚效果，却不等于公共规则已经建立了充分的参与、申诉、公开数据与持续修法机制。十八世纪没有今天这些制度并不奇怪，问题只是当我们把他的理性立法模型带到今天时，不能把“观察”当作“被观察者已经发言”。

这也迫使SDE承认自己的一个风险：我们很喜欢说“他缺E”“他漏了回写”，但只要原典中出现环境、习惯、反馈、历史差异，粗暴的“缺”就不再成立。真正有价值的诊断必须从“有没有”升级为“以什么位置进入”。贝卡利亚不是没有土，而是土常常作为立法知识的对象进入；不是没有回写，而是回写常常由立法理性吸收，而非由多主体共同制度化。

带走的话

贝卡利亚已经看见激情、概率、习惯与制度反馈。SDE不能再以“他不知道复杂性”为由批判他。剩下的窄问题是：复杂性怎样从被观察的事实，变成能够反过来约束公共判断的制度声音。

小账

他看见了什么？——严刑使人心变硬，法律过多制造犯罪，秘密告发让人彼此猜疑。

接缝在哪里？——反馈由立法者观察吸收，被治理者没有制度化的发言通道。

第 59 章

“自由人”只是规范起点，不是经验的人吗

到这里，最强辩护人终于碰到本书最危险的一刀。我们前面一直追问：社会契约以前站着的那个自由个人，从哪里来？他有没有童年？他的语言、家庭、贫困、身份怎样形成他的选择？这组问题很有发生学味道，但一个规范法哲学家可以立刻反击：你问错层次了。社会契约中的“自由人”不是心理学报告，也不是历史人类学照片，它只是一个规范起点——每个人都应当被当作自由平等的权利主体。你拿童年、教育和环境去攻击它，就像拿身高去反驳“人人在法律上平等”，根本不在一个平面。

这个反驳必须承认，而且要承认得足够彻底。规范理论经常设置一种“应当如何看待人”的位置，而不是声称现实中的人已经具备所有能力。婴儿不会签契约，重病者可能无法表达意志，有些成年人也会在极端压力下失去选择能力，但这并不妨碍法律承认他们作为人的基本权利资格。若一定要求权利资格先经过经验能力证明，反而会把最需要保护的人排除出去。

因此，“自由主体有发生”不能直接推翻“自由主体应被承认”。前者说的是能力、偏好、判断、身份怎样形成；后者说的是国家面对任何人时必须接受什么限制。一个人怎样学会说话，可以被发生学解释；但国家不能因为他还不会说话，就认为他没有免受酷刑的资格。一个人的风险偏好可能来自贫困和创伤，但国家不能因此把他降为较低等级的权利主体。发生事实和规范资格在这里必须分轨。

这也是贝卡利亚社会契约语言最强的一面。即便我们不把契约当作真实发生过的历史事件，它仍然制造了一道制度想象：把国家的刑罚权写成每个人只交出极小自由份额后形成的共同权力。这个想象的功能不是告诉我们史前人类开过一次会议，而是要求国家反过来证明自己没有拿走超出必要的东西。自由个人因此更像一个“限权坐标”，不是经验世界里已经完成的原子。

如果这样理解，第52章“那个自由的人有没有童年”就要再改一次。问童年不是为了证明规范人格是假，而是为了处理贝卡利亚理论中的另一半：他不仅使用自由资格，还大量使用经验命题设计刑罚。确定性为什么有效、及时性为什么重要、教育为什么能预防犯罪、严刑为什么令人麻木，这些都依赖实际的人会怎样感受、学习、记忆和行动。只要经验人进入政策设计，规范人与经验人之间就需要一个接口。

这个接口的任务不是决定“谁配拥有人权”，而是决定“差异什么时候可以合法进入法律判断”。例如儿童与成年人责任能力不同，重度精神障碍可能影响归责，胁迫可能影响可期待性，贫困本身却不能自动取消责任。法律必须承认经验差异，又不能把这些差异变成身份歧视。于是最难的不是发生或先在二选一，而是：基本资格一律承认，具体能力有限进入，进入方式必须公开、可证明、可审查。

这使“主体先在”正式失去作为批判词的资格。贝卡利亚把自由主体作为规范起点，本身不构成发现学错误。任何现代权利理论都需要某种不依赖国家喜恶的地位，否则权利会变成行政评分。但与此同时，如果制度要靠经验心理去预测、预防和处罚，又不能把规范主体关进一个完全不受经验影响的玻璃盒。两层既要分开，又要有桥。

从SDE角度，这里出现一个非常重要的自我纠正：发生学不能把“有来历”变成“没有规范优先性”。一条规则可以有历史来源，却在现行规范中被设为不可轻易突破的底线；一项权利可以是社会长期斗争后形成的，却并不因此成为多数人随时可以撤销的偏好。把发生来源和现行效力混为一谈，本身就是另一种倒装。

第59章阶段改判：社会契约中的“自由人”首先是规范位置，不是完整经验画像。SDE无权用“他也有童年”取消这一位置。真正需要继续研究的，是规范平等与经验差异如何在责任、预防和制度设计中建立可审查的接口。

带走的话

发生事实和规范资格必须分轨：人有童年，不等于国家可以降低他的权利资格。

小账

这一章看见了什么？——社会契约里的自由人是规范起点，不是经验画像。

接缝在哪里？——规范平等与经验差异之间，还缺一个公开、可审查的接口。

第 60 章

为了限制国家，是不是必须先假定一个不可被国家制造的人

第59章把“自由个人”从被告席上先放了下来，第60章还要再替他辩一次，而且这一次更强：也许不仅不该批评贝卡利亚预先放一个自由主体，反而必须这么做。因为如果一个理论把人完全写成环境、制度和关系的产物，国家很容易接过这套语言说：既然你是被塑造出来的，那我当然可以为了更好的公共结果重新塑造你。

这不是想象中的危险。所有治理都多少包含塑造：学校培养习惯，税制改变选择，城市设计影响出行，刑法通过威慑塑造行为。问题在于，一旦“人完全可塑”成为政治前提，教育可能滑向驯化，预防可能滑向提前控制，矫正可能滑向人格改造，风险治理可能滑向根据预测限制尚未犯罪的人。一个理论如果只会说“这也是发生”，就没有给国家留下停手的理由。

贝卡利亚的“最小自由让渡”在这里像一道古老的防火墙。国家不是先有全部权力，再善意地少用一点；相反，它只有为共同安全绝对必要的那部分强制能够获得正当性。这个方向非常关键，因为它把举证责任放在国家一边。即使某种更深的塑造可能提高平均安全，国家也必须证明自己为什么有资格进入人的身体、生命和选择领域。

所以，一个不可被国家完全制造的人，至少在规范上很可能是必要的。这个“不可制造”不是说人的人格没有社会来源，而是说国家不能因为参与了你的形成，就宣称拥有你。父母参与孩子成长，不等于孩子是父母的财产；学校塑造知识结构，不等于学生只能拥有学校允许的思想；社会提供语言，不等于个人说话必须服从一个统一答案。同样，国家提供秩序和安全，也不因此拥有对公民的总体设计权。

这正好可以区分“发生”与“占有”。某种能力由关系发生，不等于关系中的任何一方拥有结果。一个乐团共同演奏出音乐，音乐不是指挥个人的私有物；一个社会共同形成法律秩序，公民也不是立法者可以随意配置的材料。发生学如果只强调依赖关系，却不区分依赖和支配，就会把解释结构误变成政治授权。

贝卡利亚的反酷刑尤其能说明这一点。即使国家认为刑讯能更快获得信息，即使某些情形下它真的可能提高破案率，也不能仅凭效果就越过无罪者的身体。因为处罚与调查的边界先由人的法律地位决定。一个尚未证明有罪的人仍在公共保护之内，不能被当成侦查工具。这里的主体资格不是统计结论，而是限制国家使用人的方式。

死刑争论同样如此。贝卡利亚先问“国家有没有这个权利”，再问“它有没有用”。我们前面把它概括为双钥匙结构：权利来源是一把钥匙，必要与效果是另一把。即使一种措施可能有效，也不自动因此正当。这恰恰是防止公共幸福吞掉个人的最强结构。若SDE只会用发生与效果去评价，就会丢掉第一把钥匙。

因此，“自由人先在”在这里显出另一面：它可以不是形而上学的“自然原子”，而是一种政治上的不可吞并声明。它告诉公共权力：你可以描述我的处境、改善我的能力、限制我对他人的侵害，但不能因为我由社会形成，就宣布我的基本地位也完全由社会任意授予。社会可以参与人的发生，却不能垄断人的价值。

第60章给本书一次更大的反向判决：如果SDE想比贝卡利亚多说一句，就必须同时多承担一项责任——说明“发生”怎样与不可支配相容。否则，贝卡利亚那个看似先在的自由主体，反而比一切都可塑的发生论更能保护人。

> **带走的话**
>
> **为了限制国家，确实需要某种不由国家单方面制造、也不能被公共目标全部吞并的规范主体。这个位置可以有历史来源，却必须在现行制度中对国家保持优先约束。**

> **小账**
>
> 这一章看见了什么？——人若完全可塑，国家就能以公共结果为名重新塑造人。
>
> 还要追问什么？——SDE怎样说明发生与不可支配相容，分清依赖与支配。

第 61 章

如果没有一个先在自由，还有没有人权

现在问题被推到最尖：如果我们不再承认一种形而上学意义上的“天生自由实体”，人权还站得住吗？很多人会担心，一旦说权利也是历史中形成的，权利就变成风俗；既然是社会造出来的，社会当然也能收回。于是为了保护人权，我们似乎必须把自由写成社会以前就完整存在的东西。

但这里其实混了三个不同问题。第一，某个权利观念在历史上怎样出现；第二，今天我们用什么理由为它辩护；第三，在现行法律与政治秩序中，它具有多高的效力。一个权利可以有清楚的历史发生过程，却在今天被赋予最高等级的规范地位。知道“禁止酷刑”经历过漫长斗争，并不会让禁止酷刑变弱；恰恰相反，发生史让我们知道为什么这道边界如此来之不易。

所以，人权不必靠“没有来历”才能强。真正重要的是：它不能只靠当下多数人的偏好维持。一个成熟的权利秩序会把某些边界设成普通成本收益计算不能轻易穿透的硬约束，并配置独立法院、公开理由、程序保障与修法门槛。规范优先性来自制度承诺和道德理由的共同结构，而不是来自它是否在史前时代就已经作为一个现成物存在。

这让我们重新理解贝卡利亚。说每个人只交出最少自由，可以被看作一种规范构造：它不必证明真实历史中每个人都签过契约，只需要持续提醒立法者，公民不是国家先拥有后返还一点自由的臣属，而是国家必须为每一份强制说明来源的人。这个结构对人权极有亲和力，因为它天然要求国家面对个人时承担说明责任。

但“人权有发生”仍然带来一个重要提醒：任何权利清单都可能有盲点。十八世纪强调身体、生命、财产和免受任意处罚，今天还要处理数据、隐私、算法歧视、环境和人格发展。权利如果被神圣化为一张永远完成的表，新的侵害可能无处进入。发生学的贡献不是把旧权利降级，而是让权利体系保留发现新伤害、形成新保障的能力。

这正是“不可自我封顶”的意义。人权秩序需要一方面拥有硬底线，另一方面承认底线的表达可以被新经验扩充。禁止酷刑不能因为社会变化而变软，但“什么算酷刑”“哪些新技术构成非人道对待”可以随着技术和经验进一步澄清。基本价值保持，适用结构发生。这比“要么永恒不变、要么任意可改”的二选一更接近实际法治。

从SDE角度，第61章也必须防止一种发生学谬误：因为一个规范有社会来源，就推断它只是一种相对意见；或者因为一个行为有环境原因，就推断行为人没有责任。这两种推论都不成立。来源解释回答“怎么来的”，正当性回答“为什么应当”，归责回答“谁该承担”。三条线可以相互影响，却不能相互代替。

经过第57—61章，最强辩护已经替贝卡利亚赢回很多地盘。我们撤回“精确计算主义”；撤回“主体完全未发生化”；承认自由主体作为规范起点有合理性；承认某种不可被国家完全制造的位置是限权所需；承认权利有历史发生并不削弱其现行优先性。现在还能留下什么批判？只剩一个更具体的制度问题：当公共幸福需要总体判断，而权利又要求个体边界时，两者之间的理由怎样被不断公开、挑战和修正。

因此，“搭了看台，少了楼梯”还站得住，但它的含义已经变了。楼梯不是为了把看台拆掉，也不是为了让每个个体随时否决公共规则；它只是要求总体判断不要把自己当成无来源的理性。经验、异议、误判、少数负担和新技术必须有路径进入后文规则。看台负责拉开距离，楼梯负责防止距离变成统治。

> **带走的话**
>
> 自由不必是形而上学上无来历的实体，仍然可以在规范上具有优先性。SDE的贡献不能是“让人权也发生掉”，而应是解释权利怎样形成、怎样制度化、怎样在新经验中扩展，同时守住不被普通效用计算吞没的硬边界。

小账

这一章看见了什么？——权利有历史来源，并不削弱它的现行规范优先性。

还要追问什么？——底线不变时，数据、算法等新伤害怎样进入权利清单。

第 十 二 编

康德、边沁以后：贝卡利亚到底留下了什么

编　序

本编五章，请外部对手上场。第62章是康德：刑罚先问应得；第63章是边沁：把效用推成系统工程；第64章说现代刑法同时继承了两人，并用编码检验“授权先于目的”；第65章说罪刑法定为何最耐久；第66章说死刑为何没被一本书杀死。

这一编是判决二的主场。康德守住归责，边沁守住节约痛苦，贝卡利亚把一般法律、必要、比例和公共理由编在一起，站在两条路的交界处。第65、66章又把判决一落进制度：先有公开规则，国家才谈得上证明强制的来源。

读这一编，可以想一件学校里的事：处罚作弊学生，是先问他真做了没有，还是先问能不能杀鸡儆猴？也可以想：规定在你做完以后才公布，你还能怎样安排自己的生活？

第 62 章

康德：不是有用，而是应得

如果说第57—61章是让贝卡利亚自己反击SDE，第62章就让一个更强硬的外部对手上场：康德。两个人在死刑问题上的冲突特别适合做一面镜子，因为他们并不只是对“杀不杀”给出不同答案，而是对“为什么可以处罚”使用了不同语法。贝卡利亚总是先问权力来源、必要性和社会效果；康德则把处罚首先系在已经发生的违法与应得上。

在《道德形而上学》的法权论里，康德反对把刑罚主要当作制造未来好结果的工具。一个人不能仅仅为了给别人做榜样、让社会更安全而被处罚；否则国家是在把他当作实现别人目标的手段。处罚必须首先因为他已经实施了犯罪，并且刑罚要回应这项不法本身。正因为如此，康德才把报应原则推到极硬的位置：对故意杀人，死亡被他看作与所犯罪行相称的刑罚。

这正面撞上贝卡利亚。贝卡利亚问：社会契约怎么可能包含个人把生命交给国家的权利？康德的回答并不是去证明一个人事实上愿意被杀，而是把“作为理性立法者的我”和“作为具体犯罪人的我”分开。具体的犯罪人当然不愿受罚，但作为纯粹理性、参与一般法则形成的主体，他能够认可一条对所有故意杀人者同样适用的规则。于是，处罚不是从个人某次真实同意中来，而是从合理法秩序中的普遍原则来。

这一步恰好暴露两种社会契约的差别。贝卡利亚的契约语言更像限权账本：每个人只交出维持共同安全所需的最少自由，国家不能多拿。康德的“原初契约”则更像理性检验：法律应当能够被全体自由平等者作为共同意志认可，它并不要求历史上真的发生过一次交易。前者不断问“国家拿了多少”，后者不断问“这条原则能否作为普遍的法被共同意志认可”。

康德的强处，是把贝卡利亚最容易被质疑的一点堵住：如果处罚只为未来效果，犯罪人的身体是否变成社会治理工具？假如公开处死一个无辜者真的能阻止暴

乱，纯粹结果论会遇到灾难。贝卡利亚本身当然不会赞成惩罚无辜，因为他的罪刑法定、证据和无罪保护已经设了闸；但康德把这条闸提升为第一原则：处罚必须以行为人的不法和可归责为中心，而不能先看是否有用。

可康德也立起了自己的硬地板。一旦应得被放到第一位，社会效果就容易退居次位。故意杀人的死刑因此被写得极其坚硬，即使贝卡利亚指出长期剥夺自由可能有更强的威慑、死刑不可逆、国家可能误判，康德仍会说这些不是决定“应不应该罚到什么程度”的第一根据。对现代人来说，这种坚持保护了不把人当工具的尊严，也可能把报应本身推得过于封顶。

SDE在这里不能简单站队，而要看两人各自堵住什么漏洞。贝卡利亚堵住的是权力无来源和无必要的扩张：国家不能因为愤怒或传统就施加痛苦。康德堵住的是把罪犯当成政策材料：即使公共利益很大，也不能跳过他的行为、责任和应得。一个防止国家“拿太多”，一个防止国家“为了别人使用你”。两道闸不是同一道，却都限制权力。

更有意思的是两人都不是我们最粗略的标签。贝卡利亚不只是后果论，他先问国家有没有权；康德也并非完全不看人的具体处境，法权论内部仍处理归责、紧急状态、主观认识与某些特殊情形。真正的差异在优先顺序：当应得与社会效果冲突时，谁有第一句话语权。

用一个日常例子：学校处罚作弊学生。贝卡利亚式问题会问，处罚是否必要、是否能预防、是否过重、规则是否事先明确；康德式问题会先问，这个学生是否真的做了可归责的作弊行为，他应当承担什么与行为相称的后果。前者防止学校用重罚“杀鸡儆猴”，后者防止学校为了整体秩序随便抓一个人当样板。成熟制度其实需要两问都答。

带走的话

康德对贝卡利亚最大的挑战不是“死刑更严厉”，而是要求刑罚先回答应得与归责，再谈社会效果。贝卡利亚限制国家拿多少权力，康德限制国家为了公共目标怎样使用一个有责任的人。两人共同把刑罚夹在两道边界之间。

小账

这一章看见了什么？——康德要求刑罚先回答应得与归责，不许为别人使用罪犯。

接缝在哪里？——应得一旦免检，误判、不可逆与替代刑罚会被降成次要问题。

第 63 章

边沁：把效用推得更远

如果康德是从另一方向把贝卡利亚拉住，边沁则像把贝卡利亚已经打开的道路继续铺成高速公路。贝卡利亚反复问刑罚有没有必要、有没有用、是否以更少痛苦获得更大安全；边沁把这一思路系统化为效用原则：评价私人行为和政府措施，都要看它们是否增加相关人的幸福、减少痛苦。共同体利益在他那里最终还要回到组成共同体的个人利益。

边沁对刑罚有一句非常重要的基本立场：惩罚本身是恶。它之所以可能被允许，不是因为痛苦本身有道德价值，而是因为它有机会防止更大的恶。这个起点和贝卡利亚极其接近。两人都拒绝把国家施加痛苦写成天然善事；每一份额外痛苦都要说明自己产生了什么必要效果。边沁后来关于处罚“节约性”的要求，几乎把贝卡利亚的“最少必要刑罚”做成了更系统的设计规则。

他甚至把处罚的实际痛苦和处罚在人心中的表象效果区分开：对一般预防而言，真正发生作用的是人们对处罚的观念，而真实处罚只是为了维持这种观念所付出的成本。这个判断把“少受真实痛苦、保留足够威慑效果”的方向推得很远。只要能用较少实际伤害制造同等预防，就没有理由多制造痛苦。这里可以看见贝卡利亚“确定性胜于残酷性”的明显亲缘。

边沁还系统讨论什么时候不应处罚：如果处罚没有根据、没有效果、成本过高，或者根本没有必要，就应退出。处罚可能误伤无辜，可能因为发现概率过低而必须不断加重，可能因为个体感受差异而产生远超设计值的痛苦。由此，刑罚被放进一张更完整的成本—效果账本。贝卡利亚的直觉在这里变成了工程化的政策分析。

但正因为边沁走得更远，贝卡利亚身上的“自由硬闸”反而更值得重新看。若所有问题都最终折算为幸福与痛苦的总量，某些个人边界会不会成为可交易变量？边沁当然有许多复杂的法治设计，并不能用一句“多数暴政”概括；可从结构上

说，效用一旦成为最高裁判，就必须额外解释为什么某些权利不能因为总收益更大而被轻易牺牲。贝卡利亚的最小自由让渡，在这里像一条不愿完全溶入总账的线。

这也说明“自由限权型公共幸福中心”为什么比单纯“功利主义”更适合描述贝卡利亚。公共幸福是方向，但自由、罪刑法定、无罪保护和权利来源会对方向设边界。边沁则把效用系统化程度显著提高，甚至试图把不同痛苦、概率、强度、持续时间都纳入更明确的分析。两人不是简单前后相同，而是同一条道路上的不同配比。

从SDE看，边沁最大的创新是把“路”工程化。一个政策目标不再只靠直觉判断，而是拆成可比较的效果、成本、概率和替代方案。这个工具极强，因此也更需要“土”和“回写”的约束：谁定义幸福？谁的痛苦容易被低估？不可量化的尊严、污名和制度信任怎样进入？如果模型只保留能计数的变量，所谓理性就会把看不见的东西当成不存在。

AI时代尤其容易重新边沁化。今天真的可以统计犯罪率、预测风险、估算执法成本、模拟政策效果，甚至给不同措施设权重。技术让“效用账本”比十八世纪精细数百倍。可同样的问题仍在：一个模型能优化既定目标，却不会自动告诉我们目标应不应该这样设。降低再犯率可以是目标，但不能因此自动授权永久监控；减少案件积压可以是目标，但不能因此牺牲无罪者的申辩。

边沁因此既是贝卡利亚的继承者，也是贝卡利亚的一面放大镜。他把“少一点无用痛苦”推进为系统效用原则，让刑罚政策获得前所未有的可分析性；同时也把贝卡利亚没有完全解决的价值排序问题放大：当一切进入总账时，哪些东西不能被普通加减法处理？

> **带走的话**
>
> **边沁把贝卡利亚的必要性、预防与节约痛苦推成系统效用工程。贝卡利亚留下的自由边界因此反而更清楚：公共幸福需要计算，但计算不能独自决定国家对人的全部权限。**

> **小账**
>
> 这一章看见了什么？——边沁把惩罚本身是恶、节约痛苦推成系统效用工程。
>
> 接缝在哪里？——一切进入总账时，少数人的权利可能被平均成小数点。

第 64 章

现代刑法为什么同时继承了两个人

走过康德和边沁，再回头看现代刑法，会发现它很少真正选择“纯康德”或“纯边沁”。现实制度更像一台双发动机：一边坚持只有在行为可归责、法律事先规定、刑罚与罪责相称时才能处罚；另一边又持续关心威慑、预防、矫正、再犯、成本和社会安全。前一边的语言更接近应得与权利，后一边更接近效用与治理。两边经常吵，却长期共同工作。

◆ 与哈特正面划界：从“两阶段”改为“三道闸”

这里必须公开承认先占。H. L. A. Hart 在1959年的《刑罚原则导论》中已经把刑罚的一般正当化目的与处罚的分配、责任和减免问题区分开，并强调一种可接受的刑罚理论往往是若干部分冲突原则之间的妥协。本书早先把“先问能不能罚、再问怎么罚”称作自己提出的“两阶段正当化”，说得过头了。

修正以后，本书不再争“两阶段”这个名字，也不再把“第零闸——授权来源”写成自己的理论发明。更准确的做法，是把三道闸当作叙述框架：先问国家是否取得这种强制的权限，再问刑罚制度的一般正当化目的，最后问责任和具体刑罚怎样分配到这个人。这个次序有助于读贝卡利亚，但“惩罚权从哪里来”本身早已有独立理论传统。

死刑最能说明第零闸的作用。贝卡利亚并不只说“死刑没有用”，而是先问：组成主权的自由让渡里，是否包含了把生命交给国家的权限。即便我们不同意他的契约论答案，这个问题在逻辑上仍不同于“死刑有没有一般社会效用”，也不同于“某个杀人者是否应得某种刑罚”。因此，本书把三道闸写成：有没有这种权—为什么需要刑罚制度—为什么由这个人承担这种处罚。

