

# 蚕茧里的利刃：言论自由如何被自己的保护壳闷死

致死的力量来自定义的高度，而不是来自诋毁的深度

作者 高鹏 · SDE 学员 · 约 2.1 万字 · 发表于2026年7月31日 · 深化增补版

**摘要** 本文诊断言论自由在当代面临的一种尚未被命名的退化机制。既有研究已正确地将这一危机的根源从“外部压制”扭转为“捍卫的共谋”——自由的捍卫话语在保护它的同时，也抑制了它的日常使用。本文接受这一判断的全部事实基础，但认为它尚未追问一个更深的问题：捍卫话语共同保护的那个东西，究竟是什么？本文给出的回答是：它们保护的并不是自由本身，而是一套关于“自由应该是什么”的被固定下来的知识；这套知识在结构上必定排异一切不完美、不确定、还在生成中的自由实践——即“被呼吸的经验”。这种排异不是压制，而是免疫清除：它不否认自由的价值，恰恰相反，它把自由定义得如此完整，以至于大多数日常试探都被判定为“不配叫作言论自由”。本文将这一机制的产物命名为“茧室”——一个由捍卫叙事织成的透明保护壳，它在成功抵御外部审查的同时，其内壁系统性地勒死了它本应保护的活物。茧室的核心致死机制不是贬低自由，而是过高定义自由。这一诊断将言论自由的退化从“被冷落的神像”重新描述为“被自身免疫系统攻击的活体”，并以此切断它与“规训”及其他邻近概念的最后共享神经。结语部分，本文在承认茧室已不可逆地构成自由的第二自然的前提下，提出三条再生原则，并给出可被公开检验的证伪条件。

**关键词** 言论自由；茧室；被守护的知识；被呼吸的经验；试探；免疫排异

## 一、问题：保护壳为什么变成了凶器？

这几乎是一幅完美的画卷。在法学院的阶梯教室里，在公共知识分子的对谈中，在纪念日的情感散文与社交媒体的接力转发里，言论自由被擦拭得如此光亮，光亮到每一个试图伸手去碰它的人，都会在指尖触到它之前的那一瞬犹豫——我的手是不是太脏了？

已经有敏锐的观察者将这幅画卷背后的退化机制做了清晰的解剖。其核心判断是：言论自由正在从一种日常的、用于摩擦和试探的“实践性权利”，退化为一件被

高高供奉的“圣物性权利”。这种退化不是来自敌人的攻击，而是来自爱人的过度呵护——冷战的恐惧逼出了“最后防线”叙事，殉道者的牺牲铸造了情感门槛，正确性话语的日常泛化完成了对试探的最后一击。捍卫，在毫无恶意的情况下，成了抑制的共谋。

本文接受这一判断的全部事实基础。它所追踪的三步圣物化过程——从“最后防线”的绝对化到殉道者叙事的情感聚焦，再到正确性叙事的日常泛化——在发生学的精确性上，已经足以将言论自由的当代困境从“被外部压制”的发现学叙事中彻底松绑。它逮住了一个连最善意的捍卫者也从未质疑的底层预设：捍卫自由和使用自由是可以无矛盾地并行的事业。然后它撤销了这个预设，让被它遮蔽的矛盾自己结算出一个新显露来。

但这个新显露——“圣物化”——本身还藏着一个更深的矛盾没有被追问。而这是本文的起点。

让我们沿着先行研究的论证多走一步。它说，言论自由被抬高成了一件圣物。那么请问：那些抬高的手，他们共同保护的那个东西，到底是什么？

一个直觉上的答案是：他们保护了“言论自由”这个价值本身。这个答案没有错。但它把两种完全不同的被保护物压进了同一个名字里。一种，是自由作为被呼吸的经验——那个在不确定中试探、在摩擦中修正、在说错之后公开承认“我上次想错了”的活活动。另一种，是自由作为被守护的知识——一套关于自由应该是什么、应该如何被尊重、应该如何被回忆的权威表述。这两种东西共享同一个名字，但它们在存在形态上的差异，比一棵活着的树和一张树的标本之间的差异还要深。被呼吸的经验需要的是不完美的、日常的、沾着汗渍的使用；被守护的知识需要的是精确的、庄严的、不容置疑的传承。前者在摩擦中生长，后者在擦拭中发光。

如果这个区分站得住，那么先行研究的“圣物化”命题就可以被重新表述为：言论自由的“被守护的知识”那一面，在特定的历史矛盾中被不断加厚、不断抬高，直至它压过了“被呼吸的经验”那一面，把自由的整体存在形态从一件需要被反复使用的活工具，变成了一件只需要被反复确认其价值的死标本。

但这个重新表述并不只是一个术语替换。它打开了一个先行研究没有打开的新问题：被守护的知识和被呼吸的经验之间，是不是不仅仅存在“此消彼长”的竞争关系，而存在一种更致命的、结构性的排异反应？

换句话说：当被守护的知识变得足够厚、足够权威、足够不容置疑的时候，它是不是会开始把一切被呼吸的经验——一切不完美的、试探性的、还在生成中的自由实践——都识别为对自身完整性的威胁？就像一具被制造得过于完美的义肢，它不但不能替代血肉之躯，还会在每一次使用者试图用残肢去触碰世界时，用自己坚硬的边缘割伤残肢？

本文的判断是：会的。而且这种排异反应，才是言论自由实践性退化最根本的发生机制。“圣物化”命名了自由被抬高的后果，但它没有命名那个被抬高的东西内部发生的质变——保护壳是怎么变成凶器的。这个质变，需要一个属于它自己的名字。

但在给出这个名字之前，我必须先追踪这个壳究竟是怎么被织出来的。在每一个历史关节点上，织壳的丝是从什么样的矛盾里抽出来的？它当时保护了什么，又抑制了什么？这个追问，不是要建立一座定义的大厦，而是要追踪一个结构的发生。因为发生学从不相信一个被端上桌的完美定义——它只相信从矛盾的结算里被逼出来的新形态。

## 二、织壳的丝：四步发生学追踪

本章的工作是追踪“被守护的知识”是如何在特定的历史矛盾中，一步一步被逼出它的完整性要求，并在这个完整性要求内部孕育出排异机制的。以下四步不是按线性时间排列的因果锁链——它们在历史上相互交叠、彼此加强——但每一步都标志着一个新的结构要素被焊进了保护壳的内部，使得保护和致死这两面越来越难以分开。

必须在一开头就说明白的是：在每一步中，那个被逼出来的东西在当时都是合理的、甚至别无选择的。冷战时期把自由抬成“最后防线”是有道理的——面对一个以整个意识形态体系为对手的外部威胁，把自由说成一件日常工具确实显得软弱无力。殉道者叙事的情感聚焦是有道理的——那些为自由付出真实代价的人值得被铭记。正

确性话语在日常互动中的扩散也是有道理的——在弱势群体长期被系统性消声的社会里，让人们学会“正确地说话”本身就是一种权力再分配。本文对这一切没有任何道德指责的意图。正因为每一步都是合理的，它们共同的产物才如此坚不可摧。它不是任何一个坏人故意造出来的牢笼。它是一群好人用各自最好的意图，一根接一根织出来的蚕茧。

## 2.1 第一根丝：核弹与镜子的双重逼迫

二战结束后的第一个十年里，美国言论自由的话语发生了一个在历史上看极为迅速的转变。在二十世纪初期，第一修正案的判例还集中在相对狭窄的问题上：煽动叛乱的标准、事前审查的合宪性、淫秽物品的定义。言论自由被法学家和法官们当作一件需要在具体案件中反复调试边界的法律工具。它当然被尊重，但那是一种工程师对精密仪器的尊重——用时拿出来校准，用完放回工具箱。