本书自己设置的推翻条件在这里已经触发。洛克在《政府论》下篇讨论自然法执行权；A. John Simmons 1991年的《Locke and the Right to Punish》直接

把惩罚权作为独立问题处理；R. A. Duff 后来的责任与刑事审判理论也追问政治共同体凭什么有资格要求行为人就公共错误作答。因此，“第零闸”应从理论增量降级。

先看定罪。现代刑法不会因为一个人“看起来危险”就直接把他当罪犯，通常仍要求已经发生的法定行为、证明标准和主观责任。这个结构明显拒绝把纯粹未来效果当作处罚基础。即便预测某人未来再犯概率很高，也不能把概率本身当成已经完成的犯罪。这一层更靠近康德式的归责边界，也延续贝卡利亚的罪刑法定与无罪保护。

再看量刑与执行。到了刑罚选择，制度又会问很多后果问题：什么处罚能减少再犯，监禁是否制造新的犯罪能力，社区矫正是否更有效，治疗、教育、就业支持能不能降低风险，监狱成本是否值得。这里明显进入贝卡利亚—边沁路线。现代刑法不只是结算过去，也在设计未来。

比例原则本身就站在两条路交界处。说刑罚不能明显超过犯罪严重性，既可以理解为应得的上限，也可以理解为防止无用痛苦和权力过度的限权工具。不同理论会给出不同理由，却可能支持同一制度。贝卡利亚最早把比例、必要和最小强制编在一起，正因为他没有彻底把报应与效用分成后来那样清楚的两军，他反而成了一座桥。

这座桥的好处，是让现代刑法不必把一个问题全交给一种价值。一个严重故意犯罪，处罚不能只问“这样能不能吓住别人”，还要问行为人是否应当承担这种后果；反过来，确定有责也不能自动推出“越痛苦越正义”，还要问处罚是否必要、是否过量、是否存在更少损害的替代方案。应得决定“不能把谁当工具”，效用决定“不要制造没有必要的痛苦”。

但双发动机也会打架。再犯风险高的人是否应该判得更重？从预防看似乎有理由，从应得看则可能因为未来尚未发生而受到限制。惯犯制度、保安处分、少年司法、精神障碍者处遇、毒品政策都不断出现类似冲突。现代制度不是已经解决，而

是把冲突做成程序：有些因素可以进入，有些因素只能有限进入，有些因素必须排除。

这里正好回到本书的“看台—楼梯”。现代刑法拥有比贝卡利亚时代高得多的看台：统计、犯罪学、心理学、经济学、数据科学能观察大量总体规律；同时也拥有更复杂的楼梯：公开立法、司法审查、辩护、上诉、违宪控制、学术批评、媒体与社会组织都可能把个体经验和错误反馈回制度。楼梯并不完美，但它让公共理性不再只能依靠一个明智立法者。

因此，本书对贝卡利亚的评价也应进一步收敛。他不是没有楼梯，而是处在一座现代楼梯尚未建成的建筑早期。他已经打掉君主和法官任意的私人楼梯，铺出一般法律、公开规则、比例、证据和预防这些主梁；后来的宪政、民主、程序法与社会科学再不断补充反馈通道。若把“没写完二十一世纪制度”当成他的失败，显然不公。真正有意义的是看他留下的结构在哪里仍需要续建。

康德和边沁的加入，使“续建”更清楚：一条楼梯把个体责任和尊严送上看台，提醒总体政策不要跨过人的应得与权利；另一条楼梯把社会效果和实际痛苦送上看台，提醒正义不能只满足抽象对称而忽视制度后果。现代刑法的成熟，不在于找到一台唯一发动机，而在于让两种理由彼此限制。

这也让贝卡利亚重新显出独特位置。他不是纯粹的康德前身，也不是简单的边沁前身。他一方面用社会契约和最小自由份额给国家设来源边界，另一方面又用公共幸福、必要、预防和刑罚效果评价制度。正因为两条线都在，他才成为现代刑罚理论里一个持续可回返的起点。后人越把两条路线分得清楚，越能看见他那本小书里已经同时埋着两颗种子。

◆ 惩罚权来源：第零闸为何必须降级

与 Simmons、Duff 等研究对照以后，本书保留的不是“发现了惩罚权来源”这一主张，而是一项更窄的思想史判断：贝卡利亚在1764年的论证顺序中，尤其在死刑章，常常先追问国家有没有获得这种权力，再进入有用与否、如何分配的

讨论。这个“授权先于目的”的文本次序可以被检验，却不能冒充刑罚哲学的新层级。

这样处理反而使三道闸更诚实：它是一张阅读与论证清单，而不是一套宣称前无古人的理论。若后续研究发现贝卡利亚在大多数关键章节其实并不维持这种次序，本书就应放弃“问题次序”判断；若文本持续显示这一先后关系，则其价值在于重新显露贝卡利亚如何组织问题。

◆ 问题次序编码：六章里谁先出场

凡例说过，“授权先于目的”只是一个可检验的思想史判断，不是理论发明。检验的办法是编码：在剑桥版中挑出同时包含权利论证与效用论证的六章，逐章记录两件事——哪一类论证先出场，两者冲突时哪一边让路。

剑桥版章	先出场的论证	冲突时谁让路	编码
第2章 惩罚权	必要与授权同时出场：每一种不出于绝对必要的刑罚都是暴虐；惩罚权是每人交出的最小自由份额之和	授权由“必要”来界定，二者尚未分离	混合（定义章）
第16章 酷刑	授权：判决之前不能称人为罪犯；在有罪无罪未定之时，除了强力，还有什么权利允许施刑	效用与认识论证（强者脱罪、弱者受刑）随后出场，作为补强	授权在先
第27章 刑罚的温和	效用：遏制犯罪靠的是刑罚的必然，而不是残酷	未出现独立的授权论证	效用主导
第28章 死刑	授权：死刑不是一种权利，谁会把处死自己的权力交给别人	效用论证（不必要、无益）随后出场；例外按国家存亡开	授权在先
第41章 如何预防犯罪	效用：预防胜于惩罚	未出现独立的授权论证	效用主导
第46章 赦免	效用：赦免助长“或可逃罚”的希望，仁慈应写进法律本身	未出现独立的授权论证	效用主导

结果是：授权在先两章，混合一章，效用主导三章。授权在先的两章，第16章和第28章，恰好是国家直接处置身体与生命的两章；第2章是给惩罚权下定义，授

权与必要从一开始就缠在一起。效用主导的三章，处理的都是刑罚怎样运作得更好。

因此，“贝卡利亚往往把授权放在效用之前”这个笼统说法不成立，应当放弃（F9部分触发）。能够保留的是一个更窄、也更锋利的判断：当国家要直接处置一个人的身体与生命时，贝卡利亚先问国家有没有这种权；当问题只是刑罚怎样更有效时，他直接以效用立论。授权在他那里不是贯穿全书的第一原则，而是一道只在最重的强制面前才启动的闸。

这次编码由本书作者独立完成，只覆盖六章，“先出场”的判定也依赖对段落主旨的阅读。出版前应请第二位编码者按剑桥版原文复核，并把样本扩展到全书中所有同时出现两类论证的章节，例如第25章“放逐与没收”。

◆ 扩样复核：十八章里谁先出场

本版按上面的要求做了扩样。第二位编码者同样不知道本书结论（也是一个大模型实例，直接依据意大利文维基文库所收1766年本全文），任务是找出全书所有同时含有授权论证与效用论证的章节，逐章判定谁先出场。它找出核心十八章，另有六章证据较弱，单独计数。

先看一致性。作者编码过的六章，第二编码者的判定全部一致：第2章融合，第16、28章授权在先，第27、41、46章或效用在先、或根本没有独立的授权论证。两人在重叠样本上六比六。

再看扩样。核心十八章中，授权在先十章（第1、3、4、16、19、21、24、25、28、35章），效用在先三章（第30、34、36章），融合五章（第2、29、31、32、37章）。授权在先的章节确实成团，但成的不是上一版说的那一团：除了酷刑、审前羁押、死刑这些“定罪之前的人身”问题，还有一大块是“权力归谁”——法官能不能解释法律（第3、4章）、贵族能不能另罚（第21章）、庇护所能不能在法律之外（第35章）。反过来，同样处置身体的第22章（盗窃，主张以劳役抵罚）、第27章（刑罚的温和）、第36章（悬赏取人性命），却是效用先出场。

所以上一版的收窄判断——“授权只在国家直接处置身体与生命时启动”——在扩样面前不成立，F9再次触发。新的、仍然可输的表述是：贝卡利亚在两类问题上先问授权，一是权力归谁，二是定罪之前的人身；在“刑罚怎样运作得更好”的问题上，他直接以效用立论。第二编码者还提醒了一个细节：第28章虽然先说“死刑不是一种权利”，决定胜负的却是随后那句——只要证明死刑既无用也不必要，人道的事业就赢了。授权开了头，效用收了尾。

这个结果和第56章的“锁钥同手”接上了。“权力归谁”本来就是钥匙问题；贝卡利亚在这类问题上一贯先问“凭什么权利”，唯独在自己的死刑例外处没有问——把钥匙交给国家时，他没有追问国家凭什么握着它。扩样之后，这个空位比上一版更清楚。

限制同样写明：第二编码者自认对第1、19、29、36、37章的先后判定主观性较高；证据较弱的六章是否计入会改变比例（计入后为授权在先十一、效用在先七、融合六），但不改变“授权先问集中在权力归属与定罪前人身”这一方向。人类编码者的复核仍待做。

◆ 专题深描　八个问题把最强辩护做实

本编深描一　可计算，不等于可穷尽

法治为什么总喜欢数字、等级和清楚边界？并不一定因为法律相信人可以被数字穷尽，而是因为强制权越重，越需要让理由能够被第三个人复查。一个法官只说“我感觉你危险”，别人很难检验；他说“我依据这几个公开事实、这条法律和这个证明标准”，至少留下了争辩的把手。贝卡利亚追求的“比例”和“确定”，首先是把权力从私人直觉搬到公共理由，而不是宣布公共理由从此没有剩余。

这一点可以帮助我们区分三种完全不同的“计算”。第一种是排序：知道杀人通常比轻微财产损害严重，刑罚不能轻重倒置。第二种是估计：知道抓获概率、执行延迟会影响威慑，因此比较不同政策效果。第三种是穷尽：相信只要变量足够

多，就能把一个人的责任和未来完全算完。贝卡利亚前两种用得很多，第三种证据不足。今天算法治理最危险的恰恰是把前两种顺手滑成第三种。

所以，本书后来讨论AI量刑时要守住一个重要界线：法律可以使用计算，但必须保留“计算之外仍有法律判断”的位置。预测概率不是罪责，统计相关不是个体事实，平均效果不是权利来源。数字可以帮助说明理由，却不能自己成为理由。只要这条界线存在，贝卡利亚式理性化就不必等于技术统治。

本编深描二　概率进入法庭以后，谁承担错误

贝卡利亚把道德确定性解释为足够高的概率，这一判断在今天尤其值得继续。司法从来不是在绝对知识之后才行动，否则绝大多数案件都无法裁判。真正的问题是：当事实永远带着剩余不确定性，错误风险应该怎样分配。刑事法治之所以要求很高的证明标准，本质上不是假装错误不存在，而是决定宁可承受多少“有罪者未被定罪”的风险，以降低无辜者被国家强制的风险。

这和贝卡利亚反酷刑的结构相通。酷刑试图通过增加身体痛苦来换取更快确定性，可它实际上把不确定性的成本强行压到尚未证明有罪的人身上。现代算法也可能重演同样逻辑：如果模型只有70%的风险准确率，剩下30%的错误由谁承担？一个政策若把统计上的不确定性直接变成某些人的真实限制，制度就必须说明为什么这些人该替模型的不确定付费。

因此，概率不是让正义变软，而是让举证责任变清楚。越承认认知有限，越不能让国家装成绝对正确。一个制度真正成熟的标志，不是它从不犯错，而是它预先承认错误可能发生，并安排复核、上诉、补偿、撤销和纠正。这里其实就是“楼梯”的最具体形态之一：错误结果必须有路回到规则。

本编深描三　规范平等为什么不能被经验差异吃掉

现实中的人当然不同。年龄、教育、认知能力、疾病、创伤、社会资源都可能影响行动。刑法若完全无视这些差异，会制造新的不公；可如果把差异无限放大，又会让“平等主体”消失，最后每个人都被一套风险标签重新分层。现代法治一直

在这两端之间走钢丝：一方面承认责任能力、胁迫、认识错误等具体差异，另一方面拒绝仅凭贫困、种族、阶层或生活方式把某人预先当成更低等级的权利主体。

所以“规范主体—经验主体接口”不是一句抽象术语，而是一套过滤器。经验差异要进入法律，至少要回答：它和本案的认识或控制能力有什么具体关系？证据是什么？这种差异是解释行为，还是足以改变归责？如果只是统计上“这一类人平均更危险”，它就不应直接变成个体罪责。

贝卡利亚把刑法从人格审判拉回行为和社会损害，本身已经是在保护这个接口。他担心法官凭“这个人心更坏”自由量刑。两百多年以后，我们换成机器说“这个人风险更高”，结构可能仍然相似。最好的继承不是回到完全不看差异，而是把差异的法律进入方式做得更窄、更透明。

本编深描四　不可被制造的人，不等于孤立的人

为了限制国家而承认一个不可被国家吞并的主体，很容易被误解成极端个人主义：好像每个人都是封闭原子，与家庭、社会、语言无关。其实没有必要。一个人完全可以在关系中成长、依赖他人、接受公共教育，同时在规范上拥有不能被他人占有的地位。依赖并不等于从属，形成并不等于所有权。

这一区分对刑罚特别关键。国家当然可以通过教育、城市设计、社会保障和警务降低犯罪条件，但不能因此说“既然你的行为受环境影响，我就有权提前重塑你的全部生活”。公共政策可以改善机会结构，却不能把人降成政策工程材料。贝卡利亚的最小自由让渡，正好提供一个很朴素的版本：共同生活需要强制，但强制只拿它确实需要的部分。

所以，SDE真正需要补的不是“人全由自己创造”，而是“发生关系中怎样保留不可支配性”。这可能成为本卷后面给SDE自己写死法的重要条款：如果一个发生理论只会证明人彼此依赖，却无法说明为什么依赖中的人仍不能被工具化，那么它在政治哲学上是不完整的。

本编深描五　权利有历史，并不等于权利只是历史

思想史研究最常见的一种错误，是把“怎么来的”和“是不是正当”混在一起。我们发现一个权利观念在某个时代出现，就说它只是那个时代的产物；我们发现一个制度有具体政治背景，就说它没有普遍理由。这种推论并不成立。说明来源可以帮助我们理解盲点，却不能单独完成规范否定。

贝卡利亚本身就是最好例子。罪刑法定、反酷刑、刑罚必要性都有十八世纪欧洲的具体土壤，可今天继续接受这些原则，并不是因为我们仍活在1764年的米兰，而是因为它们能被新的理由重新辩护：权力可预期、身体不可被任意工具化、国家强制需要必要性。一个发生出来的原则，可以在后续历史中获得更强、更广的根据。

反过来，所谓“永恒权利”如果拒绝任何发生史，也可能隐藏排除。过去宣称普遍的人权体系并不总是把所有人真正包括在内。发生史能提醒我们，谁曾被排除、哪些经验没有进入原来的“普遍”。因此最稳的结构不是“有历史所以相对”，而是“有历史所以持续自查，因有充分规范理由所以保持硬度”。

本编深描六　康德的强闸，也有自己的风险

康德用应得与人格尊严阻止国家把罪犯纯粹当作威慑工具，这是对后果主义最强的一道闸。但强闸也会反过来产生问题：如果某种痛苦被认定为“应得”，是否就不必再问它有没有必要？如果死刑被写成故意杀人的严格对价，误判、不可逆、制度偏差和替代刑罚的存在会不会被降成次要问题？一旦报应自身获得免检地位，它也可能成为新的磐石。

贝卡利亚的价值就在这里重新出现。他强迫一切刑罚接受“有无必要、是否最少”的追问，即便某人有罪，也不能无限加码。康德提醒我们不能为了别人使用罪犯，贝卡利亚提醒我们即使面对有罪者，国家也不能把痛苦当成天然善。两个限制叠起来，比任何一边单独工作都更安全。

SDE因此不能把康德简单写成“理念中心”、贝卡利亚写成“现实中心”就结束。真正值得分析的是两种错误的互相阻止：纯结果可能工具化人，纯报应可能神圣化痛苦。现代刑法的混合结构，不是理论懒惰，而是因为两种风险都真实存在。

本编深描七　边沁的总账，怎样避免把少数人变成小数点

边沁说共同体利益最终由成员利益构成，这比把“国家利益”当神秘实体透明得多。但一旦进入总量计算，新的困难立刻出现：总幸福增加，并不能自动说明每个人都变好。十个人各得到一点利益，是否可以抵消一个人遭受巨大痛苦？如果可以，少数人的权利可能在总账里被平均掉。

贝卡利亚的自由硬闸在这里提供一种补充。最小自由让渡意味着，某些国家强制不是只要总收益更大就自动合法，还必须证明它属于共同安全所必要的授权范围。现代权利理论会用更成熟的语言处理，但结构相似：有些个人地位不是普通政策收益可以轻易买走。

这不意味着效用分析没有价值。恰恰相反，拒绝计算也会制造盲目痛苦。问题是总账必须带分布表：谁受益、谁受损、损失是否集中、是否可逆、是否有补偿、是否触及不可交易边界。边沁教我们不要浪费痛苦，贝卡利亚提醒我们有些痛苦即使“划算”也可能无权施加。

本编深描八　两阶段正当化：先问能不能罚，再问怎么罚

经过康德、贝卡利亚和边沁的对撞，本书可以提出一套很适合普通读者理解的“两阶段正当化”。第一阶段问：这个人是不是因为一个已经发生、法律预先规定并且可以归责的行为而进入处罚范围？这一步重点是罪刑法定、证据、责任、无罪保护，防止把人当政策工具。第二阶段问：即使可以罚，为什么必须用这种刑罚？有没有更轻替代？效果如何？副作用是什么？这一步重点是必要、比例、预防与最少痛苦。

这两阶段能解释许多现代争议。预测到一个人未来可能犯罪，第一阶段还没过，不能直接处罚；一个人已经被定罪，第二阶段仍然不能因此随便施加任何痛

苦。前者防止“没有罪也为了安全先管起来”，后者防止“有罪以后国家想怎么罚都行”。

贝卡利亚的重要性正是在两阶段之间。他没有使用今天这样清楚的术语，却同时追问处罚权来源和处罚效果。康德后来强化第一阶段，边沁后来强化第二阶段。现代刑法如果要真正继承三个人，不是把他们混成一个折中口号，而是让每一阶段都保持自己的举证责任。

◆ 五条正式改判

改判一：贝卡利亚不是“人可以被完全计算”的理论家。他追求的是公共可检验的规则与主要比例，不是对复杂行动的绝对数学穷尽。

改判二：自由主体作为规范起点不构成理论错误。发生学必须承认规范资格可以对经验差异保持优先，而不能用“有来历”取消人的基本地位。

改判三：贝卡利亚已经看见大量环境与反馈，“缺土、缺回写”的粗诊断撤回。仍可保留的只是反馈主要由理性立法者吸收，公众经验进入规则修正的制度路径相对薄。

改判四：“搭了看台，少了楼梯”从对象判罪降为现代续建判断。它不指责十八世纪思想家没有发明今天制度，而是提醒当代治理不能把总体理性封成免检高地。

改判五：本书对刑罚正当性的最终工作模型变成“双钥匙＋三道闸”：先有行为、法律与归责，才进入处罚资格；进入以后仍须再过权利来源、必要、比例和最少强制。任何一把钥匙都不能被另一把替代。

◆ 现实检验　三个场景看“双钥匙＋三道闸”怎样工作

场景一：校园暴力之后，能不能为了震慑“重罚一个典型”

一所学校发生严重欺凌，家长和网络舆论迅速要求“必须开除一个，才能让其他学生害怕”。从纯粹预防看，树立典型可能有效；但第一阶段必须先问，谁具体实施了什么行为，证据是否可靠，规则是否事先存在，年龄与责任能力怎样。如果只是因为某个学生名声差、以前惹过事，就把他当成最方便的样板，处罚已经在证据门前失败。

即使行为和责任都证明，第二阶段还要问开除是否必要。有没有修复、隔离、辅导、停课等较轻但有效的办法？处罚是否把未成年人永久推离教育？这里贝卡利亚会要求最少必要，康德式归责会要求不能抓错人，边沁式效果会要求不要制造比原问题更大的伤害。双钥匙不是让制度软弱，而是防止“为了大家好”省略对一个具体人的理由。

场景二：再犯风险很高的人，刑满以后能不能继续关

一个人已经服完依法判处的刑罚，风险模型认为他再次严重犯罪的概率远高于平均水平。为了安全继续限制他，看上去很符合预防逻辑，但第一阶段立刻遇到难题：未来犯罪尚未发生，旧罪刑罚已经结算。若单凭风险继续剥夺自由，就把“可能”变成新的处罚根据。康德式应得边界在这里非常重要。

这并不意味着社会什么都不能做。可以提供治疗、就业、住房支持，也可以在法律明确、比例严格、程序充分的特殊制度中处理真实而迫近的危险。但必须区分“帮助降低风险”和“把预测当成新罪”。贝卡利亚的罪刑法定在算法时代重新变得锋利：国家不能因为更会预测，就把未来提前判成已经发生。

场景三：冤案纠正以后，制度有没有真正回写

一桩案件多年后被证明错误定罪。最容易完成的动作，是撤销判决、给当事人赔偿，然后把个案归入“偶然错误”。真正的楼梯则要继续追问：错误证据为什么被相信？侦查激励是否偏向有罪结论？鉴定标准是否有缺陷？类似案件还有多少？规则需要怎样修改？如果个案纠正没有进入制度学习，回写只停在个人补偿层。

这正是本书“搭了看台，少了楼梯”最具体的含义。看台可以统计冤案率，楼梯则要求错误者的经验反过来改变证据规则、审讯程序、鉴定准入与救济机制。贝卡利亚的伟大是把刑罚置于理由之下；现代制度要继续一步，把理由置于纠错之下。一个不能从错误中改规则的法治，只是更整齐地重复错误。

◆ 理论深化　最强辩护之后的三次改判

第一，彻底撤回“精确计算主义”指控。贝卡利亚对比例的要求是主要分界与秩序，而非几何学精确；对证据的理解明确容纳概率；对人的行动又承认激情、习惯、处境和社会状态。以后再讨论算法量刑，不能把“贝卡利亚=算得越准越好”当起点。

第二，“主体先在”不再作为负面判词。社会契约中的自由个人首先是规范位置，功能是把国家置于限权关系中。经验主体当然会发生，但发生事实不能取消基本权利资格。新的问题只剩接口：经验差异怎样有限、透明地进入归责和政策，而不演化成身份歧视。

第三，“搭了看台，少了楼梯”保留，但从历史批判改成当代延伸诊断。贝卡利亚已经看见大量反馈，也不能要求十八世纪一本刑罚改革书预先写出二十一世纪参与式制度。所谓楼梯不足，只是提醒今天继续使用“公共幸福—理性立法”语言时，必须补上经验、异议、误判与实施结果持续回写规则的制度通道。

因此，本编之后的总定位式进一步收窄为：贝卡利亚建立了“权利来源限制＋公共幸福目标＋必要比例审查＋一般法律路径”的复合刑罚观；其现代延伸的最大

风险，不是他不知道人会变化，而是总体政策理性可能在缺乏充分反馈与权利硬约束时，把被治理者重新变成政策材料。康德与边沁分别从应得和效用两端，把这一风险继续显形。

◆ 压力测试　六个问题继续攻击当前判词

压力测试一：如果一个权利有历史来源，能不能因此被多数修改？不能直接推出。发生史说明它如何形成，不决定现行效力。修改仍要受更高层级的道德与法律理由约束。发生不是降级许可证。

压力测试二：如果预测模型证明某类人再犯率更高，能不能据此提前处罚？不能。预测可以配置支持、监督或资源，但处罚的核心仍需已发生行为、法定依据和个体归责。否则未来概率会被偷换成现在罪责。

压力测试三：如果重刑真的显著降低犯罪率，贝卡利亚会不会支持？仅有效还不够。他的结构还要通过权利来源、必要性、比例和最少强制。效果是一把钥匙，不是唯一钥匙。

压力测试四：如果报应性处罚对预防毫无帮助，还要不要罚？康德会把“应得”放得更前；贝卡利亚和边沁会更强地追问无用痛苦。现代制度的难题正是如何给应得设上限，又避免处罚变成纯政策工具。

压力测试五：如果公众参与导致更严厉、更情绪化的刑罚，楼梯还值得建吗？值得，但楼梯不等于直接多数决。它应包括公开理由、权利审查、证据规则、冷静期与专业评估。反馈的目标不是让激情接管看台，而是防止看台自我封闭。

压力测试六：如果SDE无法给出比贝卡利亚更清楚的规范边界，只能不断说“还有发生条件”，它有没有增加价值？没有。SDE只有在新条件能改变具体判断——例如区分可归责与应预防、区分可观察与可处罚、区分政策目标与不可交易权利——时才算升级，否则只是复杂化叙述。

第一，中心判断继续保持“自由限权型公共幸福中心”，但把“自由”进一步明确为规范硬约束，不要求它在经验上无来历。

第二，收回动作“搭了看台，少了楼梯”保留，强度降低：它不是对贝卡利亚历史任务的否定，而是对其理论在当代治理、特别是数据化治理中继续使用时的补充条件。

第三，康德提供了“不可把人仅当政策工具”的应得与归责闸门；边沁提供了“处罚本身是恶、无用痛苦必须节约”的效果闸门。现代刑罚正当性至少要同时过两关。

第四，SDE给自己的死法更加清楚：若它把“有来历”推成“没有规范优先性”，它会死于发生学谬误；若它把所有复杂性都塞进政策模型，却不能保护个体不被模型吞掉，它会死于工具化；若它只会增加变量而不能改变判断，它会死于无增量。

第五，本编之后，后续章节可以正式进入“贝卡利亚以后”的制度遗产：第65章讨论罪刑法定为什么比很多具体刑罚主张活得更久；第66章讨论死刑为何没有被一本书终结；随后转入AI时代，检验“更精确的看台”会不会制造新的任意。

> **带走的话**
>
> 现代刑法之所以长期同时使用“应得”和“效果”两套语言，不是理论不纯，而是国家处罚本来就必须回答两个不同问题：这个人为什么应当承担责任，以及国家为什么必须用这种方式承担。康德主要强化责任与应得的分配闸，边沁系统推进一般效用目的；贝卡利亚则在两问尚未完全分家时，先追问国家是否取得了这种处罚权，并把“国家必须说明理由”钉在了门口——这一追问集中出现在“权力归谁”与“定罪之前的人身”两类问题上，却在他自己的死刑例外处缺席。

> **小账**
>
> 这一章看见了什么？——现代刑法是双发动机：定罪偏应得，量刑执行偏效用。
>
> 接缝在哪里？——授权先问集中在权力归谁与定罪前人身，死刑例外处却缺席。

第 65 章

罪刑法定为什么比贝卡利亚活得更久

一本思想书最有意思的命运，是作者离开以后，书里有些东西迅速变成历史材料，有些东西却像被制度接走，进入了后来人的日常。贝卡利亚对具体刑罚形式的许多判断仍然需要争论，他关于法官解释的极端警惕今天也不可能原样照搬，可“没有事先的法律，就不能临时制造犯罪和刑罚”却越来越像现代刑法的一根骨头。为什么？

◆ 给中文读者的一块落脚石：罪刑法定与罪责刑相适应

写给中文读者，罪刑法定不能只停在十八世纪欧洲。我国现行《刑法》第3条规定：法律明文规定为犯罪行为的，依照法律定罪处刑；没有明文规定为犯罪行为的，不得定罪处刑。第5条又要求刑罚轻重与所犯罪行和承担的刑事责任相适应。这两条分别把“事先有法”和“罪责刑相适应”写进总则。

这里不主张一条简单的“贝卡利亚直接影响中国刑法”谱系。更稳妥的说法是：现代成文刑法在不同法律传统中都需要处理他曾经逼出来的那组问题——国家不能在看见具体的人以后临时发明犯罪，刑罚轻重也必须接受可公开的责任与比例约束。把贝卡利亚读到这里，中文读者不必把他当成遥远的欧洲古董；他的提问已经进入我们每天使用的刑法结构。

第一层答案很简单：罪刑法定不要求后人接受贝卡利亚全部哲学。你可以是功利主义者，也可以是报应论者；可以强调个人自由，也可以强调公共安全；甚至可以在许多政治问题上彼此对立，但只要承认国家刑罚是最强烈的公共强制之一，就会理解为什么这种强制不能在看见一个具体的人以后才临时发明。罪刑法定之所以耐久，不是因为它解决了所有刑法问题，而是因为它把不同理论都必须经过的一条入口固定下来。

第二层答案更深：它保护的不是“犯罪人特权”，而是所有人的行动可预期性。一个人早上出门时，不可能知道下午会遇见哪个法官、哪个检察官、哪一种舆论。如果国家可以在行为发生以后再说“你当时其实已经违法”，那么人无法用法律安排自己的生活。真正的自由不是永远不受限制，而是限制的边界在行动之前已经足够公开、清楚和可查。

这也是为什么罪刑法定比很多具体刑罚主张更容易跨时代。具体刑罚会随着社会条件改变：监禁的形式、财产刑的范围、社区矫正的技术、未成年人制度都会变化。可“先有规则，再有处罚”的方向并不依赖某一种刑罚工具。它更像规定了道路的铺设程序，而不是规定所有车辆永远只能是马车。

从SDE角度看，这是一种“路径遗产”。很多思想家最容易把遗产理解成一个结论S：例如“死刑不正当”“某种刑罚更有效”。结论一旦遇到新的经验、技术和制度，容易被修正。路径D则可能更耐久：谁先说话、理由何时公开、什么事实可以进入、谁承担举证、什么程序允许强制发生。贝卡利亚最耐久的贡献，恰恰不是替后世写好所有答案，而是改变了答案取得资格的路径。

这里可以做一个普通生活的比喻。一个家庭当然可以不断修改家规，可最让孩子无所适从的不是“规则会变”，而是“做完以后才告诉你原来不能这样做”。同样，学校可以修订考试纪律，公司可以调整请假制度，城市可以改变交通规则，但规则改变必须有公布、过渡和一般适用。否则管理者拥有的不是规则权，而是事后定义错误的权力。

当然，罪刑法定并不意味着文字能够把现实一次写完。第16章已经讨论过，语言进入具体案件必然需要解释。真正要守的，不是“法官一字不能解释”，而是解释不能把原本没有进入犯罪边界的行为偷偷拉进来，不能借模糊词无限加重刑罚。现代法治比贝卡利亚更承认解释，但也更清楚为什么刑法解释必须受到文本、体系、可预见性和禁止不利类推的约束。

这正好说明一种成熟思想的命运：后人不必保留它最僵硬的外壳，却会保留它逼出来的制度问题。贝卡利亚当年把危险主要看在法官个人意志上，所以希望法律

尽可能明确、法官尽可能机械。今天我们知道任意也可能来自立法、行政、舆论、数据模型甚至自动化系统，于是罪刑法定的意义反而扩大了：任何能够决定犯罪边界的机制，都必须回答它的规则从哪里来、何时生效、普通人能否知道、能否质疑。

AI时代尤其能看出这一点。假如一个模型在几百万个历史案件中学出“高风险行为模式”，但法律从未把这些模式规定为犯罪，模型不能因为预测准确就创造新的犯罪类型。数据里的相关性不是刑法里的构成要件。即使一个行为与将来犯罪高度相关，它仍不等于现在已经违法。机器可以发现模式，却不能凭模式改写处罚边界。

所以，罪刑法定比贝卡利亚本人活得更久，是因为它把一个哲学要求变成了制度闸门：国家不能先看见人，再决定规则；不能先想处罚，再回头寻找罪名；不能把预测、道德讨厌或公共恐惧直接升级成刑罚。只要国家仍然拥有处罚权，这道闸门就不会失去意义。

◆ 中文研究中的贝卡利亚：从经典接受到威慑再检验

中文学界对贝卡利亚的阅读也不只停在“刑法之父”这一标签。施鹏鹏从刑事诉讼角度梳理无罪推定、程序理性与公开审判；张智把《论犯罪与刑罚》放进苏格兰启蒙和“人的科学”背景中，提醒读者注意效用、正义与人道之间的复杂关系；谷桢则直接以严厉性、确定性和及时性重新分析威慑机制。

把这些研究放进本书，可以看到中文读者面对的并不是一位只需纪念的十八世纪先驱，而是一组仍在中国刑法和刑事程序中工作的难题：规则能否事先知道、强制能否与责任相称、程序能否抵抗错误事实、威慑能否用经验检验。也因此，本书不主张建立一条简单的“贝卡利亚—中国刑法”直接影响线，而是把他作为现代刑罚问题的一处历史源头来读。

带走的话

贝卡利亚最耐久的遗产，不是某一个刑罚结论，而是“先有公开规则，后有国家强制”的发生路径。结论会变，路径决定权力能不能合法出生。

小账

他看见了什么？——没有事先的法律，就不能临时制造犯罪和刑罚。

还要追问什么？——立法、行政、数据模型划定犯罪边界时，规则从哪里来。

第 66 章

死刑为什么没有被一本书杀死

如果思想真的像开关，1764年《论犯罪与刑罚》出版以后，欧洲的死刑就应该很快消失。贝卡利亚已经提出了权利来源、必要性和效用三重追问：个人是否把生命交给国家？死刑是否真的非用不可？它是否比长期、持续的刑罚更有预防效果？这些问题后来影响深远，全球废除死刑的范围也不断扩大。联合国人权事务委员会2018年第36号一般性意见（CCPR/C/GC/36，生命权）长期把全球趋势概括为越来越多国家在法律上或事实上放弃死刑。可直到今天，死刑仍未成为全世界的历史名词。为什么一本极有力量的书没有“杀死”死刑？

◆ 思想进入制度：至少要经过出版、翻译、政治机会与立法选择

“一本书没有杀死死刑”必须拿历史事实来写。1786年11月30日的托斯卡纳《利奥波德刑法典》废除死刑与酷刑。佛罗伦萨国家档案馆关于这部法典的专题明确指出，废除死刑的论证直接吸收了贝卡利亚的论证。这是思想进入立法的强例，却不是“书一出版，制度自动改变”：它还需要统治者的政治选择、法典编纂与既有实践条件。

俄国的例子更能说明非线性。叶卡捷琳娜二世1767年的《圣谕》广泛吸收孟德斯鸠和贝卡利亚式改革语言，试图为立法委员会提供原则，但委员会并没有完成一部新的法典。思想可以进入官方语言，却不一定立刻变成稳定制度。

奥地利哈布斯堡改革也推动了刑罚制度变化；到1787年的约瑟夫二世刑法改革，死刑在普通刑法中被大幅排除。不同地区的改革节奏说明，贝卡利亚不是一根单独的因果杠杆，而是进入了更大的行政、启蒙与国家建构网络。

还要再加一个更有挑战性的事实。Peter Garnsey 2026年的研究指出，佛罗伦萨官员Giuseppe Pelli在贝卡利亚之前已经写过一篇未完成的反死刑论文，甚至可称同类最早作品之一，却长期没有发表。这恰好支持本章主题：思想能否改变制

度，不只取决于“谁先想到”，还取决于能否完成文本、出版、流通、被翻译并进入有权修改制度的人手里。

因为制度从来不是由一个理由单独长出来的。死刑之所以顽固，不只是有人没读懂贝卡利亚，也不只是政治落后。它同时纠缠着报应、威慑、安全、受害者感受、宗教与道德传统、国家权威、战争与暴力记忆、媒体事件、公众恐惧以及司法制度对自身正确性的信心。贝卡利亚击中了其中几根承重梁，却不可能凭一篇论证把整片土一起移走。

康德的反对尤其说明这一点。对贝卡利亚来说，国家如果不能证明死刑来自合法授权并且确有必要，就不应杀人；对康德来说，刑罚首先不能只问“对未来有没有用”，而要问行为人是否因为自己的行为应当承担相称后果。两人不是在同一个计算器上争一个数字，而是在争刑罚正当性的起点。只要“应得”与“预防”这两种直觉同时存在，死刑问题就不可能仅靠效率数据结束。

公众情绪又加入另一层。严重暴力犯罪发生以后，人们要求最重处罚，往往不仅是为了降低下一次犯罪率，还在表达“有些行为不能被普通方式对待”。这种表达包含愤怒、哀悼、秩序修复和道德谴责。贝卡利亚会提醒，国家不能把愤怒直接变成刀；但如果理论完全不理解这种情感，只会说“统计上不一定更有效”，它也无法解释为什么死刑在一些社会中拥有持续的象征力量。

制度惯性同样重要。一套刑罚制度一旦建立，法律条文、司法程序、警察实践、检察政策、政治语言和公众想象会相互支持。废除死刑不是删掉一个词，而是要重新回答最严重犯罪如何表达谴责、如何隔离危险、如何保障被害者、如何防止终身监禁变成另一种无尽残酷。旧制度存在替代成本，新制度也要证明自己能承载原制度曾经承担的功能。

最关键的一点，是贝卡利亚真正改变的并不是结果，而是举证责任。过去，死刑可以借传统、主权或报复直觉当然存在；经过他的追问以后，保留者越来越需要说明：为什么必须是死刑？为什么其他刑罚不够？如何排除误判？如何证明威慑？如何解释不可逆性？这并不保证所有社会都会给出同样答案，却改变了讨论结构。

这里可以看出思想影响制度的真实方式。哲学书很少像命令一样直接造成制度转向，它更常做三件事：改变哪些问题必须回答，改变哪一方承担证明，改变哪些旧理由不再够用。贝卡利亚没能让死刑在十八世纪立即消失，但他让“因为国家一直都这样做”越来越难成为充分理由。

SDE在这里也必须管住自己的夸张欲。我们不能说“贝卡利亚的思想发生以后，废死就必然发生”。发生不是单线因果。一个理念进入社会，会遇到宗教、战争、阶层、司法能力、政治竞争和公共情绪的纠缠。相同的文本落在不同土里，会长出不同制度结果。思想只是其中一股力，不是唯一发动机。

死刑因此成了理解“回写”最好的例子。贝卡利亚影响后世法律与公共讨论，新的制度经验又反过来改变人们阅读贝卡利亚的方式。冤案、DNA证据、长期监禁、人权法、受害者权利、恐怖主义和大规模暴力事件不断向旧论证加入新条件。一本书没有终结问题，正说明它进入了一个持续回写的历史。

从本书的母题看，这一章也替“公共幸福中心”加了一道限制。公共幸福不能只理解为总体犯罪率下降，因为刑罚制度还承载着正义感、权利边界与不可逆风险。任何以“大家更安全”为名的死刑论证，都必须经过个体权利和必要性审查；任何以“应得”为名的死刑论证，也必须面对误判、制度不平等与执行后果。没有任何一把钥匙可以独自开门。

> 带走的话
>
> 贝卡利亚没有用一本书终结死刑，他更深的成功是改变死刑的举证结构——从“国家当然可以杀”转向“国家必须不断证明为什么非杀不可”。

小账

他看见了什么？——死刑要先问国家有没有这种权，再问它有没有用。

接缝在哪里？——书只改了举证结构，废死还要过出版、翻译、政治选择与立法。

第 十 三 编

AI时代，贝卡利亚最危险的愿望可能真的实现了

编　序

本编六章，把贝卡利亚放进算法时代。第67章分开预测与罪责；第68章问量刑能否被计算；第69章问他会不会喜欢算法法官；第70章看确定性走到极端的样子；第71章问预测是在预防还是在制造罪犯；第72章讨论必要性变成一个数字以后怎么办。

这一编是判决五的主场。模型给出分数，并不免除处罚者解释的义务；风险高，也不等于强制者已经举证；模型上线后还会改变自己未来的数据，制度必须能审计、停用，并补救过去受它影响的决定。

读这一编，可以带着日常问题进去：学校给学生做违纪风险分级，高风险的孩子得到的是辅导，还是重点盯防？人脸识别抓到了更多逃犯，是不是就越多越好？

第 67 章

如果机器真的可以预测犯罪风险

贝卡利亚希望刑罚更理性、更确定、更少依赖法官的脾气。两百多年以后，机器学习似乎递来一个诱人的礼物：既然人的判断有偏见，就让模型读更多数据；既然经验会忘记，就让系统记住几百万个案件；既然法官难以估计再犯风险，就让算法给出概率。问题突然变成：如果机器真的能够比人更准地预测谁更可能犯罪，我们是不是离贝卡利亚的理性刑法更近了？

先把最容易混淆的两件事分开：预测未来，不等于证明过去。刑事责任首先处理已经发生的行为。证据链要回答的是“这个人是否实施了这个行为、是否具备相应责任条件”。风险模型回答的却是“在某些数据条件下，他将来发生某类行为的概率有多大”。这两个问题可以同时重要，但不能彼此偷换。

一个极端例子最清楚。假设模型对某人未来两年实施暴力犯罪的预测准确率高达90%。这仍然不能证明他今天已经犯了罪。概率很高，只说明未来风险高；未来尚未发生，不会因为模型自信就变成过去事实。贝卡利亚的罪刑法定在这里突然获得了新的用途：国家不能因为更会预测，就把“可能”提前判成“已经”。

美国威斯康星州的State v. Loomis案经常被放在这个问题中央。2016年，威斯康星州最高法院允许量刑法院在一定限制下参考COMPAS风险评估，但明确强调风险分数不能用来决定是否监禁、不能用来决定刑罚严厉程度，也不能成为判断能否在社区安全监督的决定性因素；法院还要求量刑理由必须有风险评分以外的独立根据。这个判决的意义并不是“算法已经被法律认可”，而是恰好显示法律在努力把模型关进一个辅助位置。

欧盟《人工智能法》走得更明确。Regulation (EU) 2024/1689第5条禁止仅仅根据个人画像、人格特征或特性，对自然人实施犯罪风险评估，用来评估或预测其实施犯罪的风险；只有在已经存在与犯罪活动直接相关、客观且可验证事实的基

础上，AI作为人类评估的支持工具，才不落入这一绝对禁止的同一范围。这个规则抓住了一个核心危险：把“你像哪一类人”变成“你可能会犯罪”。

为什么这种危险比普通预测更大？因为刑事系统拥有强制权。电商预测你喜欢什么商品，错了最多推荐得烦；警方、法院预测你危险，错误可能导致更密集监控、搜查、羁押或更少自由。一旦预测与国家强制连接，模型的每一个输入变量都必须接受权利审查。

从SDE看，风险模型最容易犯的是“把差异序列D当成既成显露S”。模型从历史数据中看到一条路：某些特征组合与某些结果在过去经常一起出现。可它输出一个“高风险”标签以后，这条路就被压成一个人的定型身份。更危险的是，定型标签会回写：警察更关注他，单位更谨慎，社区更排斥，他获得的资源更少，新的数据再证明他“高风险”。如果制度不检查这条回路，预测就可能参与制造自己后来声称只是发现的现实。

这不意味着一切风险预测都应禁止。医疗也预测风险，消防也预测火灾，交通也预测事故。关键区别在于：预测进入什么决定、对谁造成什么负担、有没有独立事实基础、能否解释和纠错。一个城市根据道路事故数据增加某路段照明，与根据某个青年家庭收入、住址和朋友圈把他列为潜在罪犯，不是同一件事。前者主要配置公共资源，后者把风险贴到一个具体人的自由上。

欧洲委员会近年来关于人工智能与人权的材料也强调，预测性警务若影响逮捕或拘留，就必须回到可验证、客观且与犯罪活动直接相关的事实，并满足合法、必要、比例和保障要求。这里重新出现的，正是贝卡利亚熟悉的四个词：法律、事实、必要、比例。二十一世纪没有把十八世纪的问题消灭，只是把法官的直觉换成了模型的概率。

所以，机器越会预测，我们越要把“预测”和“处罚资格”隔开。模型可以告诉制度哪里需要更多预防资源、哪些人可能需要帮助、哪些释放方案需要更多支持，但它不能凭自己创造罪责。未来概率可以进入风险管理，不能偷偷升级成过去责任。

带走的话

预测得更准，不等于有权罚得更早。未来概率不会因为算法精确就变成已经发生的罪。

小账

这一章看见了什么？——预测未来不等于证明过去，90%的风险也不是已犯的罪。

接缝在哪里？——高风险标签回写，监控与排斥制造出模型声称发现的现实。

第 68 章

如果量刑真的越来越可计算

风险预测之外，AI还有另一条更直接的诱惑：量刑能不能越来越可计算？同样的案件不同法官判得差很多，人们会觉得不公平；同一个法官今天心情不同、经验不同，也可能影响判断。如果模型能够读完全部相似案件，给出一个“通常刑期区间”，看上去正好实现贝卡利亚追求的稳定与比例。

先承认它的真实价值。数据工具可以帮助发现异常差异：相似案件为什么某地普遍更重？某类被告是否在相同事实下承担更高刑期？某个法官的量刑是否长期偏离整体？这些统计不是敌人。恰恰相反，它们可能把过去隐藏在“司法经验”里的差异照亮，让制度第一次看见自己的不一致。

问题从第二步开始：一旦数据从“照镜子”变成“替你决定”，发生了什么？一个量刑模型必须选择变量。犯罪金额算多少权重？前科算多少？年龄要不要算？就业是否算？认罪是否算？家庭责任是否算？每一次选择都不是纯数学，它背后都有价值判断。模型给出最后的数字时，这些价值判断不会消失，只是被压进公式。

所以，计算最容易制造一种“数字客观性”：因为结果带小数点，就显得比人的语言更真。可数字只能精确表达被放进去的关系，不能自动证明这些关系值得被放进去。一个把贫困、住址或教育程度当成再犯风险代理变量的模型，也许统计上提高准确率，却可能把社会不平等重新包装成个人危险。精度和正当性是两条轴。

这也是为什么我们在前文撤回了“贝卡利亚迷信数学精确”的批评。他本人已经承认人的激情、利益和政治关系无法像几何学那样精确安排。他追求的是轻重有序、规则可预期，而不是让每个人的刑期由一台理性计算器直接吐出。AI时代真正危险的，不是把贝卡利亚做得更彻底，而是把一个他自己都没有主张的“完全可计算刑罚”反过来冠到他头上。

量刑里至少有三种问题不能被同一个数字吞掉。第一，事实问题：行为造成了什么后果、行为人是什么角色。第二，规范问题：这些事实为什么应当增加或减少责任。第三，政策问题：这种刑罚是否有必要、是否能实现预防和复归。算法可以帮助整理第一类，也可以提供第三类的经验数据，但第二类包含“为什么应当”的理由，不能因为历史案件曾经这样判就自动获得正当性。

历史数据还有一个更隐蔽的麻烦：它把过去带进未来。假如过去某类地区的被告更容易被判重，模型学习“相似案件”的时候就可能把这种差异当成常态。人类法官至少可能被新的规范说服，模型如果只是拟合过去，则会把旧偏差做成新标准。机器不是没有历史，它的历史往往就藏在训练集里。

因此，合理的算法量刑只能处在辅助位置：帮助查找同类案件、显示分布、提示异常、计算法定幅度、提醒遗漏因素；真正决定刑罚的规范理由必须能够被公开说出来。法官不能只写“系统建议三年两个月”。他要解释为什么这些事实进入了法律标准，为什么这个刑期在权利、责任与必要之间成立。

这又回到贝卡利亚的老愿望：让同样的人面对同样规则。AI可以帮我们发现“其实并没有同样”，也可能帮我们减少随机差异。但如果为了同样而抹掉真正相关的不同，它会制造另一种不公。真正的比例不是人人同一数字，而是相关差异得到相关对待，不相关差异被排除。

所以，“可计算”不是敌人，“计算替代理由”才是。现代刑法可以使用更好的数据，却不能把价值判断藏进模型后说“机器算的”。任何会影响自由的数字，都必须能被翻译回公共语言：用了什么变量、为什么相关、误差多大、谁承担错误、有什么替代方案、当事人怎样反驳。