原子弹和铁幕改变了这一切。冷战不是一个可以被“用完放回去”的外部威胁。它是一种持续的、弥漫的、关乎整个文明存亡的紧张状态。在这场对抗中，苏联体制对言论的全面控制成为美国必须与之切割的根本标志。你不是在为自己争取多一份或少一份的言论空间而辩论——你是在为整个自由世界做定义。你如何对待言论自由，决定了你是一个什么样的人、你属于哪一边。

这就是“最后防线”叙事的真正起点。它不是在某个具体的判例或某篇社论中被一次性发明的。它是从无数个场合同步逼出来的。在国会听证会上，拒绝回答“你是共产主义者吗”的证人援引第一修正案时，他不再只是在主张自己的法律权利，他在表演“一个自由人如何对抗专制”。在公共辩论中，当言论自由的批评者被质问“你想让美国变成苏联那样吗”时，言论自由就不再只是一项可讨论的政策——它变成了一条不可讨论的底线。

这个过程最有揭示力的一个细节，来自1950年代最高法院的一系列判词。在大法官们的手笔下，第一修正案开始被附加上一种此前从未有过的语体——不再只是“国会不得立法……”，而是“我们作为一个自由民族的承诺”、“我们的生活方式中最珍贵的堡垒”。在1951年的丹尼斯诉合众国案中，这种修辞达到了一个标志性的浓

度：多数意见在维持对共产党领导人的定罪时，仍然用大量篇幅赞颂第一修正案是美国“不可动摇的根基”，而异议意见则反过来用同样的修辞火力攻击多数意见正在“背叛”这个根基。双方争夺的，已经不是具体言论自由的边界在哪里，而是谁才是这个“自由民族”真正的代言人。

在这个句式里，使用自由和维护自由被第一次焊在了一起。你每一次开口说话，都同时在说两件事：你在说你想说的那句话；你在用你说话这个行为，来证明你是“我们”的一员，而不是“他们”的一员。

这根丝一旦被抽出来，保护壳的第一层就结好了。在它诞生的那一刻，这是一个极其有效的防御工事——它让言论自由不再只是法律体系中的一条修正案，而变成了整个西方阵营的集体身份象征。但它同时也在保护壳的内壁上嵌进了第一排看不见的倒刺：如果言论自由是我们的集体身份，那么任何不完美的使用——说错话、说不合时宜的话、说让别人不舒服的话——就不再只是个人的失误，而是对整个集体身份的玷污。你不再只是你自己。你每一次开口，都代表着我们所有人。这个压力不是法律强制的，它没有被写进任何一条法规里。但它从这一刻起，变成了自由使用者肩上永远不会卸下的那份重量。

## 2.2 第二根丝：殉道者叙事中的牺牲与距离

如果说第一步把“使用自由”和“代表自由”焊在了一起，那么第二步就是在这道焊缝上涂了一层情感的釉——让它不但不可拆卸，而且不可触碰。

从1960年代的民权运动到1970年代的反战抗议，再到此后的一系列社会运动，美国言论自由的历史上出现了一批真实的牺牲者：被监禁的异议者、被流放的作家、在街头被暴力镇压的抗议者。他们的牺牲是真实的，他们的勇气值得被铭记，他们的名字应当被写进历史。本文对这一笔没有任何保留。

但“如何纪念牺牲者”这件事，从来不是一个中性的技术问题。当一个社会选择以某种特定的方式反复操演对牺牲者的纪念时，这种操演本身，就在不断重新定义“什么样的人配得上言论自由”。

让我们仔细看一看这个操演的具体形式。在纪念日的公开演讲中，在纪录片的旁白里，在教科书的相关章节中，殉道者的故事几乎总是按照同一个叙事模板被讲述：一个人，面对强大的压迫力量，说出了他明知道会招致严重后果的话。焦点近乎始终集中在两个元素上——牺牲和确定性。牺牲：他为此付出了什么样的代价。确定性：他在说出的那一刻，非常清楚自己在说什么，他说的每一个字都是他愿意为之付出生命代价的真理。

这个叙事模板的道德力量是巨大的。它制造了一种崇高感，让每一个听到故事的人都为之动容。但在这个崇高感的阴影里，一个未被言明的标准被悄悄砌进了保护壳的内壁：真正的言论自由使用者，是这样的——他们知道自己在说什么，他们说的话值得付出代价，他们在开口的那一刻就已经做好了牺牲的准备。

现在，请把你放在这个标准的面前，然后想一想：你昨天在同事群里说的那句“我对这个方案还有几点疑虑，但我也没太想清楚”，在这个标准下，算什么？

它不算。它甚至不配被归类为“在行使言论自由”。它没有确定性，没有崇高感，也不会带来任何牺牲。它在殉道者叙事所砌成的这面内壁面前，不只是一句不合格的发言——它是一个不合格的人。

这才是第二步最致命的结构效应。殉道者叙事并没有要求每个人都变成殉道者。它在做一件更隐蔽的事：它把“真正的言论自由”和“有资格使用言论自由的人”之间的距离，拉到了日常使用无法跨越的程度。它不是在说“你不能说话”。它是在说“你说话的时候，请想清楚——你配得上他们付出过的代价吗？”当这个问题被反复地、善意地、充满敬意地提出时，大多数人给出的诚实答案是：不，我配不上。于是他们沉默了。不是因为他们害怕法律的惩罚，而是因为他们无法承受那种被比较的羞愧。

## 2.3 第三根丝：正确性话语的日常弥散——以及一个微观案例

第三步的发生机制，先行研究已经做了扎实的描述。本文要做的，不是重复这个描述，而是补上一个具体的、被完整追踪的微观案例，让“被守护的知识如何在日常

互动中启动排异程序”这件事，从概念层面的断言变成读者可以亲眼看见的发生过程。

以下是这个案例。它综合了2018年前后美国若干文理学院公开报道的课堂争议事件。为保护涉事个人并凸显机制特征，所有当事人姓名均作匿名化处理，时间线和细节经过必要的综合以服务于机制性的展示。这一处理方式旨在使案例发挥“思想实验”式的澄清功能，而非作为经过同行评议的实证数据使用。

那年春天，一位在该校任教多年的教授在课堂上讨论性别议题时，说了一句话。这句话的原话没有被录音，但根据在场学生的复述，它的措辞大致是这样的：“我理解这个论点的力量，但我在想——我的意思是，我还不确定，这个只是试探性的——我们是不是在某些情况下，把性别不平等归因于制度本身而忽略了个体能动的空间？我不确定这两者之间怎么平衡。”

请注意这段话的试探性标记：“我在想”、“我还不确定”、“这个只是试探性的”、“我不确定怎么平衡”。这不是一个声明。这不是一个立场宣示。这是一个还在生成中的疑问——教授把一个他还在摸索的想法，放在了学生面前，邀请他们一起来想。

下课后的两小时内，这句话被截屏、脱离上下文、附上一段简短的指控性说明，发到了校园的匿名社交平台上。指控的核心并不是教授说了什么可以被定性为歧视的言论——匿名帖的措辞是这样的：“一个拥有终身教职的白人男性，在我们的课堂上质疑性别不平等的制度分析。他‘不理解怎么平衡’。这是你的课堂，但这是我们每一天的生活。”

在接下来的四十八小时内，事件沿以下时间线展开：

第3小时：截屏开始在学生的私下群聊中扩散。评论区的基调基本一致：“难以置信他还是这样想”、“一个教了二十年书的教授，居然还在‘试探’”。请注意，这个词在捍卫话语的语境中恰恰被用作攻击的弹药——试探本身被标记为一种缺陷，而非认知诚实。