> **带走的话**
>
> 算法可以让刑罚更一致，却不能让刑罚自动更正当。数字能压缩判断，不能替判断免除理由。

小账

这一章看见了什么？——量刑模型挑选变量就是价值判断，只是被压进了公式。

接缝在哪里？——历史训练数据把旧偏差带进未来，做成新的量刑标准。

第 69 章

贝卡利亚会喜欢算法法官吗

如果只读贝卡利亚反对法官自由解释的那些段落，很容易做出一个有趣的想象：他大概会喜欢算法法官。机器不喝酒、不发脾气、不认识贵族、不因为被告说话难听就加刑；法律是大前提，事实是小前提，程序自动推出结果——这不就是他梦想的“三段论法官”终于获得了硬件吗？

这个想象只对了一半。贝卡利亚确实会被“减少个人任意”吸引。一个公开、固定、可验证的计算工具，如果只是帮助按照事先法律进行机械换算，例如罚金倍数、法定期限、时间折抵，它非常符合他的制度气质。机器在这里像尺子，价值在于人人用同一刻度。

但现代AI法官并不是普通尺子。机器学习模型往往不是“如果A则B”的显式规则，而是从大量数据中形成复杂关联。输入怎样编码、训练样本怎样选择、标签怎样产生、阈值怎样设定，都会影响结果。表面上法官个人意志减少了，背后却可能出现设计者、数据提供者、机构采购者和历史样本的集体意志。任意没有消失，只是从法庭正面搬进了系统后台。

这一点对贝卡利亚尤其致命，因为他的法治理想高度依赖“公民可以事先知道法律”。如果决定自由的规则藏在无法解释的模型参数里，即使输出稳定，也不满足他的可知性要求。人能够知道法条，却不知道系统怎样把自己评成高风险；能够看到结论，却无法检验中间道路。这样的稳定，是黑箱的稳定。

State v. Loomis案之所以持续被讨论，也正因为这种张力。COMPAS具有专有、封闭的算法部分，被告难以完全检验模型内部方法。威斯康星州最高法院并没有因此宣布一切风险工具非法，而是用限制条件把它压到辅助位置。这个做法很“贝卡利亚”：不是拒绝技术，而是不允许不可见工具成为决定自由的唯一理由。

算法法官还有一个传统法官没有的特殊问题：规模。一个有偏见的法官可能伤害一批案件，一个有系统偏差的模型可以在几秒钟内把同一种偏差复制到成千上万个决定。机器减少了随机任意，却可能放大结构任意。稳定不是天然优点，错误也可以非常稳定。

所以，如果我们真的把贝卡利亚请到今天，他大概会同时问机器六个问题：规则在行为之前公布了吗？普通人能知道影响自己的标准吗？模型只是辅助，还是实际决定？错误能够被具体指出和纠正吗？同样的人是否真的被同样对待？系统背后的人能不能用“是机器算的”逃避责任？这些问题比猜他“支持AI还是反对AI”更有意义。

从SDE看，“算法法官”揭露了器具可消性的新版本。过去，人们希望把法官个人差异消掉，于是得到一个仿佛纯粹的法律适用过程；今天，人们希望把算法当作同样对所有人工作的中性器具，于是又容易把器具背后的训练土、目标函数、评价标准消掉。器具越统一，越容易让人忘记器具参与了结果发生。

真正成熟的技术法治不是追求“没有人”，而是追求“权力可定位”。人类法官要署名、给理由、接受上诉；算法系统也需要版本、数据来源、功能边界、审计记录、责任主体和救济机制。机器参与多少，责任就要显露多少。不能因为输出来自服务器，就让权力失去地址。

因此，贝卡利亚可能喜欢的不是“算法替代法官”，而是“算法也被法律驯服”。一台能够公开规则、限制功能、留下记录、接受质疑的工具，可以帮助减少任意；一台以效率为名把理由、责任和反驳全部藏起来的机器，恰恰会恢复他最反对的那种不可见权力。

> **带走的话**
>
> 贝卡利亚未必会迷恋没有人的法官；他真正要的是没有免检权力的法官。机器若获得免检，算法只是穿上白大褂的任意。

> **小账**
>
> 这一章看见了什么？——算法减少随机任意，却可能在后台放大结构任意。
>
> 还要追问什么？——系统版本、数据来源、责任主体能否定位，权力有没有地址。

第 70 章

“确定性”走到极端，会变成什么

贝卡利亚最常被引用的判断之一，是刑罚的确定性往往比刑罚的严厉性更能形成预防。与其偶尔用极刑制造震撼，不如让违法者知道违反规则后果更稳定、更及时。这个判断直到今天仍有巨大吸引力。可是AI、摄像头、传感器、身份识别和大规模数据分析把一个新问题推到他面前：如果“确定性”真的被做到极端，会发生什么？

想象一座城市，每一次闯红灯都被识别，每一次逃票都被发现，每一次违规停车都自动扣款，每一次进入限制区域都立即留下记录。对传统执法来说，这几乎是梦：规则不再因为人手不足而偶尔执行，同样行为面对同样后果，选择性执法也可能减少。

可再往前一步，城市为了让犯罪“无处可逃”，把人脸、位置、支付、通信、车辆、社交关系全部连接起来。此时确定性不再只是“犯了罪更可能被查到”，而变成“每个人时时处在可被识别、计算和关联的状态”。预防收益增加了，隐私、匿名行动和不被持续观察的自由也开始付出价格。

这说明确定性不是单一好处，而是一条成本曲线。贝卡利亚反对的是“低确定性＋高残酷性”的旧刑罚：抓不到多数人，只能用少数人的痛苦制造恐惧。现代技术可能把它变成“高确定性＋低单次强度”，表面更温和。但如果高确定性依赖无处不在的监控，总体强制未必更小。

最简单的家庭例子也能看出边界。父母如果说“晚上十点后不要玩游戏，违反就第二天少玩半小时”，规则清楚、后果稳定，孩子容易形成预期；但如果父母为了保证百分之百发现违规，在房间、手机、手表、书桌全部安装监控，问题就不再只是规则执行，而是生活是否还存在私人空间。确定性从秩序工具变成环境结构。

因此，我们要把贝卡利亚的“确定性”重新翻译成现代语言：它不是“国家知道一切”，而是“法律后果不应任意地只落在少数倒霉者身上”。真正值得追求的是执法的一致、公平、及时与可预期，而不是全知。前者限制国家，后者可能无限扩张国家。

这一区分在技术时代格外重要。一个系统如果通过更好的案件管理提高报案处理率，是提高确定性；如果通过全体居民持续生物识别来提高发现率，则还要回答必要性和比例。目标相同，手段的权利成本完全不同。贝卡利亚的“最少强制”正好可以拿来限制他自己的“更高确定性”愿望。

也就是说，贝卡利亚理论内部其实有一套自我制动器。确定性有利于预防，但任何提高确定性的措施都不是自动正当；它仍要经过权利来源、必要性、比例和较轻替代方案。只要把这几道闸门一起读，就不会得到“为了让违法无处可逃，国家可以知道一切”的结论。

从SDE的三界看，极端确定性还会回写自我界。人知道自己始终被观察，会改变说话、交往和探索方式。有些改变是守法，有些则可能是自我审查。一个制度不能只统计“违法率下降多少”，还要看它塑造了怎样的主体、怎样的社会关系。否则现实界的监控器具会悄悄重写自我界，却不进入效果评估。

所以，第70章给“确定性”加上一条现代限制：越接近技术上的百分之百发现，越不能只问有效率多高，还要问这种确定性是怎样获得的。法治要的是可预期，不是全透明的人；要的是规则稳定，不是生活无死角。

> **带走的话**
>
> **刑罚确定性真正反对的是选择性和任意，不是要求国家成为全知者。确定性若以全面可见为代价，就会从限权工具反转成扩权工具。**

小账

这一章看见了什么？——确定性走到极端，要求每个人时时可被识别与关联。

接缝在哪里？——违法率下降时，持续观察带来的自我审查不进效果评估。

第 71 章

预测犯罪是在预防犯罪，还是在制造“罪犯”

贝卡利亚说，预防犯罪比处罚犯罪更好。放到今天，这句话几乎会自然地把我们带向预测：既然要在犯罪发生之前行动，就应该尽量知道犯罪可能在哪里、由谁、在什么条件下发生。问题是，一旦预测的对象从“风险环境”变成“具体的人”，预防与制造标签之间的距离会突然变短。

先区分两种很不一样的预测。第一种预测地点和条件：某些路段夜间照明差、某类诈骗在特定时间上升、某些交易模式可能意味着洗钱。制度据此增加照明、提醒公众、配置巡逻或加强审核，主要改变环境。第二种直接预测个人：这个人因为住址、年龄、朋友圈、历史接触和行为模式，被打成“高风险”。后一种更容易触碰自由，因为它把未来概率贴到一个人的身份上。

为什么标签会制造现实？假设某地区历史上警力本来就多，于是发现和记录的违法也更多。模型学习这些数据以后，判断该地区“高风险”，于是继续投放更多警力；更多警力又发现更多违法，新数据进一步证明模型正确。这里并不一定有人作弊，每一步都可能“基于数据”，但系统形成了一个反馈环。被观察得更多的人，会产生更多被记录的数据。

个人层面也是如此。一个青年被列为高风险后，可能受到更频繁盘查、更难获得某些机会、在社区中被视为危险。这些反应会改变他的生活路径。模型起初声称只是描述概率，制度却通过自己的回应把概率变成了环境。SDE所谓回写，在这里不是抽象哲学，而是可以发生在数据库、巡逻路线和个人机会之间的具体循环。

所以，预测性警务最危险的误区是“只要不处罚，就没有权利问题”。事实上，持续监控、重点盘查、名单管理、机会限制都可能实质改变个人自由。即使这些措施不叫刑罚，也要接受合法性、必要性、比例和程序保障的审查。欧洲委员会关于AI与人权的材料特别强调，若预测系统的信息被用于支持逮捕或拘留，必须满足基于可验证客观事实的合理怀疑要求，不能让黑箱概率代替事实。

欧盟《人工智能法》对基于个人画像和人格特征的犯罪风险预测设置禁止，也体现同一条边界：不能因为一个人“像统计上的高风险者”，就把他当成潜在犯罪人。当然，法律并没有因此禁止所有风险分析。企业欺诈、海关路线、已存在客观犯罪事实基础上的辅助评估，仍可能有空间。关键不是“预测”两个字，而是预测对象、依据和后果。

这一章也逼我们重新理解贝卡利亚的预防主义。他说预防优于惩罚，并不等于“越早干预越好”。如果所谓预防本身已经像处罚一样限制自由，它必须被同样严格地问：你根据什么？你要防什么？错误由谁承担？当事人能不能知道并纠正？还有没有更轻的环境性措施？

真正贝卡利亚式的预防，往往不是盯住“危险的人”，而是减少让犯罪更容易发生的条件。他重视教育、清楚法律、减少不必要禁止、提高制度可信度，这些都在改“土”。它们与把某些人提前锁进风险身份，不是同一条路。一个是让环境更少生产犯罪机会，一个是把未来罪责提前压到个人身上。

这也解释了为什么“风险身份”特别危险。犯罪身份至少以已经证明的行为为基础，还有刑期终点；风险身份却可能没有明确终点，因为未来永远没有彻底到来。今天没犯罪，可以被解释成“干预有效”，也可以被说成“仍要继续观察”。一个无法通过不犯罪来退出的标签，会变成自我封闭的制度。

所以，预防与制造罪犯之间要插入一道硬线：制度可以预测风险，但不能让风险本身变成罪责；可以配置帮助，但不能让帮助伪装成无限强制；可以关注环境，但不能把社会问题全部压缩成某些人的危险人格。贝卡利亚当年把刑法从“坏人的灵魂”拉回行为与社会损害，AI时代不能再把灵魂换成风险分数。

> **带走的话**
>
> **预防犯罪最危险的捷径，是把“哪里可能出问题”改写成“谁本身就是问题”。风险可以被管理，风险身份不能自动变成罪责。**

> **小账**
>
> 这一章看见了什么？——警力多处记录多，模型据此派更多警力，形成反馈环。
>
> 接缝在哪里？——风险身份没有终点，今天不犯罪也无法退出这个标签。

第 72 章

算法第一次把“必要性”变成一个数字，怎么办

第9章我们第一次问过：谁有权宣布“必要”？到了第72章，这个问题终于返回，而且换了一张脸。过去，法官、立法者或行政者说“为了安全，这是必要的”；今天，算法可能说：“模型显示，如果不采取措施，风险为78.6%。”一句模糊的必要，突然变成一个带小数点的数字。它更可信吗？

数字当然可能让讨论进步。过去说“这个人很危险”，到底有多危险没人知道；模型至少要求定义结果、回看数据、检验准确率。过去说“多巡逻有效”，也许只是经验印象；数据可以比较不同措施的效果。把一部分判断量化，会迫使制度暴露依据。这是AI可能给贝卡利亚理性主义的最好礼物。

但数字不能自己完成“必要性”。必要性至少包含一个目标、一个事实判断、一组替代方案、一种权利成本和一个价值选择。比如“若不安装全城人脸识别，某类犯罪风险为8%；安装后降到5%”。这两个数字仍然没有回答：降低3个百分点是否值得让所有人的面部信息被持续处理？有没有只在高风险地点增加照明和巡逻的办法？误识别由谁承担？数据能否被转作别用？

也就是说，概率只回答“可能发生什么”，必要性还要回答“为了避免它，我们可以对谁做什么”。前者属于事实和预测，后者属于规范授权。把78.6%直接变成“必须限制自由”，就是把事实语言偷换成权利语言。数字越精确，这个偷换越容易被忽略。

因此，我建议把现代必要性审查压成五步。第一步，目标显形：究竟要防什么具体损害，不能只说“安全”。第二步，事实显形：风险依据是什么，数据质量、误差和适用范围如何。第三步，替代显形：有没有同样有效但侵害更小的办法。第四步，权利成本显形：谁失去什么，错误是否集中落在特定群体。第五步，回写显形：措施实施后如何复查，什么结果会触发修改或停止。

这五步里，算法最多直接帮助第二步，也可以部分帮助第三步的效果比较，却不能独自替代第一、第四和第五步。目标本身是公共价值选择；权利边界不能由预测准确率生成；反馈机制更需要制度设计。一个模型如果把自己擅长的“预测”扩张成“授权”，就越过了工具边界。

这正好回应贝卡利亚最初的必要性原则。十八世纪的危险是君主或法官说“我觉得必须”；二十一世纪的危险可能是机构说“模型证明必须”。两句话的语气不同，结构却可能一样：判断者把自己的中间步骤藏起来，只把最后结论交给被治理者。真正的限权，不是把“我觉得”换成“机器觉得”，而是让必要性变成可拆、可查、可反驳的过程。

这里出现一个新的“收回动作”候选，我们可以暂时叫它“数字盖章”。当政策还处在正在成形阶段，本来需要讨论目标、替代、权利和不确定性，系统却用一个风险分数把争论提前定型：高风险=必须干预，低风险=可以放过。这个动作和我们早先撤回的“量刑尺”批判不同。我们不是说所有数字都错，而是说数字不能把规范争议伪装成已解决。

“数字盖章”也必须有推翻条件。如果一个算法系统只是提供事实估计，决定者公开说明价值判断，权利边界由法律预先规定，当事人可以质疑数据和结果，系统还有持续审计和停止机制，那么数字就没有盖章，它只是工具。只有当概率被当成授权本身、当中间规范步骤被隐藏时，才构成我们要批评的收回。

这一章最终把全书又送回贝卡利亚的起点：国家凭什么罚你？AI没有替我们回答，只是把“凭什么”里的证据和预测部分做得更强。权力来源、个体责任、必要性、比例、最少强制仍然需要人类制度承担。机器能够改变事实判断的精度，却不能替社会签署权利让渡书。

如果说贝卡利亚的时代需要把刑罚从君主的意志里拉出来，那么AI时代需要再做一次类似工作：把算法从“自然答案”的位置拉回制度工具。模型有出生日期，有训练数据，有目标函数，有使用场景，也有误差。把这些重新放回土和路，我们才不会把一个发生出来的技术结论又包装成“机器发现的必然”。

◆ 专题深描　八个制度现场：当“理性刑罚”遇见算法

专题一　同一个风险分数，放在不同程序里，法律意义完全不同

风险工具最容易被误解成一种“万能事实”。同一个70分，似乎走到哪里都还是70分，于是人们会忘记它进入不同程序以后，法律意义已经变化。审前阶段的核心问题，是有没有足够理由暂时限制自由、如何确保到庭和防止现实危险；量刑阶段处理的是已经定罪以后承担什么后果；假释或矫正阶段又涉及监督资源和复归条件。模型即使计算的是同一种“再犯风险”，也不能自动跨越这些程序边界。

原因在于，程序不是装数据的空盒子。每个程序都有自己的证明对象、举证强度和错误成本。审前把人关起来，意味着在最终定罪之前先限制自由；量刑加重，意味着国家在已经确认责任后增加痛苦；矫正资源配置则可能主要为了帮助。相同预测进入不同位置，必须重新证明它与该程序目标的关系。不能因为模型在A场景验证过，就默认它在B场景也具有同样正当性。

这正是贝卡利亚“谁有权、在什么时候、按照什么规则处罚”的现代版本。算法时代不能只审算法，还要审它被放在哪里。工具的位置本身就是法律的一部分。

专题二　平均准确率很高，不等于每一种错误都可以接受

技术报告喜欢给出准确率、AUC、召回率等总体指标。对一般预测任务，这些指标很重要；但刑事司法必须继续问：错的是哪一种？把低风险的人错判成高风险，与把高风险的人错判成低风险，后果不同。前者可能让无辜者承担更多监控、羁押或更严条件，后者可能增加公共安全风险。制度必须公开自己更害怕哪一种错误，以及为什么。

更麻烦的是，平均准确率会掩盖分布差异。一个模型总体达到80%，不代表在所有年龄、地区、性别或社会群体上都接近80%。如果某些群体样本更少、记录方式不同、历史执法密度不同，错误可能集中在那里。此时“总体更准”不能成为忽略局部权利负担的理由。

贝卡利亚当年反对用个别人的痛苦制造整体震慑，今天我们也不能用少数群体承担系统误差来换一个漂亮的总体指标。公共幸福如果只看平均数，很容易让最重的代价落在统计表格最不显眼的地方。

专题三　“可以解释”不等于“理由充分”

AI治理常说要可解释，但“解释”其实有两层。第一层是技术解释：哪些变量对结果影响大，模型怎样从输入得到输出。第二层是法律理由：为什么这些变量可以进入国家强制决定，为什么这个结果足以支持相应措施。前者回答“机器怎么想”，后者回答“国家凭什么这样做”。

假设系统能够清楚说明：某人风险升高，主要因为过去接触警察次数多、失业时间长、居住地区历史犯罪率高。这已经比黑箱进步，却不代表这些变量应当用于加重限制。与警察接触多可能来自更高的警务覆盖，失业可能反映社会处境，地区变量更可能把邻居的统计特征贴到个人身上。技术解释反而把新的规范问题暴露出来。

所以，法治不能把“可解释AI”当成最终答案。解释只是打开盒子，打开以后还要审内容。贝卡利亚要求法官不能用自己的情绪偷渡刑罚，今天同样要问模型有没有用看似中性的变量偷渡身份和结构差异。

专题四　“人类最终决定”也可能只是橡皮图章

很多制度在面对AI风险时会加一句保险：系统只提供建议，最后由人决定。听上去问题解决了，可现实中“人在环中”不一定等于真正的人类判断。如果工作人员面对一个标着“高风险”的红色提示，又没有时间、信息和权限推翻系统，人只是在屏幕上点确认。形式上是人决定，实质上仍是模型决定。

这叫自动化偏见：人容易把机器输出当成比自己更客观的结论，尤其当系统复杂、机构把它包装成科学工具时。反过来，如果人可以毫无理由忽略模型，系统又失去一致性价值。真正的制度设计必须说明人何时应当采纳、何时可以偏离、偏离需要什么理由、谁复查。

贝卡利亚会非常熟悉这种困境。他当年担心法官借“法律精神”任意加刑；今天我们既要防人借直觉推翻数据，也要防人借机器逃避判断。好的制度不是在人和机器之间挑一个主人，而是让两者的功能边界都显露出来。

专题五　商业秘密能不能挡住被处罚者的质疑

某些风险工具由商业公司开发，模型细节受到知识产权或商业秘密保护。企业当然有保护技术资产的正当利益，但当模型进入刑事司法，另一个利益会变得更重：一个人必须知道影响自己自由的依据，并有机会有效质疑。商业秘密不能自然高于正当程序。

State v. Loomis之所以重要，正是因为专有模型带来了这种冲突。法院最终没有把COMPAS当作决定刑罚严重程度的工具，并要求有独立理由。这个结果至少说明：如果系统不能完全接受对抗性检验，它就更没有资格独占决定。越不可见，功能边界越应收紧。

这里可以形成一条简单原则：透明度不足不是自动禁止一切技术，却会降低它可承担的权力重量。一个只能说“相信我”的系统，最多适合做提示，不适合做最后一锤。贝卡利亚把秘密司法视为危险，算法时代同样不能让“秘密公式”坐上刑罚王座。

专题六　模型上线以后，必须研究它怎样改变自己未来的数据

传统评估常假定模型面对的是一个稳定世界：用过去数据训练，在未来检验准确率。但刑事司法中的模型会参与改变世界。警方依据预测增加巡逻，法院依据评分改变监督强度，矫正机构依据风险分配项目，这些措施都会改变后来被记录的行为。后文数据已经不是“没有模型时的世界”，而是“模型参与过的世界”。

因此，算法审计不能只问最初准确率，还要问部署后的回写：哪些地区因为模型被看得更多？哪些群体因为高风险标签获得更多帮助，哪些反而受到更多限制？模型有没有因为自己的干预改变基准率？如果不做这种动态审计，系统会把自己制造的分布重新当成客观现实。

这正是SDE能够提供实质增量的地方。回写不是一句“技术有影响”，而是要求建立版本化审计：模型A上线前是什么分布，上线后哪些变量改变，改变来自真实风险还是执法密度，何时触发重新训练、暂停或废止。没有回写记录，所谓数据治理只是在重复第一张照片。

专题七　把风险分数用于“帮助”，与用于“加罚”要彻底分开

同一份风险信息，可以导向完全不同的制度动作。发现一个人复归困难、再犯风险较高，制度可以提供更多职业训练、心理支持、住房帮助和社区服务；也可以选择更严监控、更长限制甚至更重处罚。前者把风险理解成“需要更多资源”，后者把风险理解成“值得更多强制”。如果两者不区分，支持性工具会悄悄变成惩罚工具。

这种区分对贝卡利亚特别重要，因为他的预防思想本来就包含教育和改善制度条件，而不只是加大威慑。AI如果能够更早识别谁需要帮助，完全可能成为“改土”的工具；问题在于资源不足时，机构更容易选择便宜的控制而不是昂贵的支持。算法没有决定这一点，预算和制度激励才决定。

所以，评价一套风险系统不能只看模型本身，还要看它的下游动作。相同分数如果自动触发加罚，正当性要求最高；如果只触发可自愿接受的支持，权利风险明显不同。真正的AI法治必须把“预测—动作”这条链完整公开，而不是只展示预测准确率。

专题八　AI时代最需要的不是更聪明的法官，而是更会改错的制度

人们谈AI司法时，常把问题设成“机器和法官谁更聪明”。这其实沿用了发现学的老问题：仿佛那里有一个唯一正确答案，只要找到更好的发现者就行。本书走到这里，问题已经改变。刑事司法不仅要在某一刻作出判断，还要能在错误暴露后修改规则、补偿受害者、更新数据和重新审视制度。真正关键的是纠错能力。

机器可能比人更稳定，却未必更会承认自己错。模型通常按照既定目标优化，只要指标没有触发警报，它不会主动问“这个目标本身还对吗”。人类制度的优势

本应在这里：异议、上诉、学术批评、立法修改、媒体监督和当事人经验，可以让目标本身被重新讨论。可如果机构把模型包装成客观科学，反而可能削弱这种自我怀疑。

所以，AI时代对贝卡利亚最好的继承，不是终于造出一个不会发脾气的法官，而是把他开启的“权力必须说明理由”再推进一步：权力还必须保留被事实和异议改写的通道。一个可纠错但偶尔犯错的制度，可能比一个暂时准确却无法质疑的制度更值得信任。

普通人现场　三个例子看“预测、必要与处罚”

现场一　学校预测“高违纪学生”，会不会把孩子推成那个样子

学校如果根据迟到次数、家庭情况、过去处分和同伴关系给学生做违纪风险分级，看上去可以更早干预。问题取决于“干预”是什么。如果高风险学生得到更多辅导、沟通和学习支持，风险模型可能帮助学校把资源提前送到需要处；如果高风险学生被重点监视、取消活动资格、被老师预先当成麻烦制造者，标签本身就会改变师生关系。

这说明预防和处罚的界线不在系统名称，而在下游后果。贝卡利亚反对把人整体当成“坏人”，今天同样不能把“高风险”变成人格。学校最需要审查的不是模型有没有高精度，而是高风险标签触发了帮助还是排斥，学生能不能知道依据、能不能纠错、多久复核。

现场二　人脸识别抓到了更多逃犯，是不是越多越好

如果人脸识别能够帮助找到已有明确法律依据、已有案件事实和在逃身份的人，它与“从全体人群里预测谁将犯罪”并不一样。前者是在已知对象和已发生事实基础上提高识别效率；后者则可能从人格和行为数据推断未来罪责。两者都需要比例和准确性审查，但权力逻辑不同。

这一区分很重要，因为技术争论经常在“全面支持”和“全面反对”之间摇摆。真正的贝卡利亚式问题应当更细：你现在要完成什么法律任务？对象是不是已经有客观事实基础？系统只是提高识别能力，还是在重新定义谁值得怀疑？只有把任务拆开，必要性才有意义。

现场三　一个算法被证明有偏差以后，停不用就够了吗

假设某风险模型经过审计，被发现对某一群体显著高估风险。机构最直接的反应是停止使用。这当然必要，却还不是完整回写。还要追问：过去有哪些决定受它影响？这些人是否有重新审查机会？训练数据为什么形成这种偏差？采购和验证程序哪里失灵？以后怎样防止同类问题再次出现？

这正是“会改错的制度”和“会删除工具的制度”的区别。前者把错误变成制度学习，后者只是换一个新软件。贝卡利亚把刑罚置于事先法律之下，AI时代还需要把技术置于事后纠错之下。没有补救过去、修改流程和记录教训，所谓停用只是把旧问题封存。

◆ 理论深化　AI把贝卡利亚的五道闸门重新点亮

· 第一道闸门：行为与预测分开。已经发生的行为决定刑事责任入口，未来概率只属于风险信息。风险可以影响支持和管理，但不能自己生出罪责。

· 第二道闸门：规则与模型分开。法律规定什么是犯罪、什么强制可用；模型不能因为发现新的相关性，就自行扩张犯罪边界或处罚权限。

· 第三道闸门：一致与正当分开。机器可以减少同案随机差异，却也可能稳定复制错误。稳定输出不是正义的同义词。

· 第四道闸门：确定性与全知分开。执法可预期值得追求，全面监控不是其逻辑必然。提高发现率仍受必要和比例约束。

· 第五道闸门：概率与必要性分开。风险数据进入决策以后，仍要经过目标、替代、权利成本和反馈审查。事实判断不能跳过规范授权。

◆ 现实压力测试　六个问题攻击当前模型

压力测试一：如果算法准确率远高于人类法官，为什么不让它决定？因为决定包含的不只是预测事实，还包含权利、责任和价值权衡。准确率提高了事实工具的地位，没有自动转移最终授权。

压力测试二：如果一个人被预测为极高风险，不提前限制他是不是对未来受害者不负责？制度当然可以对真实、迫近且有客观事实基础的危险采取预防措施，但措施必须有明确法律依据、程序和比例。不能把一般统计风险直接升级为刑罚。

压力测试三：如果全面监控能把暴力犯罪降低一半，是否就必要？仍然不能只凭效果回答。要比较替代方案、权利成本、误用风险和制度回写。效果越大，只是必要性证据越强，不是权利审查自动消失。

压力测试四：如果算法比人更少歧视，是不是应该更多使用？可以，但要证明这种优势稳定存在，并保留审计、申诉和版本控制。减少一种偏差不能成为隐藏另一种偏差的理由。

压力测试五：如果公众自己愿意用隐私换安全，全面识别就正当吗？多数同意是重要条件，却不是唯一条件。刑事法治还要保护少数人和个体不被不可逆、不可退出的系统吞没。

压力测试六：如果SDE只是说“算法也有土”，有没有实际增量？只有在它能据此改变制度判断时才有价值。本编给出的增量是：把数据来源、反馈循环、风险标签回写和工具功能边界变成审查项，而不是停在“技术不完全中立”的常识。

◆ SDE对账与改判

第一，贝卡利亚的中心判断继续保持“自由限权型公共幸福中心”，但AI时代使“自由硬约束”比前文更重要。公共幸福可以使用预测、统计和模型，却不能把个人仅当作风险容器。

第二，“搭了看台，少了楼梯”在AI章节中获得更清楚的现代含义：模型擅长站在高处看总体，却必须给被评分者提供能够走上去质疑数据、解释情境、纠正错误和触发制度修改的楼梯。

第三，本编新增的动作候选是“数字盖章”：把仍需规范讨论的必要性，用风险分数提前宣布为已经解决。它只在数字替代授权时成立，单纯量化事实不构成该动作。

第四，罪刑法定被重新理解为“模型不能创造处罚资格”；必要性被重新理解为“五步可审查过程”；确定性被限制为“公正可预期”，而不是“全知监控”。这三处都是贝卡利亚进入AI时代以后仍然能工作的遗产。

第五，SDE给自己的死法继续保留：如果它因为强调回写，就否认法律需要稳定边界，会死于无规则；如果它因为强调发生，就把一切权利降成可随政策改变的变量，会死于规范坍塌；如果它把所有算法问题都解释成“土的偏差”，却无法提出具体审查程序，会死于空泛。

◆ 六条正式判决

第一判：罪刑法定之所以比贝卡利亚本人活得更久，是因为它保存的不是一个时代的具体答案，而是一条权力出生程序——先有公开规则，后有处罚资格。AI再会发现模式，也不能越过这条程序自己创造犯罪。

第二判：死刑没有被一本书终结，不证明贝卡利亚失败，而证明制度发生从来不是单线传播。一本书最深的影响往往是改变举证责任：从“国家当然可以”改成“国家必须说明为什么非如此不可”。

第三判：风险预测与刑事责任属于两种不同时间结构。责任回看已经发生的行为，预测面向尚未发生的未来。未来概率可以帮助预防，却不能因为足够高就倒装成现在罪责。

第四判：算法量刑最大的诱惑不是算错，而是让人误以为“算出来”就已经完成了正当化。数字可以提高一致性和事实判断，却不能替代“为什么这些变量应当影响自由”的公开规范理由。

第五判：贝卡利亚的刑罚确定性必须在AI时代重新限缩为“公正、及时、可预期的执行”，而不能扩张为“国家对所有人持续全知”。确定性的手段本身仍须经过必要与比例。

第六判：AI时代新的收回动作候选是“数字盖章”——把本来仍处在争论中的必要性、危险性和政策选择，用一个风险分数提前宣布为定型。推翻它的办法不是拒绝数字，而是让目标、事实、替代、权利成本和回写全部重新显露。

> 带走的话
>
> 必要性不是一个风险分数。算法可以帮助证明“可能发生什么”，却不能单独决定“因此国家可以对你做什么”。

> **小账**
>
> 这一章看见了什么？——78.6%只回答可能发生什么，不回答可以对谁做什么。
>
> 还要追问什么？——目标、权利成本与回写三步，算法都不能独自替代。

第 十 四 编

处罚走出法庭：家庭、学校、公司

编　序

本编五章，让处罚走出法庭。第73章讲父母为什么不能只靠处罚教育孩子；第74章讲公司为什么罚得狠不等于有效；第75章讲学校怎样通过规则制造“坏学生”；第76章讲制度怎样让人只怕处罚、不信规则；第77章讲自由不是“没人管我”。

这一编是判决三的主场。贝卡利亚看见的那些发生，在家里、公司和学校天天重演：重罚让员工把错误藏起来，过密的校规制造违规者，潜规则和秘密执法改变人与人的关系，解释与修复改变孩子以后的选择。

读这一编，可以想想身边最近一次处罚：罚的是那件事，还是那个人？它有没有让下一次更少出问题？没人看着的时候，你还觉得那条规则值得守吗？

第 73 章

父母为什么不能只靠处罚教育孩子

一个孩子把玩具摔坏了。父亲第一反应是：“罚！不罚下次还敢。”这句话和刑罚史上许多最古老的直觉几乎一模一样：错误之后必须跟着痛苦，痛苦越深，记得越牢。家庭不是国家，父母也不是法官，但这里藏着同一种思维习惯——把处罚本身当成教育，把受苦当成改正，把服从当成成长。

贝卡利亚真正值得带回家庭的，不是“别罚孩子”这样一句温柔口号，而是三道追问。第一，你罚他是为了什么？第二，这个目的真的需要这种强制吗？第三，这个处罚会不会制造出一个比原问题更坏的新问题？如果这三问不做，处罚就容易从教育手段滑成成人情绪的排放口。

先看最常见的场景。孩子考试考差，父母没收手机一个月。有没有可能有效？当然可能。如果手机确实占用了大量学习时间，暂时限制使用、重新安排作息，本身可能合理。但如果孩子成绩下降是因为课程听不懂、长期睡眠不足、与同学关系恶化，没收手机只是给结果加痛苦。它既没找到原因，也没创造新的能力。父母看到的是“分数低”，处罚的却是“手机”，中间那条发生路径完全没有被看见。

这正是贝卡利亚“预防优于处罚”在生活中的第一层意义。真正的预防，不是提前把处罚加重，而是去看错误在哪片土里长出来。孩子撒谎，可能是因为他发现说真话会遭遇不可预测的暴怒；孩子拖延，可能是任务太大、反馈太晚；孩子顶嘴，也可能是他开始形成自己的边界，却没有学会表达。把这些情况都塞进“不听话”三个字，再统一处罚，家庭就像一部只有一条罪名的刑法典。

更危险的是处罚会回写关系。一个孩子如果长期体验到“犯错=失去爱、失去说话权、被羞辱”，他学到的未必是规则本身，而可能是怎样不被发现。他会更会隐藏、更会编理由、更会在成人看不见时行动。父母以为处罚降低了错误，实际上可能只是降低了错误被看见的概率。贝卡利亚当年区分“刑罚严厉”与“刑罚确

定”时，已经逼我们注意一件事：行为不会只对痛苦强度反应，它会对整条路径反应。

家庭里的“确定性”也不能误读成“每次犯错都必须罚”。更好的理解是，规则与后果应当稳定、可预期。比如睡前十点以后不再玩游戏，违反以后第二天减少娱乐时间；这个后果事先知道、程度有限，而且和被违反的规则有关系。这样的结构比父母临时发怒说“以后永远别碰手机”更接近法治。孩子首先学到的是世界有边界，而不是大人的心情决定命运。

但家庭教育还比刑法多了一层：父母的目标不是把孩子永远管理成“不会犯错的人”，而是让他逐渐成为能够自己判断的人。如果所有规范都靠外部处罚维持，孩子长大后最大的技能就可能是计算“被抓概率”。没人看就闯红灯、老师不在就作弊、老板看不见就偷懒，这并不是自由，而是把规则理解成外部障碍。

所以家庭中的处罚必须有退出机制。今天父母替孩子设边界，明天应当把一部分判断权交还给他。年龄增长以后，规则需要解释，解释需要允许反驳，反驳以后还可能修改。一个永远不允许申诉的家庭，即使处罚很轻，也可能让孩子学会权力只需要服从；一个允许争辩却没有任何边界的家庭，则可能让孩子误以为自由就是所有约束都可以讨价还价。

在SDE的大白话里，好的教育至少要让四件事连起来：规则如何显露，这是S；孩子过去怎样一步步走到现在，这是差异序列D；家庭气氛、同伴关系、身体状态、学校压力构成特征纠缠E；一次行为与一次回应又会反过来改变以后，这叫“回写”。只盯着行为结果处罚，相当于只看最后长出来的一片叶子，却不看根、天气和之后的生长。

这也是为什么真正有效的家庭规则通常同时包含三个部分：边界、解释和修复。边界告诉孩子哪些行为不能做；解释让他知道为什么；修复让错误之后还有返回的路。打碎别人的东西，不只是挨骂，而是道歉、赔偿、想办法补救；伤害弟弟妹妹，不只是罚站，而是理解伤害、恢复关系、重新学习表达。处罚如果不能通向修复，就容易只留下羞耻。

贝卡利亚最适合带进家庭的一句话，不是“温柔一点”，而是“任何强制都要说明必要性”。父母当然有照护和教育责任，也不可能把所有规则都交给幼小孩子表决。但越是因为孩子弱小、依赖成人，父母越应当谨慎使用羞辱、恐惧和不可逆的惩罚。权力越大，举证责任越重。家庭正是一个人最早学习“权力可以怎样对我”的地方。

> **带走的话**
>
> 父母不能只靠处罚教育孩子，不是因为所有处罚都坏，而是因为教育的目标不是制造一个怕后果的人，而是让一个人逐渐获得能够理解规则、修复错误并承担自由的能力。

小账

这一章看见了什么？——处罚可能只降低错误被看见的概率，孩子学会的是隐藏。

还要追问什么？——处罚怎样通向修复，孩子长大后判断权怎样交还给他。

第 74 章

公司管理为什么“罚得狠”不等于有效

公司里也有自己的“刑罚史”。迟到扣钱，业绩差通报，流程错误罚款，违反规定取消奖金。管理者最容易相信一个朴素模型：成本足够高，员工自然就不敢犯错。这个模型并非完全错误。明确后果确实会影响行为。问题在于，一旦所有管理都被压成“奖罚力度”，组织就开始看不见错误是怎样发生的。

想象一家仓库，员工经常发错货。经理发现以后，把每次错单罚款从五十元提高到五百元。短期内，大家更加紧张，错误可能下降。但也可能出现另一种变化：员工不再主动报告小错，而是想办法遮住；同事发现问题不愿提醒，因为一旦上报就等于让别人损失五百元；管理层看到的数据越来越漂亮，真实问题却被藏起来。处罚越重，信息越差。

这和贝卡利亚反对酷刑的逻辑有一个很有意思的相似点。酷刑把“能不能忍痛”混进“是否有罪”的事实生产；组织中的严厉处罚，也会把“敢不敢说真话”混进质量数据。制度本来想获得更真实的信息，却因为处罚结构改变了人的行为，最后只得到经过过滤的信息。

所以现代公司管理最该从贝卡利亚继承的，不是“处罚轻一点”，而是“强制必须与真实目标对应”。如果目标是降低错单，就要问错误的来源：系统界面是否容易选错？培训是否不足？一个人同时负责太多订单？绩效考核是不是鼓励速度而牺牲准确率？流程中有没有第二次校验？只有在这些问题被排查之后，故意违反规则的责任才更清楚。

这就是必要性原则的日常版。罚款是一种手段，不是目标。手段前面必须有目的，旁边必须有替代方案，后面还要检查副作用。如果改界面、加校验、改善培训能够以更小代价显著减少错误，直接加罚就很难说是“必要”。

公司的规则也需要“罪刑法定”式的底线。员工最害怕的并不是有规则，而是规则事后出现。今天说“这件事没问题”，明天出了事故就追溯说“你早该知道”；绩效差的时候临时发明扣分项；领导不喜欢某个人，就从旧制度里翻出一句模糊条款处罚。这样的组织会把员工训练成政治动物：与其把工作做好，不如先猜领导想什么。

规则越清楚，越能把注意力从“讨好权力”转回“完成任务”。这也是贝卡利亚罪刑法定思想在组织中的普遍意义：一个人只有大致知道行为边界，才能真正安排自己的行动。不可预期的管理不是严格，而是把不确定性变成支配资源。

但公司又不能把规则写死。市场变化、技术变化、岗位变化都比十八世纪刑法快得多。这里恰好暴露了贝卡利亚式机械法官模型的边界：组织如果要求管理者只机械执行条文，也会被新问题拖死。真正成熟的制度，不是没有裁量，而是裁量有理由、有记录、有复查，而且不能偷偷变成针对个别人的临时立法。

举一个更尖锐的例子。一家公司要求客服平均通话时长不得超过六分钟，并规定连续三周超标就降绩效。规则非常清楚，看起来十分“法治”。可员工很快学会五分半就开始催客户结束，复杂问题被草草关闭，客户第二次来电，整体成本反而更高。这里的问题不是规则不明确，而是目标被代理指标替代了。刑罚越稳定，错误目标反而被执行得越好。

这给贝卡利亚的效用思维提出一个现代补丁：不仅要问“处罚有没有效果”，还要问“效果指标是不是我们真正想要的东西”。如果制度奖励的是表面数字，人就会围着数字优化。处罚越确定，优化越彻底。所谓治理失败，常常不是执行不力，而是把不该当目标的东西设成了目标。

因此，公司管理中的“确定性”应当理解成三层：规则稳定、事实认定程序稳定、纠错机制稳定。员工知道什么行为可能产生什么后果；发生争议时知道如何申诉；制度发现规则本身造成坏结果时，也有公开修改方式。只有这样，确定性才服务于信任，而不是把错误冻结。

还有一点尤其重要：处罚不能吞掉尊严。公司拥有劳动管理权，但员工不是组织的附属物。公开羞辱、人格标签、集体连坐，有时确实能制造恐惧，却会破坏合作。一个制度如果要靠让人害怕“被当众收拾”才能运行，说明它把管理问题转化成了身份统治。

贝卡利亚把刑罚从“统治者愤怒”中拉出来，公司也需要把管理从“领导不爽”中拉出来。规则应当针对行为与任务，而不是把一次失误升级成“你这个人态度不行”“你不忠诚”“你没有价值”。行为可以纠正，人格审判则会把员工推到制度对面。

> **带走的话**
>
> 公司管理中真正有效的强制，不是让处罚变得更疼，而是让目标、规则、事实、后果与纠错连成一条透明的路。罚得越狠却越没人敢说真话，是组织版的酷刑悖论。

> **小账**
>
> 这一章看见了什么？——错单罚款从五十提到五百，员工不再报告小错。
>
> 接缝在哪里？——客服六分钟指标执行得越好，真正目标越被代理指标替代。

第 75 章

学校为什么会通过规则制造“坏学生”

学校是一个特别适合观察“规则如何制造身份”的地方。一个孩子刚入学时，身上并没有写着“好学生”“差生”“问题学生”。这些名字会在一次次考试、迟到、表扬、批评、分班、座位安排和老师评价中慢慢形成。标签当然不全是学校制造的，家庭、同伴和孩子自身都参与其中。但一旦学校规则把某种差异固定成身份，身份又会反过来影响以后发生什么。

比如一个孩子因为连续迟到三次，被班主任定义为“纪律差”。此后他第四次迟到，老师会更容易理解为“果然又这样”；而一个一直表现好的学生偶尔迟到，则可能被解释成“今天特殊”。同样行为进入不同身份背景，会得到不同解释。身份开始参与事实判断。

贝卡利亚当年反对按人的地位决定犯罪轻重，也警惕法官依据印象判断人的内心。把这条线带进学校，我们就会发现，纪律制度最大的风险之一不是“有处分”，而是处分逐渐从行为判断变成人格判断。迟到是一个行为，“懒散的人”是一个身份；打架是一个行为，“坏孩子”是一个身份。前者可以结束，后者会跟着人走。

更严重的是标签会回写。一个学生如果长期听见“你就是不行”，很可能逐渐退出主流评价体系。他不再通过成绩和老师认可获得尊严，而转去同伴圈寻找另一种身份。原本偶发的违纪开始成为“我是谁”的一部分。学校随后发现他越来越难管，于是更确信最初标签准确。这是一种典型的自我强化回路。

所以，前文讨论算法风险评分时提出的“模型会改变未来数据”，在学校里早就以人类版本存在。老师心里的模型、班级的名声、处分档案和同伴评价，都会改变孩子以后面对的机会。区别只是它们没有显示成0.83的风险分数。

这并不意味着学校应该拒绝分类和记录。教育必须知道谁需要更多帮助，纪律也必须对严重行为作出明确回应。问题在于分类是为了分配支持，还是为了提前关闭机会。一个孩子阅读困难，被识别出来以后得到额外辅导，这是分类服务成长；如果被直接归入“学不会”并减少挑战性任务，分类就开始制造结果。

处罚同样如此。打架之后，学校当然需要保护受害者、查明事实、制止继续伤害。可是如果只做停课、记过，再把孩子送回原环境，没有处理冲突来源、同伴结构、家庭压力和情绪技能，处罚只是把一段行为切下来贴上标签。它没有改变“土”，也没有给“路”增加新的可能。

学校里的规则还特别容易出现“过度犯罪化”。为了管理方便，越来越多行为被纳入禁止清单：发型、鞋子、座位姿势、课间声音、社交方式、手机、表情、口头禅。规则越多，看上去越整齐，实际越可能把正常成长中的试探和冲突变成纪律案件。贝卡利亚早就提醒，法律过多会制造犯罪；学校同样会因为规则过密而制造“违规者”。

当然，学校不是自由市场。儿童与青少年需要更强的结构，老师也承担安全责任。SDE分析不能把“发生”偷换成“什么都随孩子自己长”。教育恰恰需要人为设计环境。真正的问题是，设计有没有给变化留下空间。一个十二岁的孩子被处分，不应当像刻在石头上一样决定十七岁时别人怎样看他。

这就要求学校的纪律至少有三个“去标签装置”。第一，处分针对行为，不使用永久性人格称呼。第二，记录有期限和修复机制，完成改正后不无限跟随。第三，教师评价需要多来源，避免一个人的印象成为全体老师共享的“事实”。这些并不是什么高深哲学，而是在制度里给回写留出正向出口。

还要注意程序公平对孩子的影响。一个学生即使最后受到处分，如果他觉得自己被听见、事实被认真核查、规则事先清楚、不同学生标准一致，他更可能把结果理解为规则后果；如果他觉得老师先认定“就是你干的”，再去找证据，处分会被理解成权力敌意。程序不只决定结果是否公平，也在教育孩子“社会的规则值不值得相信”。

从这里看，学校其实是法治的第一课堂，不是因为它教宪法，而是因为孩子每天都在经历微型权力：老师能不能临时改规则，班干部有没有特权，处分能不能申诉，优秀学生是否同样受规则约束，犯错后有没有回来的一天。一个社会未来怎样理解权力，很大程度上就在这些小场景里被塑形。

> 带走的话
>
> 学校最危险的处罚不是记一次过，而是把一次行为写成一个人的永久身份。好的纪律让错误结束，坏的纪律让错误成为名字。

> **小账**
>
> 这一章看见了什么？——同样是迟到，进入不同身份背景，就得到不同解释。
>
> 接缝在哪里？——规则过密把成长中的试探变成纪律案件，处分记录跟着人走。

第 76 章

一个制度怎样让人只怕处罚、不相信规则

一个制度表面上可能非常“有效”：查得严、罚得快、违规率低。可如果人们只有在摄像头下才守规则，一离开监督就立刻寻找漏洞，这个制度到底成功了没有？贝卡利亚重视威慑和确定性，但如果我们把他的思想只读成“让违规成本稳定变高”，就会漏掉更深的问题：规则为什么值得人们在没人看见时仍然遵守？

先区分两种秩序。第一种秩序靠外部成本维持。你不乱停车，因为怕罚款；不迟到，因为怕扣工资；不作弊，因为怕处分。第二种秩序除了成本，还有认同：你觉得规则大体合理，别人也在遵守，违反它会伤害共同生活。现实制度通常是两者混合。问题在于，如果第一种长期吞掉第二种，人会越来越擅长计算而不是信任。