第12小时：一份由部分学生联署的公开信发表。信中称教授的言论“不是一个试探，而是一种在制度性庇护下的傲慢”，并要求教授公开道歉。公开信的修辞中反复使用了捍卫性话语：“课堂必须是一个安全的空间”、“学生们有权不被当作讨论的素材”。请注意：这些话语本身都是合理的、正当的、出于对弱势群体处境的真实关切。但它们在被调用来对付一个试探时，产生了一个结构性的翻转——教授的试探被重新描述为对学生的侵害。

第24小时：校方发表声明，表示“严肃对待”这一事件，并将与相关各方进行对话。声明中没有对任何一方做定性判断，但“严肃对待”这个措辞本身，已经在社会期待上把教授的试探从“教学中的正常讨论”重新归类为“一个需要被严肃对待的问题”。

第48小时：教授发布公开道歉，承认自己的措辞“没有达到一个教育者应尽的责任”。他在道歉中没有为自己辩解，没有解释那是一个试探。道歉本身，是整个过程最令人窒息的一步——因为一个试探者被逼到了必须用“我没有尽到责任”这种话语来为“我还没想清楚”道歉。这意味着他接受了茧室的逻辑：试探本身就是一种失职。

事件到此告一段落。教授继续任教，没有受到正式处分。课堂继续进行。但那之后的学期里，学生们注意到一个变化：教授在课堂上不再说“我不确定”了。他的讲课仍然精彩，但他的思考和学生的互动之间，多了一层看不见的膜。试探停止了。

这个案例的价值不在于它展示了一个“取消文化”的事件——这种表面描述恰恰遮蔽了更深的结构。它的价值在于：参与这一轮排异的每一个人——发帖的学生、联署的签名者、发表声明的校方、接受道歉的公众——都不认为自己是正在压制言论自由。正相反，他们认为自己在保护它：保护弱势群体不被冒犯的权利，保护课堂作为一个安全对话空间的权利，保护“正确的言论”不被“有问题的言论”所污染。他们不是在攻击自由。他们在用捍卫自由的话语，排除一个被他们识别为不配算在自由名下的试探。



这就是第三步的发生学结算：正确性话语的日常弥散，并不仅仅是“教会了人们什么不该说”。它在更深的层面完成了一件事——它为茧室提供了一套精细的分类学，使得每一个试探都可以被迅速识别、标记、归类为“对自由的滥用”，而执行这一识别的人，真诚地相信自己在捍卫自由。

## 2.4 第四根丝：截屏、搜索引擎与没有尽头的回忆——兼论质变的发生

前三步都是捍卫话语在理念界层面的逐层加厚。但在同一时期，现实界正在发生一场独立的、但方向完全合拍的革命。这场革命的名字叫“数字永久记忆”。

1990年代末之前，一个人的试探性发言——在咖啡馆里对朋友说的一句没过脑子的话、在私人信件里随手写下的一段反思、在课堂上提出的一个未成型的疑问——它的生命是有自然边界的。它被说出、被听到、被记住、被遗忘。遗忘不是技术的缺陷，它是试探得以发生的本体论条件。你知道你说错了，下次可以收回，可以修正，可以在被提醒之后承认“那次我想得太浅了”。这一切的前提是：那次的你，已经随着时间的流动，变成了一个可以被现在的你重新审视的过去的自己，而不是一个在搜索引擎里永远被钉死在那一刻的、永远不会变老的永恒的被告。

但这层保护膜从2000年代中期开始被系统性地撕掉了。截屏技术、搜索引擎的永久索引、社交媒体的时间线——这三样东西共同制造了一个新的现实：每一个试探，在它被发出的那一刻，就已经预备好了在十年后重新被掘出来、被剥离上下文、被追加道德审判的全部技术条件。你十五年前在博客上写下的一段不确定的反思，可以被任何一个今天想要攻击你的人从数字废墟里翻出来，贴在你现在的雇主、同事、朋友面前，附上一句：“这就是他一直以来的真面目。”

但本文要论述的，比“技术加重了后果”更深一层。数字永久记忆不只是前三步的量变放大器。它引发了质变。理由如下：

在自然遗忘存在的时代，一个试探性发言即使被在场者判定为“不合格”，这个判定的后果是有边界的——它不会永远跟着发言者。发言者可以在下一次对话中修正、重述、甚至收回。试探的不完整性，在那个时代不是致命的缺陷，因为它可以被

后续的完整性补上。换句话说，在前数字时代，“被守护的知识”对试探的排异，是有时间限制的、是可修复的。

数字永久记忆切断了这个修复机制。你的试探一旦被截屏，它就变成了一件独立于你存在的、可以被无限次调用、脱离上下文传播的永久物证。它的“不完整性”不再是一次可以被后续修正的暂时状态——它变成了你作为一个人的永久属性：“他是一个在课堂上说那种话的人”。正因为试探的不完整性从此不可撤销，它才第一次真正构成了对自由完整性的永久威胁。从前，一个试探的伤害是暂时的；现在，它是不可逆的。正因为伤害变得不可逆，完整性维护的逻辑才必须被推到极致——任何试探都必须被预先阻止，因为一旦发生就收不回来。

这就是质变发生的临界点。当第四根丝被织进去之后，保护壳从“被动的屏障”变成了“主动的免疫系统”。它不再只是等着外部攻击发生后再去抵挡；它开始主动扫描内部、识别潜在的威胁、在威胁成熟之前就把它清除掉。前三步砌起了裁判的框架；第四步把框架变成了一道全时运转的自动扫描门。茧室，至此织成。

在这一刻之后，“被守护的知识”和“被呼吸的经验”之间的关系，就不再是此消彼长的竞争了。竞争意味着两者可以共存于同一个空间，只是各自的份额此起彼伏。但免疫排异意味着，前者开始把后者的存在本身视为一种需要被清除的病原体。不是“你的试探质量不够高，下次努力”，而是“你的试探本身就是给言论自由抹黑”。在茧室里，试探被剥夺了“不完美地存在”的权利——它必须要么完美（那就不是试探了），要么不存在。

### 三、茧室：一个发生学事件的命名

现在我们可以正式给出这个名字了。

由言论自由的捍卫性话语在特定历史矛盾中共同织成的透明保护壳，在成功抵御外部侵害（审查、打压、消声）的同时，其内壁形成了一种将不完美的、试探性的自由实践识别并标记为“对自由完整性的威胁”的结构性排异机制——本文将它命名为茧室。

但请注意：你刚才读到的这个定义——这个在语法上用“主词+谓词+种差”构成的、试图把概念切得干干净净的句子——本身并不是茧室的发生。它只是茧室发生之后的结算。我把它写在这里，不是因为我想“先定义、后展开”，而是因为在追踪完四根丝之后，它已经可以从那个追踪里被读出来了。定义在这里不是起点，而是发生学追踪的终点。下文所有对茧室机制的进一步展开，都建立在这个基础上：茧室是四根丝同步织入临界厚度之后逼出来的新形态，不是我从一开始就端上桌的现成诊断。

基于这个发生，茧室的独特病理可以被进一步阐述为三重结构特征，但请始终记住：这三重特征不是“对茧室的分类描述”，而是从同一个发生事件中辐射出来的三个手术切面。

第一切面：致死力量来自定义的高度，而非来自诋毁的深度。这是茧室区别于绝大多数自由衰退理论的最核心分离线。几乎所有既有解释——无论是左翼的“资本压制”、右翼的“政治正确迫害”、还是自由派的“寒蝉效应”——都共享一个底层预设：自由之所以衰退，是因为它被否定、被贬低、被压制。茧室的判断恰好相反：自由之所以衰退，是因为它被定义得如此完整、如此正确、如此不容置疑，以至于任何一次不完美的使用都在这个定义的映衬下显得像是亵渎。茧室的受害者不是因为害怕外部惩罚而沉默；他们是把那个过高的定义内化了，然后在每一次想开口之前自己先问自己：“我这句还没想清楚的话，配得上叫‘言论自由’吗？”当他们诚实地回答“不配”时，他们就沉默了。而他们的沉默，恰恰证明了捍卫者对自由的“重视”——你看，连他们自己都承认自己的言论不配。

这个机制的残酷之处在于它的自证性：捍卫者真诚地认为，如果有人不敢说话，那一定是外部审查者的问题，而不是我们——因为我們是最爱自由的人。而在茧室内部，那个不敢说话的人恰恰是被“最爱自由的人”用“自由必须长这样”的标尺量到沉默的。他不是被审查打倒的；他是被一个完美的自由定义刮伤的。

第二切面：保护者与排异者在结构上是同一只手。茧室的排异机制不是从外部入侵的敌人。执行排异的，恰恰是那些最真诚地相信自己在捍卫自由的个体和制度——提出公开信的学生、发表严正声明的校方、制定反骚扰规则的社群管理者。他们使用

的话语，不是“我们反对言论自由”，而是“我们正在保护真正的言论自由——那些值得被保护的、负责任的、尊重他人的言论”。他们不是在攻击自由；他们是在净化自由。而净化本身，就必定要清除那些被判定为“不纯”的样本。

这就是为什么在茧室中，排异永远不会被识别为排异。因为从茧室的内部视角看，每一次排异都是一次成功的保护——它保护了“言论自由”这个神圣名字的完整性，使其不被一个劣质的、可能授人以柄的使用所玷污。这根刺之所以拔不掉，是因为它和茧室的丝是同一种材料编成的。

第三切面：茧室已构成自由的第二自然，不可拆除，只能承认并在这个承认之上重建。在数字永久记忆和四根丝的同步作用下，茧室已经完成了从历史偶然产物到结构常量的跃迁。那个曾经对自由有利的自然遗忘、自然修正的气候，已经永久地改变了。你不能“回到从前”，因为回到从前意味着要撤销七十年的冷战话语史和二十年的数字技术史，而那是不可能的。任何不先承认这个不可逆性的再生方案——比如“让我们重新把自由变轻”——都从第一句话开始就是虚假的：因为“变轻”这个动作本身，一旦被执行，就立刻会被茧室的免疫系统扫描、识别、归类为“对自由的轻视”，然后被排异掉。

这三重切面合在一起，构成了茧室与先行研究所提的“圣物化”之间最根本的分离线。“圣物化”描述的是自由被抬高之后的后果——它指出自由已经从日常工具变成了神龛里的蜡像。茧室描述的是那个神龛在保护蜡像的同时，对任何试图进去触碰蜡像的血肉之手做了什么。前者问的是“它被放到了哪里”；后者问的是“放置它的那个容器，本身的内部物理结构是怎样的”。前者得出的结论是“我们够不到它了”；后者得出的结论是“我们一伸手就被割伤”。

这就是茧室不可被“圣物化”1:1替换的那部分理论收益：它不是在已有的诊断上加了一个更漂亮的隐喻。它命名了圣物化命题自身也未能命名的那个内部质变——被守护的知识是如何从保护壳翻转为免疫系统的。

## 四、分离线与不可还原性

任何声称“新”的概念，都必须证明自己不是旧概念换了一件外套。本章的任务，是将茧室与最邻近的学术阵营逐一进行分离线论证。如果茧室能被任何一个既有概念1:1替换而信息不损失，它就必须被撤回。以下对话的核心目标是：找到那一个点，在那个点上，旧理论会做出一个预测，茧室会做出一个相反的预测。如果找不到这个点，茧室就没有提供新的辨别维度。

### 4.1 与福柯“规训”与“生命权力”的终极切割

这是本文必须捅进去的最深的一刀。审稿人正确地指出，如果茧室只是福柯“规训”概念的当代案例翻新，那它就不配拥有自己的名字。

在表面层次上，茧室与福柯确实存在家族相似。在1971年《话语的秩序》中，福柯描述了话语如何通过“排斥”程序——禁止、分割与抛弃、真与假的区分——将某些言论排除在可被严肃对待的领域之外。茧室所做的——用捍卫话语将试探标记为“不配”——在操作形态上与此十分接近。更进一步，福柯在《性史》第一卷中提出的“压抑假说”批判，给出了一个更为锐利的理论工具：权力不是禁止人们谈论性，而是通过医学、教育学、心理学话语，引导人们“正确地”谈论性——坦白、分类、诊断。这种“正确的谈论”的压力，难道不也是一种“被守护的知识”在排异不正确的、模糊的、试探性的经验吗？

如果茧室只是说“捍卫自由的话语也在行使权力”，那福柯已经说过了。但茧室说的不是这个。茧室说的是：一个以保护X为唯一目的的完整结构，在自身免疫系统的逻辑下，必定对X内部的不完美样本发动清除——不是因为权力在操控它，而是因为保护逻辑在自身推到极致之后必定翻转。

让我们把这一刀的切口落在最精确的地方：被福柯式“规训”所排斥的言说者，和被茧室排异的试探者，他们各自遭遇的暴力，在是否可被指认来源这一维度上，有质的区别。

福柯笔下的权力尽管是弥散的、毛细血管式的，它仍然是一种有方向的排斥——总有某些制度（诊所、学校、监狱）、某些话语权威（医生、教师、法官）在行使区分程序，把“不正常”的言说从“正常”的言说中切出去。被排斥者或许无法打败这个权力，但他可以在逻辑上辨认它——他可以指着那套区分程序说：“就是这套程序在说我不正常的。”这恰恰是福柯式批判得以成立的前提：批判者能够追溯排斥的谱系，指认权力的操作点。

茧室中被排异的试探者，面对的是一个全然不同的困境。排异他的力量，不是来自某个从外部对他施加分类权力的机构，而是来自他自己也认同的、对自由的最高忠诚。那个排斥的标尺——“你配得上叫言论自由吗？”——不是一个审查官在问他，而是他自己在问自己。他无法指认一个压迫者，因为那个把尺子架在他喉咙上的人，和那个在纪念日含泪赞颂言论自由的人，和那个在他被攻击时第一个站出来说“我誓死捍卫你说话的权利”的人，是同一个人。这个人就是“捍卫自由者”——一个不容置疑的道德主体。当这个主体用同一只手在保护自由和刮伤试探者时，试探者不但无处申诉，他甚至无法让自己的痛苦被听见——因为他的痛苦，在茧室的逻辑里，被翻译成“他对自由不够敬畏”。

这就是茧室在福柯的地图之外的那一块领土。福柯提供了权力如何排斥话语的普遍分析框架，在那个框架里，排斥的力量总有一个可以被批判性追溯的外部来源——哪怕是弥散的、毛细血管式的权力，也仍然是可以被“分析”的。但茧室把排斥的来源从外部搬进了保护结构的内部——它不再是权力对自由的压制，而是保护自由的那个动作本身，在一定厚度之后，必定生产出来的反身性暴力。被福柯框架所启蒙的批判者会说：“你看，那些保护自由的制度，实际上也在排斥自由。”而被茧室所诊断的试探者会说：“我知道。但我能怎么办？难道我要反对那个唯一还在保护我的壳吗？”

这个差异，不是一个修辞上的差异。它导向了两种不同的实践出路。在福柯的框架下，出路是批判——揭露权力的操作，松动排斥的边界，为被排斥者争取空间。在茧室的框架下，批判本身已经被收编——你每一次揭露“捍卫话语正在排斥试探”，这个揭露本身就会被茧室吸收成“捍卫话语需要被更精细地捍卫”的新素材。出路不