处罚主义最典型的副作用，就是把规范变成价格。罚款一百元以后，有些人会理解成“禁止”；另一些人却会理解成“只要我愿意付一百元，就可以做”。当所有规则都被翻译成成本，规范意义就会被市场式计算侵蚀。规则不再告诉你“我们为什么共同这样生活”，只告诉你“违规多少钱”。

贝卡利亚的功利取向容易被误读成这种纯计算模型，但前文已经证明，他并没有把刑法缩成一张价格表。他反复强调法律公开、社会共同利益、限制任意和预防犯罪。这些都说明规则要能被公共理解，而不仅仅是制造恐惧。刑罚的确定性是为了减少任意，不是为了把每个人训练成精密的违规收益计算器。

制度信任从哪里来？第一是可预测。同样行为大致得到同样处理。第二是可解释。人知道规则为什么存在。第三是可参与。规则出现明显问题时有修改通道。第四是可申诉。事实被认错或适用不当时能纠正。第五是权力自己也受规则约束。只要这五点长期缺失，再重的处罚也只能制造顺从，很难制造信任。

一个交通治理例子很清楚。如果某路口限速突然从六十降到三十，却没有明显标识，摄像头仍高频抓拍，驾驶人即使后来交了罚款，也很难把处罚理解成安全规

则。他更容易理解为“设套”。下一次他面对限速标志时，心里出现的不是道路安全，而是“哪里有摄像头”。制度把规范训练成了躲避执法的游戏。

反过来，如果限速与事故风险对应，标志清楚，执法标准一致，处罚收入和治理目标没有冲突，人即使不喜欢罚款，也更容易承认规则有理由。程序和理由不是软装饰，它们决定处罚在人的意识里被编码成“共同边界”还是“外部敌意”。

这也是秘密执法为什么特别伤信任。贝卡利亚反对秘密控告，不只是因为可能冤枉人，还因为它改变人与人的关系。一个制度如果大量依靠告密、隐藏标准和不可见审查，会让每个人都开始猜测谁在观察自己。规则不再是公开边界，而变成一张看不见的网。短期内人可能更谨慎，长期则会降低合作和诚实。

现代组织中还有一种更隐蔽的形式：规则很多，但真正决定结果的是“潜规则”。明文写着按绩效晋升，实际靠关系；制度写着可以申诉，申诉者却会被认为“不服管理”；政策说人人平等，重要人物总能例外。这样的制度会教会人一个最危险的道理：正式规则只是给普通人看的，真正聪明的人应当寻找例外。

这时，再增加处罚只会加深犬儒。因为大家不是不知道规则，而是不相信规则背后的权力也会遵守规则。法治最深的敌人并不总是无规则，而是“双层规则”：一层写给被管理者，一层留给掌权者自己。

贝卡利亚的现代意义因此可以重新说：他要求国家把处罚权写进一般法律，真正做的不是增加条文，而是取消权力的私人入口。规则只有在权力也被它约束时，才值得普通人相信。否则，处罚的确定性只是“你一定会被管”，不是“大家都受同一规则约束”。

从SDE看，信任是一种回写结果。规则实施以后，人们不断观察：是否公平、是否一致、是否允许纠错。这些经验沉回环境，变成下一次面对规则时的预期。一个制度今天不公，明天不仅多一个受害者，还多一层不信任的土。反过来，一次公开纠错也不仅解决一个案件，它告诉所有人“这套制度会改错”，信任因此可能重新生长。

所以好的制度不只要会处罚，还要会道歉、撤销、补偿和修订。只会证明自己永远正确的权力，会把每一次错误都变成新的敌意。贝卡利亚没有用“制度信任”这个现代词，但他把公开法律、有限刑罚和程序限制绑在一起，已经在为这种信任铺地。

> **带走的话**
>
> 一个制度若只能靠“被抓就倒霉”让人守规矩，它得到的是服从，不是法治。真正稳定的秩序不是处罚无处不在，而是规则在无人监督时仍被认为值得遵守。

> **小账**
>
> 这一章看见了什么？——罚款可能被读成价格：愿意付一百元，就可以做。
>
> 接缝在哪里？——双层规则：一层写给被管理者，一层留给掌权者自己。

第 77 章

自由不是“没人管我”

我们一路把“自由”作为贝卡利亚理论的重要硬约束，很容易让普通读者产生一个误会：自由是不是就是“别管我”？国家少管一点、父母少管一点、公司少管一点、学校少管一点，自由就多一点。这个理解有一半是真的，也有一半危险。

真的一半是，权力确实需要边界。一个人如果随时可能被搜查、惩罚、羞辱、监控，他很难拥有稳定生活。贝卡利亚之所以重要，就是因为他把国家处罚权从“天然拥有”改成“必须证明”。自由首先意味着，有些地方权力不能随便进去。

危险的一半是，如果把自由只理解成“没有约束”，共同生活会迅速解体。交通规则限制了我随便开车，却使所有人更可能安全到达；合同限制我随时反悔，却使合作成为可能；考试规则限制我抄袭，却使成绩还有公共意义。很多约束不是自由的敌人，而是自由能够互相兼容的条件。

所以，自由不是零规则，而是规则不能把人变成纯粹被管理对象。一个自由人应当大致知道规则，有机会表达、申诉和改变规则，并在不伤害他人的范围内拥有自己的生活空间。这里的关键词不是“无人管理”，而是“不被任意管理”。

贝卡利亚的刑法自由恰好是这种自由：你不是免于一切处罚，而是免于没有法律依据、没有必要、过度、秘密和任意的处罚。国家仍然可以处罚，但它必须走一条预先公开的路。这种自由并不浪漫，却非常坚硬。

这也能帮助我们理解为什么“自由”和“公共幸福”在贝卡利亚那里不必二选一。自由如果只是个人退出一切共同义务，公共生活不可能维持；公共幸福如果可以随时牺牲个体，个人又只是总体计算里的材料。成熟的结构是：公共幸福提供方向，自由提供边界。目标可以是减少犯罪、增加安全，但有些手段即使有效，也因为过度侵犯人的地位而不应使用。

这里还需要把前文形成的“规范主体与经验主体”区分再讲清楚。规范上，我们把每个人当作有自由资格的人，不因为他贫穷、受教育少、过去犯过错，就降低基本尊重。经验上，每个人实际能否作出选择，却受到家庭、健康、知识、金钱和环境影响。前者不能因为后者而消失，后者又不能因为前者而被假装不存在。

例如，一个人在极度贫困中犯罪，理解他的处境不等于自动取消责任；但如果制度完全不看贫困、教育和社区环境，只说“人人都有自由，所以后果自负”，自由就被抽空成一句形式口号。反过来，如果因为处境困难就认为他没有任何主体性、所有行为都只是环境结果，也会把人降成被因果推动的物体。

这正是“自由是一种能力，也是一种资格”的双层结构。资格要求制度尊重你、给你程序、把你当作能回应理由的人；能力则需要现实条件支撑。教育、信息、基本生活保障和免于恐惧，都会影响一个人实际使用自由的程度。贝卡利亚更强地守住了资格边界，后来的社会理论则继续追问能力从哪里来。

放回普通生活，自由也不是家庭里“孩子想干什么就干什么”，不是公司里“员工不受任何管理”，不是学校里“学生可以拒绝所有规则”。这些都只是把权力撤掉，却没有回答多人如何一起生活。真正的自由需要边界互相咬合：我的选择不能把别人变成纯粹承担者，别人的管理也不能把我变成没有声音的对象。

所以“没人管我”是一种负面的想象，而“我能在可理解、可质疑、可修改的边界内形成自己的生活”才更接近本书到这里要保住的自由。这种自由不排斥规则，反而需要更好的规则；不排斥共同目标，反而要求共同目标不能跳过个人。

到这里，“普通人编”其实把整本书重新压回一句很朴素的话：无论国家、父母、老师还是老板，只要你有权让别人承受不利后果，你就必须解释为什么。权力不因为出发点善良就自动正当，规则不因为写在纸上就自动合理，处罚不因为有效就自动适度。

带走的话

自由不是没有人管，而是没有任何人可以只凭身份就拥有无限管理你的资格。

真正的自由，生活在边界、理由、程序与可修订之间。

小账

这一章看见了什么？——自由不是零规则，而是不被任意管理。

还要追问什么？——自由资格之外，自由能力要靠教育、信息与基本生活保障。

第 十 五 编

给贝卡利亚写死法，也给SDE写死法

编　序

本编五章，给贝卡利亚和SDE各写死法。第78章写主死法：把“最少必要”焊成永久答案；第79章写三处次死法；第80章列出能推翻本书的证据清单；第81章问SDE是不是也先放了一个“发生”；第82章给SDE列出五种死法。

这一编是判决五反过来对准自己的地方。制度必须允许纠错，那么继承贝卡利亚的制度要保留复查与退出，本书要把F1至F11的触发状态摆出来，SDE也要写明自己在哪些证据面前认输。判决四的锁钥同手，也在F2、F11的记录里接受检验。

读这一编，可以问问身边那条“当初必要”的规定：它现在还必要吗，有没有定期复查、到期退出？再问自己：你最确信的一条看法，什么证据出现时你会真的改口？

第 78 章

贝卡利亚体系的主死法候选：把“最少必要”变成永久答案

现在轮到我们给贝卡利亚写“死法”了。所谓主死法，不是找他历史上哪一句话已经过时，而是问：如果沿着他的体系一直走，哪一种内部倾向最可能让这套思想失去继续生长的能力？经过前面七十多章反复改判，本书不再把“自由主体先在”作为主死法。那个判断太粗，而且已经被我们自己收回了一部分。现在更准确的主死法是：把“最少必要”从一种持续审查程序，冻结成一个已经完成的答案。

◆ 把“主死法”从人物判决改成可检验的制度风险

经过第56章的反证以后，本章标题中的“主死法”必须降一档：它不是说贝卡利亚本人把“最少必要”永久焊死，而是说任何继承贝卡利亚语言的制度都有这种死亡风险。一个本来要求持续比较替代方案的“必要”，一旦变成固定口令——“公共安全需要”“系统风险要求”“算法显示高危”——就会从限制权力的钥匙变成替权力开门的万能钥匙。

三方程能把风险定位得更精确：S是当前被承认为必要的制度显露，D是比较替代方案、积累经验、复查成效的差异序列，E是风险分布、资源条件、权利关系与社会信任的特征纠缠。所谓“焊死”，不是S稳定本身有错，而是S稳定以后阻断D继续变化、切断E的新信息回写。介生态被过早宣布结束，才是真正的“封顶”。

123原理因此给出三个可观察征兆：第一，张力出现时不再列替代方案；第二，新结果只被解释成旧政策正确，而不允许反例改变政策路径；第三，退出机制、复查期限和错误补救被取消。只要制度仍保留这些装置，稳定就不等于死亡。

同一失效图谱也可以用六路径复核。对已经固化的“必要”结论，从S→D→E逆拆最有效；对正在设计的新政策，则可从E→D→S先看风险与资源条件，再比较手段，最后允许制度显露暂时成形。若无论何种任务都只走一条固定顺序，SDE自己也会把方法冻成新的“必要性”。

再做一次去术语测试：行政法的比例原则、政策评估和日落条款已经能够说明很多“必要性不能永久化”的道理。SDE不能把这些常识重新命名后算原创。它若还有用，只在于把“稳定显露—路径冻结—纠缠失声—回写中断”合成一张可以跨刑罚、算法和组织管理复用的失效图谱，并允许图谱被具体制度反证。

贝卡利亚最锋利的词之一就是“必要”。不是为了传统，不是为了报复，不是因为君主喜欢，而是因为共同安全确实需要。这个词把刑罚从权力欲望拉回理由。可所有强大的概念都有自己的危险：它越成功，越容易被后人当成免检标签。

一旦“必要”从问题变成结论，贝卡利亚的思想就会反转。国家说“这是必要的”，法官说“这是必要的”，算法说“风险高所以必要”，公众说“为了安全当然必要”。原本用来限制权力的词，开始替权力盖章。这个变化不需要篡改贝卡利亚原文，只需要把审查过程拿掉。

所以“最少必要”必须永远保持比较结构。至少要问：目标是什么？有哪些替代手段？为什么较轻手段不足？强制的副作用是什么？新证据出现后是否需要调整？只要这五问还在，必要性就是活的；如果五问消失，只剩“必要”两个字，体系就开始死亡。

主死法因此不是“功利主义”，也不是“社会契约”。真正致命的是封顶：一个时代暂时找到的最小强制，被后来人误写成永恒最小值。十八世纪认为某种监禁必要，今天未必；今天认为某种数字监管必要，明天也可能被更轻技术替代。必要从来依赖目标、条件和可用方案，它本身有历史。

这个判断还能解释贝卡利亚为何具有持续生命力。他最好的部分从未给出一份永久刑罚菜单，而是改变举证责任。死刑、酷刑、秘密审判之所以被他攻击，不只

是因为“我讨厌”，而是因为国家必须证明它们为何不可替代。只要举证责任仍在，理论就有更新空间。

主死法出现的标志有三个。第一，政策只引用“公共安全”而不展示替代方案。第二，制度只讨论效果，不讨论权利成本。第三，规则一旦建立，缺乏定期复查和退出机制。这三点一旦同时出现，“最少必要”就从活原则变成固定口号。

这里需要注意一个反驳：任何制度都不可能每天重新证明全部规则，否则社会无法运转。对。必要性不是要求永久不稳定，而是要求稳定中保留可触发的重审条件。比如新技术出现、更轻方案成熟、错误率显著上升、权利成本超出预期、社会目标变化，这些都可以成为重审触发点。稳定与可修订不是敌人。

用SDE的话说，贝卡利亚的体系不能只允许从混沌走向定型，还必须允许定型被回写重新打开。如果秩序态被误当终局，回写就死了；回写一死，“必要”便从一个动态关系变成一块磐石。这才是最符合全书总原则的主死法。

因此我们给它一个家常名字：

“把临时钥匙焊死在锁里。”

原本，必要性是一把钥匙，用来检查这道门现在能不能锁、锁到什么程度。可如果钥匙被焊死，后人只会看到“门本来就该锁”。发生重新变成发现，贝卡利亚也进入他自己帮助推动的那条哲学史悖论。

◆ 把失效图谱真正跑一遍：刑事风险评分

拿书中已经讨论过的刑事风险评分作压力测试。第一步是“稳定显露”：一个由历史数据生成的风险分数被写进量刑或处遇材料后，它很容易从辅助信息变成“这个人就是高风险”的稳定判断S。日落条款至多会问这项制度几年后是否自动失效，却不一定追问这个分数在案件里怎样取得了事实地位。

第二步是“路径冻结”：一旦高风险阈值固定，决策者可能不再持续比较社区支持、治疗、就业干预等较轻替代方案，D从“比较手段”缩成“达到阈值即进入

更强控制”。第三步是“纠缠失声”：历史执法差异、地区资源、数据缺失与群体暴露程度被压进一个整洁分数，E的来源不再可见。

第四步是“回写中断”：模型改变警务和司法决策，新的高强度执法又生产更多记录，记录再回到模型。若制度只定期检查模型是否仍然“预测准确”，却不检查它是否改变了数据生成环境，就可能把自己制造的相关性当作外部现实。这里正是失效图谱比单纯日落条款多看的一层：它不只问“何时终止”，还问“什么被冻住、什么纠缠被消音、结果是否正在重造下一轮输入”。

这个案例并不能证明SDE优于所有行政法或算法治理理论；比例原则、影响评估和偏差审计完全可以覆盖其中许多问题。它能保留的有限增量，是强迫四个位置同时接受审查，并把“政策改变自身证据环境”设为一项不能被效果数字替代的问题。

◆ 与反馈回路研究划界（F10）

失效图谱的第四步——模型改变执法，执法产生记录，记录回到模型——并不是本书先看见的。哈考特2007年的《反对预测》已经论证，按风险画像分配执法会产生“棘轮效应”，让被画像的群体在记录里越来越多；Lum 与 Isaac 2016年用奥克兰的毒品犯罪数据模拟了预测警务怎样把警力反复派回同一批街区；Ensign 等人2018年更从数学上证明，只用自己发现的事件来更新的预测警务会陷入“失控的反馈回路”。“回写中断”这一步单独拿出来，已被充分研究。

失效图谱若还有自己的东西，只能是四步之间的顺序：回路之所以失控，是因为在它之前，替代方案的比较已经停止——路径先冻结，回写才中断。一个仍在认真比较社区支持、治疗和就业干预等较轻手段的机构，即使用了风险模型，它的记录也会不断被替代手段的结果稀释；只有当“达到阈值即进入更强控制”变成唯一路径，模型收到的记录才只剩一种来源。这个顺序主张不限于算法，学校的零容忍、公司的末位淘汰，都可以按同一顺序去查。

它的失败条件因此是：如果在已记录的失控反馈案例里，替代方案的比较一直在进行、路径并未冻结，回路照样失控，那么“路径冻结先于回写中断”不成立，失效图谱应降为对既有反馈回路研究的重述。本书把它记为F10，状态为未检验。

> **带走的话**
>
> 贝卡利亚最可能死在自己最强的原则上——当“最少必要”不再接受新的事实、替代方案和权利经验回写时，限制权力的原则会转化为替既有权力提供合法性的固定印章。

> **小账**
>
> 这一章看见了什么？——最少必要一旦不再比较替代方案，就成了替权力盖章的口令。
>
> 接缝在哪里？——路径冻结先于回写中断的顺序主张，已被Ensign等人先占。

第 79 章

次死法：三个没有彻底走完的地方

主死法是一条，次死法却可以有几条。它们不是同等严重，也不一定真的已经把贝卡利亚体系杀死，更像三处长期暴露的薄弱地带。只要后人继续补，它们就不致命；如果被忽略，则会和主死法合流。

第一处次死法：主体接口不足。

前文已经反复修正“主体先在”的说法。贝卡利亚并不需要先证明一个经验上完全脱离社会的自然人真实存在，才可以主张每个人在规范上应被尊重为自由者。规范资格和经验生成必须区分。但区分以后，问题仍然存在：一个人的现实选择能力怎样进入责任与政策？

如果制度只守规范资格，不看能力差异，会出现形式平等：人人都被当作自由选择者，于是贫困、知识缺失、创伤、成瘾、教育差距都被视为与责任无关。反过来，如果过度强调能力差异，又可能走向家长主义：国家说你受环境影响太深，所以我替你决定什么对你好。两边都危险。

贝卡利亚给了现代刑法一个坚固的自由边界，却没有完整发展“自由能力”理论。这不是他的错误，更像历史任务尚未展开。后人需要建立接口：哪些处境影响责任，哪些影响量刑，哪些只应该影响社会政策；在承认处境作用时，怎样仍保留人的回应能力。

第二处次死法：公共幸福的可见性不足。

贝卡利亚多次诉诸公共幸福、社会利益、共同安全。前文已经把他的中心诊断收紧为“自由限权型公共幸福中心”。问题是，公共幸福很容易被总体指标代表：犯罪率下降、社会成本降低、秩序提升。但平均意义上的改善可能让少数群体承受集中代价。

比如一项高压治理让总体犯罪下降，却对某些社区造成持续误伤；一套风险模型提高总体准确率，却系统性高估特定人群。若只看总量，公共幸福可能把少数人的自由吃掉。因此公共幸福必须被拆开看分布：谁受益，谁承担代价，代价是否可逆，受影响者是否有发声位置。

贝卡利亚的自由硬约束能部分抵抗这个问题，但还需要更明确的“分布审查”。否则，总体利益仍可能通过“必要”重新压回个人。

第三处次死法：回写机制不足。

这是全书最明显的SDE补充。贝卡利亚极其重视制度怎样改变人的行为：确定性影响威慑，秘密控告改变关系，法律过多会制造犯罪，教育可以预防犯罪。他已经看见很多回写。但这些洞见散落在各处，没有被统一成一个制度学习机制。

现代制度必须比十八世纪更进一步：政策实施后要收集结果、识别副作用、允许申诉、公开失败、修改规则、补偿错误。没有这一层，制度容易把初始设计当作永恒正确。尤其AI时代，模型和制度会不断制造新的数据，如果不追踪回写，系统会把自己的影响误认为外部世界的自然事实。

这三处次死法可以压成三句话：

一，给了人规范地位，却没有充分说明现实能力怎样形成。二，给了公共幸福方向，却没有充分展开利益与代价怎样分布。三，看见制度会改变行为，却没有把“改错”建成持续程序。

它们彼此相关。主体接口不足，会让政策看不见谁最难使用自由；分布不足，会让平均幸福掩盖集中伤害；回写不足，则让这些问题发生以后也难以重新进入制度。三者一旦同时存在，贝卡利亚最初限制权力的目标反而可能被一套自信、理性、有效率但不够会改错的治理体系侵蚀。

但次死法也告诉我们怎样复活。主体接口可以通过能力与处境分析补；分布问题可以通过权利审查、平等审查和群体影响评估补；回写不足可以通过复查期限、公开数据、救济和制度学习补。补法并不是把贝卡利亚扔掉，而是沿着他已经打开的门继续走。

带走的话

贝卡利亚的三处次死法，不是“他想错了”，而是“他走到那里以后，后来人不能停”。凡把他的十八世纪答案直接当现代终点，才真正使这套思想死亡。

小账

这一章看见了什么？——主体接口、公共幸福分布、制度回写，三处都没走完。

还要追问什么？——平均幸福提高时，集中伤害落在谁身上，他们能否发声。

第 80 章

什么证据会推翻本书

一本解构别人的书，最容易在最后一章获得一种危险的自信：我们已经看清贝卡利亚哪里开门、哪里关门，仿佛站到了他看不见的位置。按照本书的方法要求，这里必须反过来问：什么证据会让我们承认，本书的核心判断错了？如果没有答案，这本书就正在把自己的发生偷偷打造为发现。

◆ 统一推翻清单：把承诺变成状态记录

本书此前分散列过多份推翻条件。现在只保留F1—F11一张状态表；方法性失败另编号为M1—M4，不再与推翻条件混写。触发不是失败，而是方法真正工作。经检验，F2完成晚年文本复核并因“门槛位置”部分触发，F3由直接规范材料补强，F6以米兰案与卡拉斯案完成对照；F8、F9分别检验诊断工具与问题次序，本版经不知结论的编码者复核后各有改判；新增的F10、F11分别检验失效图谱的顺序主张与“锁钥同手”的历史预测：前者被先占，后者部分触发。

编号	推翻条件	状态	处理结果
F1	贝卡利亚若系统说明主体如何在教育、激情与制度中形成，“主体先在”应撤回。	已触发	教育、激情与制度章节迫使本书改为“规范资格／经验能力接口”。
F2	若自由始终只是公共幸福的工具，自由硬闸判断应撤回。	部分触发（门槛位置）	1792年Voto确认普通犯罪处硬闸未松动；但例外开在“足以颠覆国家”之人与罪，且由受威胁的国家自行判定。中心诊断保留，补入“门槛位置”（第56章）。
F3	若贝卡利亚后来有持续行政与制度修正实践，“少了楼梯”不得作为人物总判。	已触发并补强	1771年后行政实践加上1792年改革委员会死刑意见，足以否定“贝卡利亚拒绝制度反馈”的人物总判；仅保留1764文本与现代制度化反馈诊断。

编号	推翻条件	状态	处理结果
F4	若“最少必要”在贝卡利亚自身文本中已被处理为持续复查程序，“主死法”应撤回。	部分触发	文本有经验、概率、预防与改革意识，但现代化的复查、退出架构并未成型；改写为制度继承风险。
F5	若最强竞争解释（尤其Hart混合论）已完整覆盖本书“两阶段”结构，原创主张应撤回。	已触发	两阶段先占被承认；“第零闸”降级为“授权先于目的”的问题次序判断，其检验见F9。
F6	若去掉SDE术语后解释力不变，SDE部分应降级。	已检验（有限正面，待分离）	米兰案（有铭刻物、无有力申诉者，约一百五十年）与卡拉斯案（无铭刻物、有伏尔泰申诉，约三年）方向一致，但两项条件共变，无法分离；“环境持续性”保留为候选变量，失败条件见第23章。
F7	若更强反馈机制反而稳定放大恐慌、偏见或多数暴政，“反馈越多越好”应撤回。	未检验	后续需以具体参与机制、危机决策与算法反馈案例检验，不能预设反馈必然改善。
F8	若“否决—门槛—位置—方向”锁钥问题不能在答案已知的思想家身上读出正确中心，诊断工具应降级为修辞。	已检验（盲编码部分一致）	作者校准三例后，交不知结论的编码者对六位思想家盲编码：边沁、罗尔斯、贝卡利亚一致；康德不一致，作者漏读《法权论》6:334的紧急例外，据此补入第四问“门朝哪边开”。人类盲测仍待做（第44章）。
F9	若贝卡利亚在同时含权利与效用论证的关键章节中多数并不先问授权，“授权先于目的”的笼统判断应放弃。	部分触发	作者六章编码与第二编码者在重叠样本上六比六一致；扩至十八章后，授权在先10、效用在先3、融合5，“身体与生命”版收窄被推翻，改为“权力归谁与定罪前人身两类问题授权先于目的”（第64章）。
F10	若在已记录的失控反馈案例中，替代方案比较持续进行、路径并未冻结，回路照样失控，“路径冻结先于回写中断”应撤回，失效图谱降为既有反馈回路研究的重述。	已触发（先占）	Ensign 等2018给出的失控条件——模型只用自己发现的事件更新——实质就是“路径冻结”；顺序主张被先占，失效图谱降为教学框架（第78章）。

编号	推翻条件	状态	处理结果
F11	若在行政主导的死刑恢复中，普通犯罪先于国家安全罪回来，或看不出先后，“锁钥同手”的预测力应撤回，降为施米特与“任何人不得做自己案件的法官”的重述。	部分触发（顺序预测）；F11′待检	不知假设的编码者整理1786—1970年36件废止—恢复事件：行政主导（把握中等以上15件）9件先走国家安全之门、1件相反（奥地利1933）、5件同时，顺序预测部分失败；事后发现的“行政主导首批几乎必含国家安全类（14/15），民意主导几乎从不含（1/15）”记为新假设F11′，以1970年后样本检验（第56章、附录）。

先明确当前版本的核心判断。经过前文展开，它已经不是最初那句“贝卡利亚把国家发生化，却把主体先在化”。我们已经收回这个过强判断。现在更谨慎的版本是：贝卡利亚以公共幸福为总体方向，以自由作为国家强制的硬约束；他成功把刑罚权从天然权力改写为必须证明来源、必要与比例的有限权力，但他没有把规范自由资格、现实自由能力、利益分布与制度回写完整接成一套持续学习的模型。

F6已完成首次检验并扩样：1630年米兰案与1762—1765年卡拉斯案对照显示，判决被铭刻于环境者纠错更慢，但铭刻物与外部申诉力量同时变化，两案无法把二者分开；“环境持续性”只能保留为待分离的候选变量。F2用1792年死刑意见完成晚年复核：例外门槛仍然极高，但例外开在“足以颠覆国家”之人与罪，而判定者正是国家自己。F8、F9的结果分见第44章与第64章：诊断工具通过三例校准，又经盲编码部分改正并补入“方向”；问题次序判断在第二编码者扩样后再次改写。F10、F11分见第78章与第56章：失效图谱撤回增量；“锁钥同手”的顺序预测在36件全样本编码中部分失败，留下待检的F11′。

◆ 方法性失败 M1—M4

M1　选择性阅读。如果我们只挑能证明“看见—收回”的句子，忽略贝卡利亚自己反对我们的文本，这本书即使每句都有出处，也只是检察官卷宗。真正的前期研究必须先把反对自己的证据列出来。

M2　历史倒装。如果我们用今天的权利理论、算法治理和司法解释学直接要求十八世纪文本回答所有现代问题，再把回答不出来当成缺陷，就是把后来长成的树倒插回种子。现代问题可以拿来测试，但必须标明“这是今天继续追问”，不能伪装成贝卡利亚当时已经失败。

M3　概念膨胀。如果“发生”“回写”“土”可以解释任何结果，无论发生什么都说“这也是发生”，SDE就失去可反驳性。真正有用的概念必须告诉我们，什么情况下它不适用，什么证据会迫使改判。

M4　单一中心强迫症。如果我们为了完成“中心诊断”，硬把贝卡利亚所有思想压成一个中心，忽略他内部可能确实存在无法完全统一的自由、效用、人道和秩序张力，母题就会变成削足适履。一个思想家可以有主轴，但主轴不是把剩余都消灭。

因此，本书的终审不能只问“有没有写得更深”，还要问“有没有留出被推翻的位置”。真正可靠的判断，不是因为它看上去坚固，而是因为它提前写明哪一块证据一来，自己就必须让路。

到这里，我们终于可以给前文写作做一次总改判。最初的贝卡利亚像一个“自由中心主义者”；后来我们发现公共幸福并不是边缘词，于是改成“自由限权型公共幸福中心”；再往后，主体先在的批评被收窄，真正的不足转向主体接口、分布和回写。每一次改判都说明书没有沿最初预设直线前进。原典在反过来改变我们的判断，这正是本书应该有的样子。

◆ 理论深化　从“处罚”回到“规则如何长进人心”

第73—77章表面上离刑法很远，实际上做了一次必要的反向验证。如果贝卡利亚的原则只能在国家刑罚里成立，那么它可能只是一套专业技术；如果放进家庭、公司、学校仍能帮助我们区分任意与规则、强制与必要、处罚与修复、服从与信任，它才具有更普遍的思想价值。

这五章也纠正了一个潜在误解：本书不是主张“少处罚就是进步”。有些家庭需要更明确边界，有些公司需要更稳定责任，有些学校面对欺凌必须及时干预。真正问题从来不是强制数量，而是强制能不能解释、能不能比较、能不能修复、能不能被纠错。无边界的宽松和无边界的严厉一样，都可能让人失去可预期世界。

从这一轮生活压力测试里，我们可以提炼出“贝卡利亚五问”的日常版本：第一，究竟发生了什么，而不是我感觉对方是什么人；第二，我想保护什么目标；第三，为什么需要这种不利后果；第四，有没有更轻且更有效的办法；第五，这次处理会怎样改变以后的人与关系。五问走完，处罚才从情绪变成治理。

本编还让“自由”获得了更完整的定义。自由不是没有约束，而是一个人不被任意约束，并拥有理解、回应和参与改变规则的可能。这个定义既保留贝卡利亚的限权传统，也允许现代社会讨论教育、能力和现实条件。它把“别管我”升级成“别把我当没有声音的对象”。

主死法与次死法的区分也在这里最终固定。主死法只有一条：必要性封顶——把持续比较变成永久答案。次死法有三条：主体接口不足、公共幸福分布不足、制度回写不足。前者会直接让限制权力的原则变成新权力的印章，后三者则让体系在复杂现代社会中失去精度。

◆ 八条正式判断

第一判：处罚的价值不由痛苦强度决定，而由它是否真正改变问题的发生条件决定。家庭、学校和公司若只增加痛苦而不改变“土”和“路”，处罚越重，隐藏和对抗可能越多。

第二判：规则的稳定性是自由条件，但稳定不是僵死。好的规则既让人事先知道边界，也让新的事实能够通过公开程序重新进入。

第三判：学校和组织最危险的处罚，是把一次行为转成永久身份。行为可以结束，身份标签却会回写未来，使制度亲手制造后来继续处罚的对象。

第四判：一个制度若只制造“被抓就怕”，获得的是服从，不是信任。信任来自规则可预测、可解释、可申诉、可修改，而且权力自身同样受规则约束。

第五判：自由不是“没人管我”，而是“不被任意管理”。共同规则可以限制选择，同时成为人人能够互相兼容地行动的条件。

第六判：贝卡利亚的主死法不是他不够现代，而是“必要性封顶”。任何一代人一旦把暂时找到的最小强制宣布成永久答案，发生就重新被写成发现。

第七判：贝卡利亚的三处次死法分别在主体、分布和回写：人的现实能力怎样进入责任，公共幸福的代价由谁承担，制度犯错以后怎样真正学习。后人若补不上这三处，理性治理会越来越聪明，却未必越来越自由。

第八判：本书自己的可信度取决于它能否被证据推翻。凡是只允许证据支持自己、不允许证据迫使改判的方法，最终都会成为它最初反对的那种“发现学”。

进一步深描一　处罚为什么容易变成情绪出口

处罚最容易失控的时刻，不是制度认真比较了多种方案之后仍然选择强制，而是权力者感到受冒犯的时候。父母觉得孩子“挑战权威”，老师觉得学生“不给面子”，老板觉得员工“故意顶撞”，国家觉得行为“蔑视秩序”。这时处罚的真实目标很容易悄悄改变：从解决问题，变成恢复权威者被刺痛的尊严。

这正是贝卡利亚把犯罪尺度从“谁被冒犯”转向“社会受到什么损害”的现代意义。一个人对权威不礼貌，可能确实需要纠正，但不能因为权威者更愤怒，就把行为本身无限放大。日常管理中最值得警惕的句子往往是：“你这不是犯错，你这是态度问题。”一旦进入“态度”，规则的边界就开始模糊，处罚也更容易追逐人格。

解决办法不是要求掌权者没有情绪。父母会生气，老师会受伤，领导也会觉得被冒犯。真正的制度性成熟，是在情绪最强时仍然有一个暂停装置：先区分事实、损害、规则和情绪，再决定后果。尤其不可逆或严重的处罚，应当尽量避免在愤怒高峰即时作出。

这和贝卡利亚强调刑罚应由法律而不是临时意志决定是一脉相承的。一般规则的价值之一，就是替人在情绪最剧烈的时刻保留一点距离。规则不是为了消灭人，而是防止人的瞬时状态直接变成别人的命运。

进一步深描二　“修复”为什么比“原谅”更难

当我们说处罚应该通向修复，很容易被误解成“犯了错就算了”。事实上，修复往往比简单惩罚更难。惩罚只需要让犯错者承受不利，修复却要求把被破坏的关系、财物、信任和能力逐项处理。

孩子弄坏同学的物品，罚站十分钟很容易；让他理解损失、道歉、用自己的方式补偿，并重新学习如何处理冲突，更耗时间。员工造成项目失误，扣奖金很简单；让团队分析流程缺陷、修复客户关系、重新设计校验步骤，更复杂。学校学生欺凌他人，记过可以一纸完成；让受害者获得安全，让行为人真正停止伤害，让旁观者不再助长，需要整个环境参与。

因此，处罚主义有一个常被忽略的优势：它管理成本低。权力者只要决定轻重即可，不必真正理解问题。修复则要求知道发生了什么、损害是什么、怎样补救、如何避免再发生。也正因为如此，一个制度越只靠处罚，越可能说明它缺乏处理复杂原因的能力。

贝卡利亚把刑罚目的从报复拉向预防，已经走出了第一步。现代制度还需要再走一步：预防不仅是让人害怕再犯，也可以是恢复能力、修复关系和改变环境。这里“少罚”不是核心，“少制造下一次问题”才是核心。

进一步深描三　公平不是“所有人一模一样”

在家庭、公司和学校谈规则一致，很容易滑向另一极端：既然公平，就对所有人完全一样。迟到一次都罚十元、考试晚交都扣十分、请假程序人人一致。形式上非常整齐，却可能忽略真实差异。

贝卡利亚反对身份特权，并不意味着所有情形都不许区分。真正关键的是区分有没有公共理由。因为某人是贵族所以轻罚，是身份特权；因为某人行为造成的风险不同、责任能力不同、事实条件不同而作区分，则可能是规范所要求的精细化。

现代制度尤其要区分“平等适用”与“机械相同”。一个身体有障碍的员工需要不同工作安排，不是特权；一个有阅读障碍的学生获得额外考试时间，也不是破坏公平。相反，如果为了表面相同拒绝合理调整，规则本身会制造不平等。

这对本书的自由观也是一次校正。自由不是每个人面对完全一样的外部条件，而是每个人都被当作具有同等规范地位的人，并在必要时获得能真正使用这种地位的条件。规范平等与能力差异并不冲突，困难在于怎样防止“照顾差异”重新变成身份偏见。

因此制度需要公开说明差异化标准：什么因素相关，为什么相关，谁来判断，如何复查。只要理由能够公开接受质疑，差异并不必然破坏法治。最危险的是未经说明的差异——同样行为，因为关系、印象或身份被悄悄处理成不同结果。

进一步深描四　为什么“处罚及时”也可能被滥用

前面我们肯定了贝卡利亚关于及时性的洞见：行为与后果相隔太久，关联会变弱，程序本身也可能成为额外负担。但“越快越好”同样可能变成新的口号。学校当天处分、公司当天开除、网络舆论当天定罪，看上去效率极高，却可能牺牲事实调查和申辩。

及时的真正意义不是抢速度，而是避免无理由拖延。它必须和准确、公平一起工作。一个错误但迅速的决定，不比一个稍慢但充分核查的决定更好。贝卡利亚反对漫长而不确定的司法等待，我们今天不能把这个原则倒装成“程序都是拖累”。

这里再次出现自由与效用的双层结构。快速处理可以降低成本、增强规则关联，但程序权利是硬约束。最合理的制度不是“最快”，而是“在保障必要核查和申辩的前提下，不无故拖延”。

这也适用于AI。机器能在几秒内给出评分，不意味着国家几秒内就应作出决定。技术速度不能自动改写法律速度。越快的工具，越需要制度主动留出审查时间，否则效率会把程序吞掉。

进一步深描五　规则为什么必须允许“例外”，又为什么例外最危险

所有一般规则都会遇到异常情形。医院规定晚上十点后禁止探视，但危重患者家属可能需要例外；公司禁止外带设备，但紧急维修可能例外；学校禁止手机，但学生健康监测设备可能依赖手机。没有例外，规则会伤害它原本想保护的价值。

问题是，例外又是任意最容易钻进去的地方。一旦“特殊情况”没有标准，关系好的人总能特殊，普通人只能守规则。于是制度陷入一个困难：没有例外会僵死，例外太多会腐烂。

贝卡利亚对司法解释的警惕正可以放到这里重新理解。他害怕的不是一切灵活，而是例外权被个体掌握后缺乏控制。现代制度的解决办法不是假装没有例外，而是把例外程序化：明确条件、要求理由、留下记录、允许复查，并定期检查某种“例外”是否已经频繁到应该修改一般规则。

从SDE看，例外是现实回写规则的一种入口。规则是已经稳定的显露，异常事实提示特征纠缠和差异序列发生变化。如果每次异常都硬塞回旧规则，制度失去适应；如果每次异常都任意处理，显露又无法稳定。成熟法治的技术就在两者之间：让例外能够进入，但不能偷偷改写全部规则。

进一步深描六　“给第二次机会”不是软弱，而是制度能力

处罚制度有一种深层诱惑：越不可逆，越显得有权威。开除、永久记过、终身禁入、彻底封号，都会让人感觉规则“有牙齿”。但不可逆也意味着制度承认错误的能力下降。一旦事实认错、环境改变、行为人真正改正，原处罚仍然继续支配未来。

贝卡利亚反对死刑时触碰到的核心之一，就是不可逆。死刑不仅重，而且一旦执行，所有后来的新证据、悔改和制度纠错都失去意义。这个结构可以类推到生活

中较轻但同样长期的标签：学生一次严重违纪是否应该永远影响升学评价？员工一次失误是否应该终身失去晋升机会？一个平台账号一次处罚是否应永久封禁？

“第二次机会”并不意味着不承担后果。它意味着后果完成以后，制度承认一个人还可以重新进入。真正的修复性秩序必须有“归队机制”：处分何时结束，记录何时封存，信誉如何恢复，重新获得资格需要什么条件。

这也是回写的另一面。过去行为当然会影响未来，但影响不是命运。如果制度只允许负面记录单向累积，不允许新的行为改写旧身份，它就把人变成档案的囚徒。一个会成长的人，被一个不会忘记的制度永久冻结，这本身可能成为新的不公。

因此，“会处罚”只说明制度有力量；“会结束处罚”才说明制度有边界；“会在证据变化时撤销处罚”则说明制度有理性。贝卡利亚真正开启的，不只是限制刑罚的力度，也是限制刑罚占据一个人生命的时间。

普通人现场　三个场景检验“处罚—信任—自由”

场景一　公司“末位淘汰”真的提高效率吗

末位淘汰看起来把后果变得极其确定：总有人必须离开，于是所有人都更努力。但如果岗位高度协作，员工会不会因此减少帮助同事？如果评分高度依赖主管印象，会不会刺激讨好？如果团队整体已经优秀，仍然强制淘汰末位，会不会把正常差异制造成失败？

一个制度不能只看第一层效果“大家更拼”，还要看第二层回写“大家怎样重新组织行为”。贝卡利亚提醒我们，处罚制度会塑造行动。组织设计同样如此。竞争可以提升产出，也可能制造信息封锁、短期主义和内部敌意。判断制度是否有效，不能只数被淘汰后绩效曲线有没有上升，还要看它留下了什么样的合作土壤。

场景二　学校“零容忍”为什么听起来最公平

“打架一律停学三天”，规则简单、明确、人人一样。它最大的优点是减少老师随意区别对待。但现实中，自卫和主动欺凌可能都出现肢体冲突；长期受欺负的学生反击，与多人围攻同学并不具有相同结构。零容忍通过消灭裁量减少任意，却也可能通过消灭情境制造新的不公。

因此，真正需要限制的不是一切判断，而是无理由的判断。学校可以保留清楚底线，同时要求对角色、先后、伤害程度和持续性进行事实调查。贝卡利亚把法官从任意中拉回规则，我们今天还要补上一句：规则要能容纳与责任真正相关的差异。机械相同并不是公平的最高形态。

场景三　社区把“信用差”居民列为重点对象

某社区把欠费、投诉、夜归频率、访客次数等数据综合成“治理风险分”，低分住户接受更多检查。系统没有判任何人犯罪，甚至可能真的帮助管理者提前发现纠纷。但一旦“低分”成为怀疑资格，居民就会发现自己需要为一个综合标签不断自证清白。

这时前文的“数字盖章”和本编的“身份回写”连在一起：分数先改变管理方式，管理方式又制造更多记录，记录再证明分数正确。破解这种循环，不是简单删除数据，而是限制标签功能：什么信息可以用于服务提醒，什么不能触发强制；多久复核；居民怎样查看和纠正；哪些变量不得因为与身份高度相关而进入判断。技术一旦参与权力，最重要的不是它会不会算，而是它被允许做什么。

◆ 阶段收束

走完第73—80章以后，《高鹏讲贝卡利亚：普通人都能懂》已经从“国家凭什么罚你”扩展成一个更普遍的问题：凡是一个人或一个机构拥有让别人承受不利后果的权力，它就不能只凭身份、善意或效率为自己正名。它必须说明目的、事实、必要性、比例和退出机制，并允许新的经验回头改写旧规则。

这不是把刑法原则机械搬进所有生活领域。家庭不是法院，公司不是国家，学校处分也不等同刑罚。不同场景的权力来源、强度和程序要求不同。但贝卡利亚留下的那种“先问凭什么”的姿势，可以跨场景工作。它让我们对处罚多问一步，对规则多问一步，也对自己的批判多问一步。

到这里，贝卡利亚已经不再只是十八世纪那位反酷刑、反死刑的青年作者。他更像是一个把“处罚必须解释”钉进现代制度入口的人。而本书真正要做的，也不是把他塑造成今天所有答案的祖先，而是把他开启的追问继续向前：国家为什么能罚，学校为什么能处分，公司为什么能制裁，父母为什么能限制；这些权力怎样获得边界，又怎样在错误以后学会改变。

> **带走的话**
>
> 一本书真正的反身性，不是最后写一句“我们也可能错”，而是把会让自己改判的证据提前列出来，并承诺证据出现时真的改。只有这样，解构才没有偷偷变成新的免检区。

> **小账**
>
> 这一章看见了什么？——F1至F11逐项列出状态，F2揭出例外之门由国家自判。
>
> 还要追问什么？——F7反馈会不会放大恐慌与偏见仍未检验，F11′待新样本。

第 81 章

SDE是不是也先放了一个“发生”

一路写到这里，本书最容易犯的错误终于轮到自己。我们用“发生”去追问贝卡利亚：刑罚权怎么来的，犯罪尺度怎么形成，法律怎样塑造行为，程序怎样影响案件事实，主体能力怎样在教育、贫富和制度里长出来。每一次追问都很锋利，因为它把一个看似天然的东西重新放回时间、关系和条件。

可一把刀用得太顺手，就会产生新的危险：凡是别人说“这是真的”“这是权利”“这是正义”，我们都回答“先别急，它是发生出来的”。如果这种回答自己不用说明来历、不接受反驳，那么“发生”就会变成全书最大的一块磐石。我们一边批评哲学家把自己的发生打造为发现，一边把“万物皆发生”打造为新的发现。

所以第81章先问一个不舒服的问题：SDE凭什么拥有最后解释权?

第一，发生能够说明来历，不自动说明正当。一种歧视有清楚的历史，一种酷刑制度也有清楚的形成路径，知道它怎么来的，并不能推出它应该继续存在。反过来，一项权利即使有历史发生，也不意味着它今天只能被当作偶然习惯。发生解释和规范理由必须分开，否则SDE会把“能解释”偷换成“能裁判”。

第二，发生并不自动比结构高贵。稳定的法律、权利、程序和概念之所以重要，正因为人不可能每天从零开始重新协商一切。贝卡利亚要求一般法律、罪刑法定和处罚比例，很多时候就是在为自由制造稳定结构。SDE如果只会把稳定重新溶解成流动，就会失去法治最重要的成果。发生学不是“反定型”；它真正要问的是，定型有没有来历、有没有边界、有没有回写通道。

第三，SDE自己的观察也有土。我们今天特别看重反馈、参与、算法风险、分布不平等和主体形成，与二十一世纪的数据技术、复杂治理和权利语言密切相关。若把这些关切写成超历史的真理，再回头批评十八世纪的人没有完整拥有它们，就会犯本书一直反对的倒装：把今天长出来的尺子，假装成从来就放在那里。

第四，SDE的三方程、三界、意义三律和九格，也不是世界自己给出的说明书。它们是研究工具，是一套让问题显形的工作结构。工具有效，必须靠它是否增加解释力、是否能提出可输的预测、是否在原典反证面前改判，而不是靠术语本身听起来完整。前文连续撤回“量刑尺”“主体先在”“自由中心”等判断，正是为了证明这套方法至少愿意被对象反过来修改。

因此，“发生”在本书里只能获得一种有限地位：它是一项持续追问来历和回写的研究纪律，不是世界的终极答案。它可以要求任何理论说明“怎么长出来”，却不能因此垄断“应该怎样”。它可以拆掉免检区，却不能宣布自己没有免检区。

◆ SDE的三条自限

第一，降格声明。本书关于贝卡利亚的中心诊断、收回动作和死法，都是可被后续原典、历史研究和制度经验推翻的工作判断，不是先验真理。第二，立足点自首。本书的关切来自AI时代、现代权利法治与发生学方法的交叉处，因此特别敏感于反馈、算法、分布和主体形成；这既是优势，也是偏向。第三，不封顶条款。贝卡利亚不是被本书“最终解释”的对象，SDE也不是西方刑罚思想的终点。下一次更好的研究，完全可能把本书今天的“发现”重新查成一次发生。

> 带走的话
>
> 当“发生”被用来解释一切时，它就最接近变成新的“发现”。SDE真正的自我保护，不是声称自己只是工具，而是让自己的判断有来源、有失败条件、有改判记录。

小账

这一章看见了什么？——发生能说明来历，不自动说明正当，也不比结构高贵。

接缝在哪里？——SDE自己的关切也有土，来自AI时代与现代权利语言。

第 82 章

给SDE自己写死法

如果一种理论只会告诉别人怎么死，却从不说明自己会怎么死，它迟早会变成新的免检区。于是最后一章不再给贝卡利亚加任何罪名，只给本书和SDE列出它们的死法。

第一种死法，叫“发生万能化”。任何概念都被解释成历史条件和关系的产物，最后没有一个判断能够区分解释与辩护。只要发生过，就被当成合理；只要能找到来历，就觉得任务完成。这时发生学失去规范能力。作废条件很清楚：如果SDE不能独立说明为什么某些已经发生的制度仍应被废除，为什么某些有历史来历的权利仍应优先，它就不能单独承担法律正当性理论。

第二种死法，叫“回写万能化”。每看到反馈，就宣布传统理论缺少回写；可若所有制度变化都能被解释成回写，概念就失去区分力。真正有意义的回写必须指出：什么结果回到什么结构，通过什么通道改变后续选择，什么证据表明这种改变发生。如果只说“社会会变化”，那不是理论，而是常识。

第三种死法，叫“九格填满症”。九格是为了逼研究者检查遗漏，不是为了证明每个对象都恰好占满九个格子。如果为了形式漂亮，把任何材料都硬塞进创造、自由、幸福以及理念、现实、自我，九格就从诊断工具变成分类游戏。它的失败条件是：删掉某一格以后，理论解释力没有任何变化；或者不同研究者可以任意填表而没有可争论标准。

第四种死法，叫“后见之明”。站在二十一世纪，用现代法治、心理学和AI经验批评十八世纪思想家没有今天的制度，会显得总是正确，却没有思想史价值。发生学尤其应该警惕这一点，因为它最强调“土”。如果不把贝卡利亚放回他的土，就没有资格用土去批他。

第五种死法，叫“永不认输”。一套方法如果每遇到反证就换一个更抽象的说法继续成立，它实际上不可证伪。本书必须把前面的改判留下来：我们撤回“贝卡利亚迷信精确量刑”；收窄“主体先在”；把“自由中心”改为“自由限权型公共幸福中心”；再把“搭了看台，少了楼梯”从历史指控降为现代延伸诊断。这些撤回不是修辞，而是SDE继续有资格使用的条件。

因此，给SDE自己写的总死法可以压成一句：当它只剩下一把永远不会输的刀，它就已经死了。

一套活的方法，必须允许三件事同时存在：有承重判断，所以不是无穷犹豫；有推翻条件，所以不是教条；有改判记录，所以不是把失败偷偷改写成成功。贝卡利亚之所以适合做这场实验，正因为他本人也在做类似事情：用必要性、比例、一般法律和公共幸福把旧刑罚送进理由审查。我们只是把同一种审查再转向审查者。

本章不再另列一套推翻条件。统一清单及其“已触发／未触发”状态见第80章。SDE若拒绝接受其中已经触发的改判，本书最后这一章就只剩姿态。

> 带走的话
>
> 真正的反身不是在书末写一句“我们也可能错”，而是把自己会在哪些证据面前输写清楚。不能输的方法，不配拿来判断别人。

> **小账**
>
> 这一章看见了什么？——发生万能化、回写万能化、九格填满症等五种死法。
>
> 还要追问什么？——删掉某一格后解释力变不变，填表有没有可争论的标准。

结 语

国家凭什么罚你?