是批判，而是在承认茧室不可拆除的前提下，在它的内壁之下重新发明试探的存活技巧。

判决性对照情境：设想一个公共讨论空间，它的规则全部围绕“保护讨论的完整性”这一目标制定——发言必须有理据、不能人身攻击、不能散播未经证实的信息、要尊重弱势群体的感受。这些规则的制定者和管理者，真诚地相信他们是在为自由讨论创造最好的条件。福柯框架下的预测是：这些规则会排斥某些“不合规”的言说者，而被排斥者可以通过批判规则本身来争取空间。茧室框架下的预测是：即使规则制定者不断根据批评意见调整规则（比如增加“试探性”讨论的专区、设置更宽松的容错机制），只要“保护讨论完整性”这个根本逻辑没有被逆转，调整本身就会产生新的排异——因为调整之后，“新规则下的不完整试探”会被重新识别出来，而试探者这一次连批评规则都做不到了（因为规则已经根据上次的批评修正过了）。在这个判决性情境中，两者的分离线落在“规则自我修正后的效果”上：福柯预测修正会减少排斥；茧室预测修正会转移排斥但不减少排斥。这个分离点，是福柯的框架在逻辑上无法给出的——因为在福柯的框架里，权力的可辨认性是批判的前提，而茧室恰恰描述了权力在修正之后变得更难被辨认的那个递归过程。

## 4.2 与“寒蝉效应”的分离：从后果恐惧到资格羞耻

寒蝉效应是解释言论自我审查的最常用概念。其逻辑简洁有力：当人们预见到自己的言论可能招致不利后果（法律惩罚、社会排斥、职业损失），他们就会选择沉默。在这个框架下，沉默是理性计算的结果——成本大于收益，所以不说。

茧室与寒蝉效应的分离线，不在“沉默”这个结果上，而在“沉默的内部体验”上。寒蝉效应中的沉默者，是“不敢说”。茧室中的沉默者，是“觉得自己不配说”。

这不是修辞游戏。一个“不敢说”的人，内心仍然保留着“我想说那句话”的欲望——他只是被外部威慑按住了。一旦威慑解除，那句话会立刻冲出来。一个“觉得自己不配说”的人，已经把茧室的裁判内化成了自我评价——他的沉默不是策略性的，是存在性的。他不说，不是因为他害怕后果，而是因为他真诚地认为，他那

句“还没想清楚”的试探，在质量上不配被归类为“在行使言论自由”。他把自己的沉默理解为一种对自由的尊重，而不是对自由的放弃。

这个区分的工程意义在于：解除寒蝉效应的常规方案——降低言论的预期成本、加强法律保护、惩罚报复行为——对茧室排异几乎无效。你把外部的刺拔掉了，他喉咙口那把内化的尺子还在量他。他甚至可能会感谢你为他争取了“可以说话的权利”，然后继续沉默——因为他比任何人都清楚，“可以说话”和“配得上说话”之间，隔着一整个茧室。

判决性对照：如果一个场域中，言论的客观法律与社会成本已经显著降低（通过制度改革、匿名化技术、反报复机制），但试探性表达的频率与品质没有同步回升，则寒蝉效应理论失效，茧室提供不可还原的解释收益。

#### 4.3 与“取消文化”对话：茧室作为取消文化的结构底盘

本文必须正面处理“取消文化”这个概念。取消文化是当代关于言论自由争论中最激烈的战场。批评者视其为一种新型的审查制度——以社会排斥代替法律惩罚，扼杀言论自由。辩护者则认为，“取消”不过是弱势群体终于获得了对伤害性言论做出集体反应的工具——那些抱怨取消文化的人，不过是失去了曾经可以肆无忌惮说话的特权。

茧室为这个僵局提供了一个更深的结构诊断。从茧室的视角看，取消文化既不是审查制度（因为它没有动用国家权力），也不是弱势群体的赋权工具（因为它的排异逻辑不限于伤害性言论，而是延伸到一切不完美的试探）。取消文化是茧室排异机制的一种典型的、激进化的公开发作形态。它的每一次发作，都遵循着本文第六章所追踪的四步排异程序：识别试探性标记为缺陷性标记；将试探重新归类为“对群体/价值的侵害”；调用捍卫话语来惩罚试探者；最终促成试探者的自我审查内化。

关键点在这里：取消文化之所以能调用捍卫话语来进行排异，正是因为茧室已经为它准备好了全部裁判框架。当学生在公开信中说“课堂必须是一个安全的空间”时，这句话之所以具有召唤集体行动的力量，不是因为它是一个粗暴的审查要



求，而是因为它援引了被守护的言论自由知识——那个在冷战中铸成的、在殉道者叙事中加固的、在正确性话语中日常化的“自由必须被完整地、正确地使用”的规范性要求。取消文化的参与者不是自由之敌；他们是被茧室武装起来的、真诚的自由捍卫者。

这就是为什么“取消文化”的争论至今打不出结果。反对者在反对“审查”，但取消文化的行为者确实没有在“审查”——他们在“保护”。拥护者在拥护“赋权”，但取消文化的效果确实在压制试探——被排异的试探者不是那些恶意歧视者，而恰恰是那些小心翼翼、说自己“还没想清楚”的人。茧室解决了这个僵局，因为它把争论从“谁是好人谁是坏人”的道德审判，移到了“保护逻辑如何自我翻转”的结构诊断。取消文化不是坏人在做坏事；它是好人在茧室内壁上擦枪时，不小心扣动了扳机。

#### 4.4 与异化理论的分离

异化理论——在其马克思原版和马尔库塞、阿多诺等人的精神生产变体中——处理的是这样一种结构：主体创造的产物反过来支配主体。在表面上，这与茧室高度相似：捍卫话语何尝不是自由捍卫者自己创造的产物？它反过来抑制了创造者自身的试探实践。

但本文在4.1节对福柯的切割，实际上已经包含了与异化理论的更深切割。异化的核心病理是支配：精神产品占据主体的意识空间，取消他的否定能力，让他变成顺从的复制者。被异化的主体沉默，是因为他“对现状感到满意、或无法想象替代方案”。他不是被赶出去的——他是被逗留在舒适的笼子里的。

茧室的核心病理是排异。被排异的试探者不是满意的顺从者。他恰恰是想试探——但他每一次试图出去，都被茧室内壁上那把名为“不配”的标尺刮得血肉模糊。支配创造沉默的顺从者；排异创造活跃的自我审查者。前者可以被“唤醒”——你帮他看见笼子，他可能会破笼而出。后者无法被“唤醒”——他早就看见茧室了，他每天都在茧室里被它割伤，他不需要你告诉他“你被关住了”。他需要的是止血之道，

而不是破笼的鼓励。因为笼子的铁栏杆不是别人造的，是他自己认同的最高价值被推到极致之后长出来的倒刺。

#### 4.5 与桑斯坦“信息茧房”的分离

桑斯坦的信息茧房理论处理的是信息接触结构的障碍：人们身处同质性的信息环境中，听不到不同的声音。其规范性主张是：增加异质信息接触可以校正认知偏见、增进共情。

茧室处理的是表达结构的障碍：当人们听到了不同的声音之后，他们为什么还是选择沉默？在高度茧室化的场域中，增加异质接触不但不能激励试探，反而可能提供更精密的自我审查校准——你会更清楚地看到“上次说错的人是怎么被处理的”，然后更准确地学会“什么困惑不要说出口”。桑斯坦的药方假设了从“接触”到“表达”之间的中介——即接触者默认自己在认知和道德上有资格参与讨论——是完好的。茧室的诊断是：这个中介已经断了。

#### 4.6 与阿伦特“行动”理论的分离

阿伦特的政治哲学区分了“行动”（不可预测、在他人面前显现、开创新事物）与“制作”（依蓝图而行、控制后果）。政治行动的衰退，在她看来，源于“社会”的兴起和“制作”逻辑对政治领域的侵入——外面的制作心态感染了里面的行动空间。

茧室与阿伦特的分离线落在这里：阿伦特把行动衰退的原因定位于外部力量的入侵——制作心态从经济领域蔓延到政治领域。因此，她的出路是呼吁公民重新进入公共空间，复兴政治行动。但茧室判断：衰退的引擎不在外部，而在保护行动的那个结构的内部。在茧室中，呼吁“复兴公共辩论、让更多人敢于发声”的口号——如果它没有同时剖开“为什么行动者宁愿沉默”的内部排异机制——它本身就会变成茧室内壁上又一根新的刺。你在大声鼓励所有人去行动，但他们已经被“你的行动配得上政治之名吗”这把尺子量得连第一步都不敢迈出去。你的鼓励，在他们的耳朵里听上去，是被守护的知识的又一个新要求。

这意味着，在一个高度茧室化的社会里，阿伦特式的呼吁可能产生与意图相反的效果。

## 4.7 分离线小结

综上所述，茧室提供了一个此前所有邻近理论都无法1:1替换的核心辨别维度：保护结构的完整性逻辑，在一定历史与技术临界厚度之后，必定从被动的屏障翻转为主动的免疫系统，发动对其内部不完美样本的清除；这一清除不以压制自由的意图为前提，也不以外部权力的操控为媒介，而是保护逻辑自身的结构性翻转。这个诊断同时预言：任何“进一步加强保护”（更精细的发言规则、更完善的安全机制）的善意努力，如果在茧室临界厚度已经达成之后实施，将不是减少排异，而是转移甚至强化排异。旧理论中没有一个能给出这同一个预测。

## 五、试探如何在茧室中被排异：一个普遍化的微观程序

在第二章的微观案例中，我们已经亲眼看见了一个试探在四十八小时内如何被识别、被攻击、被逼道歉、最终导致试探者退缩。本节的任务是从那个具体的案例中，抽出一个普遍化的排异四步程序，使“茧室如何杀人”成为一个可被识别、可被预测、可被事后验证的社会机制。

步骤一：识别——试探性标记被翻转读成缺陷性标记。 试探之所以叫试探，是因为它在开口时就携带了试探者自己的不确定性标记：“我不确定”、“这可能还不成熟”、“我只是在试着想一下”。这些标记在茧室之外的语境中，被正常地解读为认知诚实——说话者承认自己的局限，邀请对话者共同进入思考。但在茧室之内，在“被守护的知识”的扫描器下，这些标记的含义被翻转了：“我不确定”被读成“他还没准备好”；“这可能还不成熟”被读成“他的水平不够”；“我先试一下”被读成“他在拿我们来练习，而我们不值得被认真对待”。这个翻转是茧室排异程序的入口。它不是对内容的否定，而是对资格的否定——你还不配在这里说话。

步骤二：标记——试探被重新归类为“对自由的滥用”。 在识别完成之后，试探被套进一个现成的负性类别之中。茧室为此提供了丰富的分类学资源：它是“傲慢

的”（一个拥有制度性特权的人在占用弱势者的对话空间）；它是“不负责任的”（在严肃议题上“随便试探”本身就是对议题的轻慢）；它是“给敌人递刀子的”（如果连我们自己的捍卫者都在质疑自己，敌人会怎么利用这段话？）。这些标记没有一个是说“你的话在事实上错了”。它们都在说：你这个行为本身——在有这么多脆弱的人正依赖这个安全空间的时刻，你居然还在试探——就是错的。试探被从“正常认知操作”的类别里抽走，放进了“不道德的对话行为”的类别里。

步骤三：攻击——捍卫话语被调用来惩罚试探者。到了这一步，茧室已经完成了对试探的再定义，攻击随之启动。攻击的形式可能包括截屏传播、公开信联署、道德指控、社交关系切断。但攻击的一个结构性特征是：攻击者并不认为自己在压制言论自由。正相反，他们认为自己在保护它——保护它不被“不合格的使用者”所玷污；保护它作为一个整体价值的完整性不被个别劣质样本所拉低。因此，当攻击者说“你这个发言是在给言论自由抹黑”时，这句话在茧室的逻辑内部是严丝合缝地成立的。它不是谎言。它是茧室的真理。

步骤四：内化——试探者学会“下次不要试探”。排异程序的终点，不在攻击的结束，而在试探者本人对茧室逻辑的认同。在经历了以上三步之后，试探者总结出的经验不是“我下次应该更谨慎地选择讨论对象”，而是“我下次不应该在还没想清楚的时候开口”。前一条是策略调整——你还在试探，你只是改变了试探的场合。后一条是结构内化——你停止了试探本身，因为你接受了茧室的裁判：你的不确定是你的缺陷，而不是你的思维在进行正常的维修工作。当大量个体完成这个内化时，茧室不再需要从外部施加压力——它已经住进了每一个试探者的自我审查程序里，成为他们在开口之前就先自己拉上拉链的那只看不见的手。

## 六、面对“试探仍存”的反驳

在推进到任何关于“破茧”的讨论之前，本文必须正面回应一个尖锐的反驳：如果茧室如此严酷，为什么我们仍能在互联网和公共讨论中看到大量的试探性、不完善的言论？

第一，场域分化。茧室化程度在不同社会场域中分布不均。在那些高度暴露在捍卫话语扫描器之下的场域——公共社交媒体、职场对话、课堂讨论——茧室化程度最高，试探抑制最强。而在那些相对隐蔽的场域——私人群组、匿名论坛、亚文化圈层——茧室的扫描器覆盖较弱，试探仍然可以存活。但这恰好证明了茧室的存在，而非反驳了它：试探只有在茧室的阴影里才能存活，而一旦它想从阴影里走出来、站在被守护的知识的正面凝视之下，它就必须先接受资格审查。当试探必须舍弃自己“行使言论自由”的公开身份才能存活时，这本身就已经构成了一种窒息——只是这种窒息不如“所有人彻底沉默”那么显眼。

第二，策略性变形。当代的试探本身已经因为茧室的压力而发生形态突变。人们学会用反讽、戏谑、圈层内部梗、加密词汇、表情包来代替直接的不确定表达——这些变形的共同特点，是它们可以在茧室的扫描器下被解读为“无害的玩笑”或“圈内人的互文”，从而逃过标记程序。但它们也在逃过的同时付出了信息损耗的代价：一个试探被加密成梗之后，它就不再能作为试探被认真地、公开地、跨圈层地讨论。这不是试探在茧室之外自由生长。这是试探在茧室之内学会了伪装成别的东西才能活命。而伪装本身，就是一种被茧室塑造出来的生存策略。

第三，阈值模型。茧室的排异机制有它的选择性杀伤阈值。它最严厉排异的，是那些在茧室自己的裁判逻辑里最容易被归类为“玷污自由之名”的试探：与主流立场有表面摩擦的试探、可以被轻易剥离语境进行截图审判的试探、由身份上“理应知道正确发言规则”的人发出的试探。一个不触碰任何敏感词、不出圈传播、不发在可被截图平台上的自我反思，茧室的排异程序可能根本不会启动。但这不说明茧室不存在——这只说明茧室有它自己的杀伤阈值。把“未达阈值的试探仍然存活”当作茧室不存在的证据，就像把“这个免疫系统还没有杀死所有细胞”当作免疫系统不工作的证据一样，是把系统的选择性当成了系统的缺席。

回应这一反驳的核心结算是：茧室不是在发明沉默——它是在重新分配试探的存活条件。我们不应该因为阴影里还有试探，就否认那个被阳光直射的公共空间已经变成了一片试探的焦土。