我们从一间可能使用酷刑的十八世纪法庭出发，最后走到了算法、学校、公司、家庭，也走回了研究者自己。两百多年过去，工具、制度和语言都变了，但贝卡利亚留下的那个问题没有过期：国家凭什么罚你?

这个问题的锋利，不在于它预设国家不能处罚。恰恰相反，任何共同生活都需要处理伤害、欺诈、暴力与破坏规则的人。贝卡利亚真正改变的是举证责任。过去，处罚权像国家身体上的器官；从他以后，处罚越来越需要说明来源、目标、比例、程序和必要性。贝卡利亚在全书结论中把这些要求压成一句定理：刑罚应当公开、及时、必要，在既定情形下尽可能轻，与罪行相称，并由法律规定（原典锚点：剑桥版第47章“结论”）。国家不能只说“我有权”，还要解释“为什么在这里、用这种方式、对这个人、做到这个程度”。

全书因此形成了一条长链。处罚权必须有出生证明；犯罪不能只是权威受冒犯；刑罚必须有梯度；规则必须事先公开；解释不能偷偷造罪；程序不能用酷刑制造真相；威慑不是疼痛一个点，而是确定性、及时性与强度组成的路径；死刑必须同时通过权利与必要的双重审查；法律会反过来制造身份、信任与新的行为；预防、教育与公共幸福说明刑法从来不是只在判决那一天工作；AI又把贝卡利亚追求的可计算、可预期推到一个危险的新高度。

经过原典不断反击，本书自己的判断也发生了变化。我们最早想把贝卡利亚写成一个把国家发生化、却把自由主体留在发生之前的人。后来发现这太粗。他明明知道教育、习惯、荣誉、秘密制度和法律确定性会塑造行动者。于是批评缩小为规范主体与经验主体的接口；再后来，公共幸福显示出比自由更宽的目标位置，我们把中心改为“自由限权型公共幸福中心”。最后又发现，不能把十八世纪没有参与式治理和现代反馈机制直接判成历史缺陷，于是“搭了看台，少了楼梯”被降格为

一种面向今天的延伸诊断。本版最后又把自由硬闸拆开来看：锁是硬的，钥匙却在国家手里。

现在，全书可以给贝卡利亚留下五条最终判断。

◆ 五条最终判决

第一判：贝卡利亚最大的革命，不是把刑罚变轻，而是逼刑罚权拿出出生证明。 国家处罚人不再只是起点，而成为必须持续证明的结论。

第二判：他的现代性不只是反酷刑、反死刑，而是把刑罚从情绪化报复改写成受一般法律、必要性、比例和公共理由约束的制度行为。

第三判：他已经多次看见发生。 工具会改变证据，规则会改变行为，秘密会改变社会关系，过度禁止会制造新的违法，教育会改变人的选择。把他简单写成静态理性主义者，是误读。

第四判：贝卡利亚更适合被理解为“自由限权型公共幸福中心”；而他那道自由闸是锁钥同手的——锁由自由铸成，钥匙在国家手里。

公共幸福提供制度方向，自由、罪刑法定、必要性和比例构成不可轻易越过的强制闸门；两者之间始终存在张力，正是这套思想仍然活着的原因。但这道闸门并非处处同高：对普通犯罪，它几乎封死；在“足以颠覆国家”的非常情形，它留了一道门，而判断门该不该开的，恰恰是国家自己。第9章问过“谁有权宣布必要”，这处接缝在全书最后一道闸门前又回来了。锁和钥匙在同一只手里，本书把这叫作“锁钥同手”。这一层，是本书用“谁上锁、锁多硬、钥匙在谁手里、门朝哪边开”这组问题读出来的；它还接受了一次历史检验：按事先写定的条件，“死刑先从国家安全这扇门回来”的顺序预测部分失败；留下的是一个待检的新假设——钥匙在行政手里时，国家安全这扇门几乎总在首批打开的门里（F11′）。

第五判：他真正留给今天的，不是一套永远正确的刑罚清单，而是一种权力姿势——处罚者必须解释，强制者必须举证，制度必须允许纠错。

AI时代越能预测、分类和干预，这条要求越不能降低。

五条判决合起来，就是本书的判词，与前言、导读和第 56 章一字不差：**他逼国家为手里那把刀出示出生证明，却把最硬一道闸的钥匙，留在了国家手里**。

随身带走三个问题：你凭什么？非这样不可吗？钥匙在谁手里？

如果要把全书压成一句更普通的话，那就是：一个社会是否文明，不只看它怎样对待好人，也看它在确信某个人做错事以后，是否仍然愿意约束自己的手。

贝卡利亚从国家手里夺下的，并不是那把刀。他夺下的是一种更隐蔽的特权——“这把刀不需要解释”的特权。

而这也应当成为本书对自己的最后要求：任何一把理论的刀，包括SDE，都不能免于解释。

参 考 书 目

一、原典与权威版本

1. Cesare Beccaria, Beccaria: “On Crimes and Punishments” and Other Writings, ed. Richard Bellamy, trans. Richard Davies, Cambridge University Press, 1995（online 2012）. 本稿英译核验与章序的主要底本。
2. 贝卡利亚：《论犯罪与刑罚》，黄风译，北京大学出版社，2008（47章版；ISBN 978-7-301-14228-8）。中文引文出版前以此版统一终审。
3. Cesare Beccaria, Dei delitti e delle pene, 1764；原典的版本差异以剑桥版“Note on the Texts”及出版前意大利国家版终审为准。
4. Immanuel Kant, The Metaphysics of Morals, Doctrine of Right, Ak. 6:331–337（刑罚、报应与死刑部分）.
5. Jeremy Bentham, An Introduction to the Principles of Morals and Legislation（效用原则与刑罚限制）.
6. Jean-Jacques Rousseau, Du contrat social, 1762, Book I ch.6；Book II ch.5（完全转让与生命/死亡权问题）.
7. Pietro Verri, Osservazioni sulla tortura（1770年代成稿，1804年身后首次出版）；与1630年米兰“涂油者”案相关。
8. Cesare Beccaria, Paolo Risi & Francesco Gallarati Scotti, Voto sulla pena di morte, 24 January 1792，in Atti di governo, a cura di R. Canetta, Edizione nazionale delle opere di Cesare Beccaria, vol. XVI, t. 2, Milano: Mediobanca, 2009, p. 972 起；英译见 Garnsey (ed.), Against the Death Penalty, 2020。
9. Alessandro Manzoni, Storia della colonna infame, 1840.
10. Voltaire, Traité sur la tolérance, 1763（卡拉斯案）.

二、刑罚哲学、思想史与社会科学研究

1. Ronald Dworkin, “Rights as Trumps,” in Jeremy Waldron (ed.), Theories of Rights, Oxford University Press, 1984, pp. 153–167.
2. Robert Nozick, Anarchy, State, and Utopia, Basic Books, 1974, esp. “Moral Constraints and the State.”

3. Michael S. Moore, “Torture and the Balance of Evils,” 1989；后收入 Placing Blame: A Theory of the Criminal Law, Oxford University Press.
4. A. John Simmons, “Locke and the Right to Punish,” Philosophy & Public Affairs, Vol. 20, No. 4 (1991), pp. 311–349.
5. R. A. Duff, Punishment, Communication, and Community, Oxford University Press, 2000.
6. Franco Venturi, Settecento riformatore, Torino, 1969.
7. 张智：《贝卡里亚启蒙思想的苏格兰渊源——以〈论犯罪与刑罚〉为中心》，《华东师范大学学报（哲学社会科学版）》2017年第3期，第42—47页。
8. 施鹏鹏：《贝卡利亚刑事诉讼思想论略》，《暨南学报（哲学社会科学版）》2008年第3期，第40—45页。
9. 谷桢：《如何走出贝卡利亚？——解析刑罚的严厉性、确定性和及时性》，《东岳论丛》2023年第12期，第182—190页。
10. Richard Bellamy, “Introduction,” in Beccaria: “On Crimes and Punishments” and Other Writings, Cambridge University Press, 1995, pp. ix–xxx.
11. H. L. A. Hart, “Prolegomenon to the Principles of Punishment,” in Punishment and Responsibility, 2nd ed., Oxford University Press, 2008, pp. 1–27（原文1959）.
12. H. L. A. Hart, “Bentham and Beccaria,” in Essays on Bentham: Jurisprudence and Political Philosophy, Oxford University Press, 1982, pp. 40–52.
13. Bernard E. Harcourt, “Beccaria’s On Crimes and Punishments: A Mirror on the History of the Foundations of Modern Criminal Law,” 2014.
14. Michel Foucault, Discipline and Punish: The Birth of the Prison, 1975.
15. Edwin M. Lemert, Social Pathology, 1951；Howard S. Becker, Outsiders, 1963（标签与偏差身份研究）.
16. Daniel S. Nagin, “Deterrence in the Twenty-First Century,” Crime and Justice, Vol. 42 (2013), pp. 199–263.
17. Peter Garnsey, Rethinking Capital Punishment: The Pre-History of the Abolition of the Death Penalty, Cambridge University Press, 2026.
18. Peter Garnsey (ed.), Against the Death Penalty: Writings from the First Abolitionists—Giuseppe Pelli and Cesare Beccaria, Princeton University Press, 2020（含1792年Voto英译）.
19. Tullio Padovani, La pena carceraria, Pisa, 2014, p. 47（“真正的废死者是利奥波德”一说）.

20. John Rawls, A Theory of Justice, Harvard University Press, 1971; rev. ed. 1999（第一原则的词典式优先）.
21. Paul Pierson, “Increasing Returns, Path Dependence, and the Study of Politics,” American Political Science Review, Vol. 94, No. 2 (2000), pp. 251–267.
22. Maurice Halbwachs, Les cadres sociaux de la mémoire, Paris: Alcan, 1925.
23. Keith A. Findley & Michael S. Scott, “The Multiple Dimensions of Tunnel Vision in Criminal Cases,” Wisconsin Law Review, 2006, pp. 291–397.
24. Carl Schmitt, Politische Theologie: Vier Kapitel zur Lehre von der Souveränität, München/Leipzig: Duncker & Humblot, 1922；英译 Political Theology, trans. George Schwab, MIT Press, 1985.
25. Giorgio Agamben, Stato di eccezione, Torino: Bollati Boringhieri, 2003；英译 State of Exception, University of Chicago Press, 2005.
26. Clinton L. Rossiter, Constitutional Dictatorship: Crisis Government in the Modern Democracies, Princeton University Press, 1948.
27. John Ferejohn & Pasquale Pasquino, “The Law of the Exception: A Typology of Emergency Powers,” International Journal of Constitutional Law, Vol. 2, No. 2 (2004), pp. 210–239.
28. Bruce Ackerman, “The Emergency Constitution,” Yale Law Journal, Vol. 113, No. 5 (2004), pp. 1029–1091.
29. General Agreement on Tariffs and Trade (1947), Article XXI；WTO Panel Report, Russia — Measures Concerning Traffic in Transit, WT/DS512/R, 5 April 2019.
30. John Stuart Mill, On Liberty, London: John W. Parker and Son, 1859.
31. Bernard E. Harcourt, Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age, University of Chicago Press, 2007.
32. Kristian Lum & William Isaac, “To Predict and Serve?”, Significance, Vol. 13, No. 5 (2016), pp. 14–19.
33. Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger & Suresh Venkatasubramanian, “Runaway Feedback Loops in Predictive Policing,” Proceedings of Machine Learning Research, Vol. 81 (2018), pp. 160–171.

三、历史制度、判例与现行法

1. UN Human Rights Committee, General Comment No. 36 (2018), Article 6: Right to Life, CCPR/C/GC/36.
2. Treccani, “Beccaria, Cesare”（生平、行政职务及1792年死刑Voto条目）.

3. Treccani, “Verri, Pietro”（米兰启蒙、反酷刑写作与相关生平材料）.
4. Archivio di Stato di Firenze, “La legge Leopoldina del 1786”（1786年11月30日托斯卡纳刑事改革，废除死刑与酷刑）.
5. 托斯卡纳摄政会议1790年6月恢复死刑令及1795年8月30日斐迪南三世扩大令；参见 Biblioteca Forteguerriana, “Fame e fede: il 1790 e la crisi dell’esperienza riformatrice”。出版前核对原始法令。
6. Codice penale per il Regno d’Italia（Codice Zanardelli）, 1889；Legge 25 novembre 1926, n. 2008, “Provvedimenti per la difesa dello Stato”；Codice penale, R.D. 19 ottobre 1930, n. 1398（Codice Rocco）.
7. 瑞士联邦宪法（1848；1874年第65条及1879年全民公决修正）.
8. Protocol No. 6 (1983) and Protocol No. 13 (2002) to the European Convention on Human Rights；Decreto legislativo luogotenenziale 10 maggio 1945, n. 234.
9. 奥地利1933年11月恢复即决审判死刑的政府紧急令；参见 Wien Geschichte Wiki, “Abschaffung der Todesstrafe”。第56章与附录所用其余各案的来源链接，随本书网页版资料页公布。
10. 艾奥瓦州1872年废除死刑与1878年恢复死刑立法；科罗拉多州1897年废除与1901年恢复立法（据两州死刑史资料，出版前核对法令号）.
11. Catherine II, Nakaz / Instruction to the Legislative Commission, 1767.
12. 约瑟夫二世刑事改革及1787年刑法典，作为哈布斯堡刑罚改革背景材料。
13. State v. Loomis, 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749.
14. Regulation (EU) 2024/1689 (EU Artificial Intelligence Act), Article 5(1)(d).
15. 《中华人民共和国刑法》（现行修正文本），第3条（罪刑法定）、第5条（罪责刑相适应）.

四、SDE方法

1. 《sde-philosopher-for-everyone 合订本 v1.5》：看见—收回—接缝—中心诊断—反证—推翻条件—反身审查。
2. SDE Universes 网站现行本体论材料：S＝Show（显露），D＝差异序列，E＝特征纠缠；S＝F(D,E)，D＝G(S,E)，E＝H(S,D)；123原理与回写机制。

附 录

附录一　判词演变：原典如何一次次把本书顶回来

本表不把改判当作失误清单，而把它作为全书方法证据：一个判断只有在遇到更强材料时真的收回，才配称“可证伪”。

初判	反证／竞争解释	改判后表述	依据章节	F编号	当前状态
贝卡利亚相信“精确量刑尺”	原典明确承认政治计算只能处理概率，不能几何精确	追求可比较、可预期的比例结构，不等于数学精确	第13—14、57章	原典反证	撤回初判
“国家发生，主体先在”	教育、激情、制度塑造主体的文本证据	规范自由资格与经验自由能力需要建立接口	第40、56章	F1、F3	实质改判
自由中心主义	公共幸福贯穿处罚、预防与教育；解释史已有契约/功利双线	公共幸福定方向，自由设硬闸；中心用否决权＋例外门槛检验	第44、47、50章	F2	保留为解释性压缩
“搭了看台，少了楼梯”是贝卡利亚本人的缺口	1771年后长期行政实践、1791年司法改革职务	仅指1764文本未形成现代制度化反馈架构	第56章	F3	强判断被推翻

初判	反证／竞争解释	改判后表述	依据章节	F编号	当前状态
本书提出“两阶段正当化”	Hart一般正当化目的／分配的混合论先占	承认Hart先占；“第零闸”降级为贝卡利亚“授权先于目的”的思想史问题次序	第64章	F5	理论增量撤回，解释框架保留
主死法＝贝卡利亚把必要性焊死	其改革实践与文本中的经验、概率意识	改为继承贝卡利亚语言的制度可能发生的“必要性封顶”风险	第78章	F4、F6	降级为可检验风险
自由硬闸没有缺口	1764年死刑章与1792年Voto的“足以颠覆国家”例外；Padovani等的批评	硬闸对普通犯罪几乎封死；例外开在“足以颠覆国家”之人与罪，由国家自行判定	第30、56章	F2	部分改判
贝卡利亚往往先问授权、再问目的	作者六章编码；第二编码者十八章扩样：授权在先10、效用在先3、融合5	在“权力归谁”与“定罪之前的人身”两类问题上，授权先于目的	第64章	F9	收窄

初判	反证／竞争解释	改判后表述	依据章节	F编号	当前状态
自由硬闸的例外只是门槛问题	1764年死刑章与1792年Voto；施米特、阿甘本、“任何人不得做自己案件的法官”	锁钥同手：锁由自由铸成，钥匙在国家手里，门朝死刑开；与罗西特、自裁条款、有利/不利类推划界后，只以“位置×方向”复合读数保留增量	第44、56章	F2、F11	诊断保留；顺序预测部分失败，改为待检的F11′
失效图谱是SDE的增量	哈考特“棘轮效应”；Lum & Isaac；Ensign等“失控的反馈回路”及其失控条件	回写中断与顺序主张均已被先占；失效图谱只作教学框架保留，不再计为增量	第78章	F10	撤回增量
“否决权＋例外门槛”足以诊断中心	只比门槛，读不出贝卡利亚与罗尔斯的差别	补入第三参数“例外位置”，经三例校准；盲编码改正康德读数，再补第四问“门朝哪边开”	第44、56章	F8	扩充、校准并经盲编码改正
推翻条件只是尾声声明	第80章逐条检验	十一项统一清单，记录已触发／部分触发／未检验	第80章	F1—F11	方法开始执行

附录二　死刑废止—恢复事件编码表（1786—1970）

本表由一位不知道本书假设的编码者（大模型实例）按统一口径整理，用于第56章F11的检验。收录标准：和平时期对普通犯罪或一切犯罪废除死刑（法律废除或正式暂停），后又在法律上恢复。“类别”按首批恢复罪名的法律文字编码。把握为“低”的4件未计入第56章的主统计。各条来源见本书网页版资料页，出版前应逐条核对原始法令。

法域	废除	恢复与形式	主导	首批恢复罪名	类别	把握
托斯卡纳	1786，全部	1790，大公诏令	行政	煽动民众暴乱	国安	高
奥地利（哈布斯堡）	1787，普通程序	1795，皇帝特许令	行政	叛国	国安	高
罗德岛	1852，全部	1872，州议会	立法	服无期徒刑者杀人	普通	中
罗马共和国→教皇国	1849，全部	1849，复辟恢复旧法	行政	旧法整体恢复	两类	低
托斯卡纳	1847—48，全部	1852—53，大公令与刑法典	行政	冒犯君主、预谋杀人	两类	低
秘鲁	1856，宪法	1860，宪法	立法	加重谋杀	普通	中
艾奥瓦	1872，全部	1878，州议会（私刑）	立法	谋杀	普通	高
瑞士（联邦）	1874，全部	1879，全民公决	立法	谋杀（政治罪明文排除）	普通	高
缅因	1876，全部	1883，州议会	立法	谋杀	普通	中
哥伦比亚	1863，宪法	1886，宪法	行政	叛国、弑亲、谋杀、纵火等	两类	中
科罗拉多	1897，全部	1901，州议会（私刑）	立法	谋杀	普通	高

法域	废除	恢复与形式	主导	首批恢复罪名	类别	把握
亚利桑那	1916，公民创制	1918，公民创制	立法	谋杀	普通	高
华盛顿州	1913，全部	1919，州议会	立法	一级谋杀	普通	中
密苏里	1917，全部	1919，州议会（私刑）	立法	谋杀	普通	高
田纳西	1915，谋杀	1919，州议会	立法	谋杀	普通	低
俄勒冈	1914，公投	1920，公投修宪	立法	一级谋杀	普通	高
葡萄牙	1911，宪法	1916，国会立法（参战后）	混合	战区叛国等军事罪	国安	高
意大利王国	1889，普通刑法	1926，国家防卫法	行政	危害国王、首相与国家安全	国安	高
阿根廷	1921，全部	1930，军政府戒严令	行政	危害治安、财产、人身	两类	低
奥地利	1919，普通程序	1933，政府紧急令（即决法庭）	行政	谋杀、纵火、恶意毁坏财物	普通	高
西班牙	1932，全部	1934，议会立法	混合	恐怖主义、暴力抢劫	两类	中
堪萨斯	1907，全部	1935，州议会	立法	一级谋杀	普通	中
巴西	1889—91，普通罪	1937，瓦加斯颁行宪法	行政	分裂国土、叛乱、颠覆及极端残忍杀人	两类	中
罗马尼亚	1864—66，和平时期	1938，王室独裁宪法	行政	危害王室与高官、政治谋杀、抢劫杀人	两类	高
南达科他	1915，全部	1939，州议会	立法	谋杀	普通	中
挪威	1902—05，民事罪	1941—42，流亡政府临时令	行政	叛国、通敌	国安	高
荷兰	1870，民事罪	1943，流亡政府王令	行政	战争罪、叛国	国安	中
丹麦	1930/33，民事罪	1945，议会立法	立法	通敌叛国	国安	高

法域	废除	恢复与形式	主导	首批恢复罪名	类别	把握
苏联	1947，和平时期全部	1950，最高苏维埃主席团令	行政	叛国、间谍、破坏	国安	高
新西兰	1941，谋杀	1950，政府立法	立法	谋杀	普通	中
阿根廷	1921，全部	1956，戒严令	行政	军事叛乱	国安	中
锡兰	1956—58	1959，议会立法（总理遇刺后）	混合	谋杀	普通	高
古巴	1940，宪法	1959，革命政府根本法	行政	前政权军警罪行、叛国、间谍	国安	高
特拉华	1958，全部	1961，州议会	立法	谋杀	普通	中
巴西	1946，宪法	1969，军政府法令	行政	颠覆、恐怖主义、武装叛乱	国安	高
阿根廷	1921，全部	1970，军政府法令	行政	绑架、持械袭击军警设施	两类	中