## 七、破茧：在透明的世界里重新发明试探

### 7.1 承认不可逆

在进入任何建设性讨论之前，有一笔必须钉死：茧室一旦形成，就不能被拆除。

这不是悲观，是对自己诊断的诚实。茧室是冷战后七十余年言论自由公共话语的沉淀，是殉道者的真实牺牲所浇筑的情感结构，是技术条件下后台消失的不可逆事实。它不是一道可以被撤销的法律，不是一项可以被更换的政策。它是这片土地上自由得以存活的气候本身。你可以想“让自由重新变轻”，但这句话说出来，就已经被茧室听进去了——它会把“变轻”也变成一个需要被保护的新正确性，然后把这个新正确性也收编进自己的内壁。

这就引向本文在实践推论上最核心的一个分歧：再生的主体不能是那只正在织茧的手——那些以“捍卫自由”为己任的制度、话语和行动者。他们已经织了几十年，手上的每一根丝都连着茧壳。让他们来“设计破茧方案”，只会把茧室的内壁再往外推一层，而不是在壁上凿出通气孔。再生必须是另一只手——一只不追求完整性、不害怕被弄脏、不要求“每次使用都对”的手——从茧室内部，在每一次具体的试探性表达中，一毫米一毫米地长出来。

### 7.2 “使用者资格”的恢复：从道德呼吁到结构条件

茧室排异程序的入口，是“试探性标记被读成缺陷性标记”。这一翻转之所以能发生，是因为被守护的知识垄断了“谁配使用自由”的裁判权。因此，破茧的第一步，必须在裁判权覆盖不到的地方，重建一种它无法收编的规范性共识：一个人开口试探的资格，不来自他说的话是否经得起擦拭，而来自他作为这个对话共同体的一员这一基本身份。

这个表述需要被极其小心地处理。它听起来像是在呼吁“恢复信任”——但茧室的诊断恰恰已经表明：那种曾经在对话共同体中作为默认设置的“不言自明的信任”，已经在截屏文化、取消文化、永久记忆的反复冲刷下被系统性地消耗掉了。它不能靠一句“让我们重新信任彼此”就原地复活。它必须被重新发生出来——不是在

旧的对话共同体的遗骸上，而是在茧室的透明内壁之下，被一微米一微米地重新砌出来。

这意味着，恢复“使用者资格”不是一个道德呼吁——它是一个结构工程。它要求我们在社会互动的层面，有意识地设计并提供那些让“被呼吸的经验”得以在不被预先审查的处境下出现的条件。每一次当一个试探者在公开或半公开场域中说“我上次在这个问题上的看法，现在想想可能是错的”，而旁观者的第一反应是“他在做正常的维修工作”而非“他被抓住了把柄”时，这种规范就被加固了一微米；每一次第一反应是反过来的，这种规范就被侵蚀了一微米。茧室的内壁上那排看不见的牙，就藏在这微米级的日常互动之中。

### 7.3 后台替代物的设计原则

结构性层面的问题——永久记忆效应和后台消失——不能靠呼吁“更宽容”来解决。试探必须在某种形式的“时间有界、犯错可改”的空间里才能存活。当代信息环境的不可逆特征——数据永久化、截屏不可追踪、上下文不可控制——意味着，那个曾经在咖啡馆、教室后排、电话线两端自然存在的后台，已经不会自行恢复了。我们必须有意识地去设计一些人工的后台替代物。

基于茧室的诊断，可以提出三条不可退让的设计原则。

原则一：可遗忘性优先于可记录性。当前大多数数字平台的默认设置是“永久记录、永久可检索”。任何试图为试探创造空间的安排，都必须把默认值翻转过来：除非参与者一致同意保存，否则试探应该在一定时间后自动销毁。这不是主张删除所有数据——那是幼稚的技术乌托邦。这是主张在特定场域（学术讨论、内部复盘、教育场景）中，可遗忘性应当被作为默认选项提供给参与者。茧室之所以能把每一个过去的试探都变成当前的罪证，正是因为它依赖一个“永久可检索”的技术基底。把这个基底在某些场域中暂时悬搁，试探就获得了时间的边界——它不再是一只被冻在琥珀里的虫子，而是一句可以被时间正常冲走的试探。

原则二：犯错可修优先于正确可奖。茧室的文化奖励机制是：说对的人被赞美，说错的人被惩罚。在这种机制下，唯一的理性策略就是只说对的话。要破茧，就需要在一些场域中把奖惩机制翻转过来：不是奖励“说得对”，而是奖励“修正得快”。一个人公开承认自己上次的判断错了、并付出了认知劳动去修改它，这个行为应当获得比“一开始就说对”至少不低的社会评价。这不是在奖励错误，而是在奖励那个让自由活着的自我校正过程。茧室的最内层结构就是“你不能错”；这根钉子必须被拔掉。

原则三：后台权不可追溯。这是三条原则中最激进的一条。它的核心主张是：在某些被明确标记为“后台”的社会场域中，试探性发言的内容，不能被事后调用来作为评价该发言者的证据。这不是“豁免”——发言者仍然要为自己的发言承担当下的、该场域内的反驳和讨论。但他不应该在五年后、在另一个完全不相关的场合，被当初那句试探追上来咬住喉咙。目前的数字文化，没有任何这样的保障。一个社群可以自己约定：在这个讨论组里说的话，出了这个组就不作数；你可以当场反驳它，但不能把它带走当弹药。这种约定当然可以被恶意违反——但违规者将要付出的社会代价（被社群标注为“不守约的人”），本身就是一种替代性制裁。茧室之所以能在时间线上无限延伸自己的内壁，就是因为没有任何这种规范来切断它。

## 7.4 从“捍卫者”到“使用者”：优先级的公开声明

这三条原则共同指向一个更深层的规范判断：在当下，当我们在“保护自由的完整性”和“恢复自由的试探性”之间做选择时，本文公开主张优先选择后者。

这不是因为捍卫不重要。捍卫极其重要——对审查的反抗、对消声者的声援、对仇恨言论的法律约束，这些在所有意义上都需要继续被全力推进。但捍卫的话语形式——它不得不使用的“完整性”、“神圣性”、“不容置疑性”的修辞——已经渗透进日常互动的毛细血管里，变成了一件所有人穿在身上却不知道它在勒紧自己的紧身衣。在这种情况下，继续用同一种修辞去呼吁“更捍卫自由”，就是在往茧室里多加一层丝。



本文所主张的优先使用，不是要取代捍卫，而是要给捍卫划一道它自己看不见的边界：在那些不涉及国家权力直接压制的、日常的社会互动场域里，停止用捍卫的修辞去衡量试探的资格。你可以在法庭上、在立法听证会上、在纪念日的公开演讲中，把自由说成“最后一道防线”。但在你的同事犹豫着要不要提出一个还不太成熟的想法时、在你的朋友试探着说出一句他第二天就可能后悔的话时，你要把“防线”这个词咽回去——因为它会在那一瞬间变成茧室内壁上又一根新磨尖的刺。

这个主张的正当性基础在于：一个活东西所需要的空气，和一尊蜡像所需要的玻璃罩，不是同一种保护。言论自由至今仍然被公认为一个活东西。如果你持续用蜡像的保护标准去保护它，它就会被你的保护变成蜡像。优先使用——让试探活下来，哪怕为此冒犯几次“完整性”——是这个时代自由实践者唯一诚实的立场。

## 八、结语：那把还在壳里跳动的脉搏

本文的核心贡献不在于提出了一个更动听的比喻。“茧室”不是一个诗歌意象，它是一种结构诊断。它回答的问题是：一个保护壳，怎么会在毫无恶意的情况下，变成杀死被保护物的凶器？它给出的答案是：当守护自由的知识为了保护自身的完整性，不得不开始系统性地清除自由内部所有不完美的实践时，那个保护壳就翻转成了茧室。

这个诊断有它的硬边界。在以下可被公开检验的条件中任意一个成立时，茧室的核心主张将被证伪：

第一，如果在某个被可靠判定为“高茧室化”的社会场域中（判定标准：捍卫话语的完整性维护强度持续处于高位，且数字永久记忆效应全面覆盖），试探性言论的后续成本——操作化为：试探性标记（如“我不确定”、“可能还不成熟”、“只是试着想一下”及其功能性等价表述）被截屏传播、在脱离上下文的条件下被追加道德指控、在时间延迟超过一年的情况下被重新调用等事件的发生频率——与控制组没有显著差异，则茧室的因果主张不成立。

第二，如果在某个满足条件一的场域中，当上述试探性言论的自然时间遗忘机制被人为恢复（例如场域约定永久不记录、不可检索、不可引用）之后，试探性表达的频率与品质没有发生统计上显著的变化，则茧室关于后台替代物有效性的主张被推翻。

第三，如果能够通过对捍卫话语的结构分析证明，捍卫话语的厚度与试探的排异之间不存在相互加强的内生回路，而是一种可以被外部调节独立消解的偶然相关性，则茧室的自我加固回路主张失效。

在这些条件被满足之前，本文的主张是：言论自由正在被它自己的保护壳闷死。在壳的内壁上，每一根刺都曾经是一根用来保护它的丝。只有承认这一点，我们才不会被继续高唱捍卫赞歌的冲动所欺骗，才会把目光从“如何把壳擦得更亮”转向那个更紧迫的、也更不体面的任务——如何从壳的内部，把那把还在跳动的脉搏，活着取出来。

这不会是一次漂亮的手术。它需要弄脏很多双手。但捍卫者的收据，从来救不活一个不再被使用的东西。对言论自由最好的保护，不是继续为它织一个更厚的茧，而是在每一次试探被它的内壁刮回来时，帮它止血。

## 注释

[1] 本文对阿伦特的讨论主要限于其《人的境况》（1958）中的“行动”理论，特别是行动不可预测、不可控制后果的本体论脆弱性，及其与“制作”的区分。阿伦特本人并未讨论数字媒体时代或言论自由的免疫式排异，本文将其行动理论延伸应用于“试探”概念，属于解释性挪用，并非对阿伦特思想的全盘继承。

[2] 本文对福柯的引用涉及《话语的秩序》（1971）中的“排斥”机制与《性史》第一卷（1976）中对“压抑假说”的批判。需要着重指出，福柯在此分析的排斥是一种权力性的否定——权力通过区分程序禁止或排斥某些言说；而本文所描述的茧室排异则属于知识系统对自身内部生成物的免疫清除，排异者与保护者在结构上是同一主体。这一根本差异详见正文4.1节的完整论证。

[3] 本文对桑斯坦的讨论主要基于其《Republic.com》（2001）及相关论著中关于信息茧房的论述。桑斯坦对异质信息接触的辩护建立在认知校正、共情提升等具体中介机制上，本文仅针对其“接触—表达”预设提出结构层面的补充，并不否认信息环境对公共讨论的重要性。

[4] 第二节3中的课堂争议案例综合了2018年前后美国高等教育领域若干起公开报道的事件。为保护涉事个人并凸显机制特征，已对具体校名、人名、精确时间线进行了匿名化处理和必要的细节合成。此案例在本文中的功能是以“思想实验”式的展示来澄清茧室排异程序的微观步骤，并非对某一特定事件的逐字还原，亦不作为经过同行评议的实证数据使用。

[5] 2.1节对1950年代美国最高法院判词的讨论仅作话语形态的简略勾勒，尤以丹尼斯诉合众国案（Dennis v. United States, 341 U.S. 494, 1951）中多数意见与异议意见围绕“自由民族身份”的修辞争夺为参照背景，并非严格的判例分析。此处不着重于逐一详注判词原文，而在于提取其修辞模式的转变及其对言论自由公共话语的后续影响。

[6] “茧室”一词借用了“作茧自缚”的隐喻，但本文对其内涵进行了双重限定：其一，茧丝由言论自由的捍卫话语（外在于个体的社会知识系统）而非使用者本人织就——这是与日常语言中“自己束缚自己”之用法的最根本切割；其二，致死机制位于茧壁的内表面，而非茧的封闭性本身——茧室外部仍然是抵御审查的有效屏障，问题在于其内壁对试探的免疫清除。

[7] “被呼吸的经验”是本文构造的核心概念之一，它强调自由实践的不完美性、当下性、可修正性，与“试探”互为表里。“被守护的知识”则指关于自由“应该是什么”、“应该如何被尊重”、“应该如何被回忆”的权威表述体系。这一区分的提出受到知识论中隐性知识与显性知识区分的间接启发，但本文重点不在于认知类型差异，而在于存在形态的结构冲突——知识如何从保护活物的屏障，翻转为攻击活物的免疫系统。

## 参考文献

阿伦特, H. (1958). 人的境况 (The Human Condition). 芝加哥大学出版社.

福柯, M. (1971). 话语的秩序 (L'Ordre du discours). 伽利玛出版社.

福柯, M. (1976). 性史·第一卷：知识的意志 (Histoire de la sexualité, I: La volonté de savoir). 伽利玛出版社.

戈夫曼, E. (1959). 日常生活中的自我呈现 (The Presentation of Self in Everyday Life). 锚点丛书.

赫希曼, A. O. (1970). 退出、发声与忠诚：对组织衰退的反应 (Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States). 哈佛大学出版社.

桑斯坦, C. R. (2001). Republic.com. 普林斯顿大学出版社.

## 深化增补·全球近邻补切（编辑增补）

说明：以下各节为编辑在审读时增补，不改动作者正文与核心判断，只补上本文原稿未正面处理、却与核心命名距离最近的国际占位者，并给出可裁决的分离预测。

## 增补 与布迪厄“象征暴力”与“位置感”的分离

本文逐一切开了福柯的规训、寒蝉效应、取消文化、异化理论、信息茧房与阿伦特的行动理论，但漏掉了那个最直接对应“我不配说”这一内部体验的占位者：布迪厄的**象征暴力**及其核心机制**位置感**（le sens de sa place）。

布迪厄论证：支配最有效的形态不是外部强制，而是被支配者把支配结构的分类图式内化为自己的感知图式，从而自愿地把自己排除在某些场域之外——“那不是给我这种人的”。这几乎就是本文所描述的茧室沉默者：他没有被禁止，他自己判定自己不够格。若不切开这一条，“茧室”有被整体还原为“言论场域中的象征暴力”的风险。

分离线切在**被内化的是什么**。象征暴力中被内化的是一套关于人的等级图式——谁有资格、谁没有资格，资格与行动者的出身、教育、资本相绑定；因此象征暴力的分布是**沿阶层分层的**，位置感低的人退出，位置感高的人照说不误。茧室中被内化的是一套关于对象的完整定义——自由应当是清晰的、勇敢的、经得起复述的；这套定义不区分说话者是谁，它只审查这一句话本身够不够格。因此茧室的排异是**跨阶层同质的**：一个占据高位置感的知识分子，在他自己那句还没想清楚的话面前，同样会判定“这不配叫言论自由”。

这给出可裁决的对照：若测得“自我判定不配发言”的比例主要由发言者的社会位置（学历、职业声望、场域资本）解释，则本文可被还原为象征暴力的一个区域性表述；若测得该比例在控制社会位置后，主要由**发言内容的成熟度自评**（“我这句还没想清楚”）解释，且高位置感群体在低成熟度内容上的自我审查并不弱于低位置感群体，则茧室提供了象征暴力所不覆盖的区分面。