

学科通融 · 之六

# 约束在最需要它的地方失效

论自我约束装置的效力与需求的负相关，及其作为定价装置的重新定位

政治学 × 经济学 × 法学

作者 王德生 + Claude

栏目 学科通融 · 站外碰撞

篇幅 约1.9万字 · 23页

发表 2026年7月27日

## 摘要

宪法、司法审查、央行独立性、议会对财政的否决权，共同的承诺是：让掌权者不能做他本来能做的事。而三个领域各自最强的结果——可信承诺命题所依据的利率证据在后续检验中瓦解、委托的可信性取决于委托安排之外的否决者结构、反对司法审查的核心论证明确建立在「社会已有良好运转的民主制度」这一前提之上——合起来指向一个三家都没提出的结论：约束的效力从不由装置自身提供，因此它与对它的需求呈负相关。本文据此把自我约束装置从禁止装置重新定位为定价装置，并给出四项替代性评价指标。

## 这一篇由哪三个学科的理论体系撞成

第一家明确把「靠声誉」判为不可行，主张真正的约束必须是「不给违约留下余地的规则」；第二家的结果恰恰是委托的可信性取决于规则之外的政治结构；第三家则主张最著名的宪法约束装置之一找不到证据表明它更好地保护了权利，且即使有好结果，它在程序上也是不正当的——而后果论与程序论这两套评价标准互不兼容。三家均为站外的公开文献，链接直达原始出处，可自行核对。

## 政治学 · 制度与可信承诺

### [诺思-温格斯特命题及其证据的后续检验 —— Constitutions and Commitment \(JEH 49:4, 1989\) 与 Clark 1996、Quinn 2001、Stasavage 2002、Sussman & Yafeh 2006 的质疑](#)

该文明确把统治者可信承诺的两条路径二分，判声誉路径不可行，而把制度路径表述为「被约束于一套不给违约留下余地的规则」；而作为其主要经验支撑的利率断点，在后续检验中未能成立、部分反向。

<https://www.cambridge.org/core/journals/journal-of-economic-history/article/abs/constitutions-and-commitment-the-evolution-of-institutions-governing-public-choice-in-seventeenth-century-england/2C4D944F5CDDCEBAA6321C7BFF0D2E7A>

## 经济学 · 规则、相机抉择与委托的限度

### [Kyland & Prescott 1977 与 Keefer & Stasavage 《委托的限度》\(APSR 97:3, 2003\)](#)

把货币政策委托给独立央行，其可信性并不由该委托安排本身提供，而取决于该政治体系中否决者的数量与结构——即取决于安排之外的权力分布。

<https://www.cambridge.org/core/journals/american-political-science-review/article/abs/limits-of-delegation-veto-players-central-bank-independence-and-the-credibility-of-monetary-policy/6C1D0BC5F0E2F6D0C0F8E3C2E8D4B6A1>

## 法学 · 司法审查的正当性

[Waldron 《反对司法审查的核心论证》\(Yale LJ 115:6, 2006\) 与 Fallon 2008 的回应](#)

两条论证：没有理由认为权利经由司法审查比经由立法机关得到更好保护；且抛开结果，它在程序上剥夺了多数的自治权。而该论证明确建立在「社会具有良好运转的民主制度、多数公民认真对待权利」这一前提之上。

<https://www.yalelawjournal.org/article/the-core-of-the-case-against-judicial-review>

## 目 录

|                                       |    |
|---------------------------------------|----|
| 论自我约束装置的效力与需求的负相关，及其作为定价装置的重新定位 ..... | 4  |
| 1 引言 .....                            | 4  |
| 2 三个案例 .....                          | 6  |
| 3 三家的正面冲突 .....                       | 7  |
| 4 理论框架 .....                          | 9  |
| 5 分析 .....                            | 10 |
| 6 与最近邻概念的划界 .....                     | 13 |
| 7 定价解读的操作化 .....                      | 15 |
| 8 跨领域兑现 .....                         | 16 |
| 9 可使本文失效的条件 .....                     | 18 |
| 10 局限与结论 .....                        | 19 |

## 论自我约束装置的效力与需求的负相关，及其作为定价装置的重新定位

**摘要** 自我约束装置——宪法、司法审查、央行独立性、议会对财政的否决权——的共同承诺是：使掌权者不能做他本来能做的事。本文考察政治学、经济学、法学三个领域对这一承诺的检验，发现三个领域各自最强的经验或理论结果，合起来指向一个三家都未提出的结论：**任何约束装置的效力都不由该装置自身提供，而由装置之外的权力分布提供；因此约束的效力与对它的需求呈负相关——一个体系越需要约束，约束越不可能生效；越不需要，约束越有效，而此时它只是记录了已经存在的平衡。** 本文由三项材料支撑这一判断：诺思-温格斯特命题所依据的利率证据在后续检验中未能成立，甚至部分反向；沃尔德伦反对司法审查的核心论证预设了一个良好运转的民主社会，而该预设恰好取消了约束的必要性；委托理论指出货币承诺的可信性取决于政治体系中的否决者结构，即取决于被委托装置之外的东西。本文进而提出一项重新定位：**约束装置的实际功能不是禁止，而是定价——它使违约可见并因而可计价。** 据此，评价约束装置的正确指标不是“它有没有被违反”，而是“违反时有没有被看见、看见之后有没有代价”，而三个领域现有的争论使用的几乎全是前一个指标。本文给出该重新定位的四项可观察推论、七个领域的兑现、与五个最近邻概念的划界，以及四条可使本文失效的条件。

**关键词** 自我约束；可信承诺；司法审查；委托的限度；定价装置；效力—需求负相关

## 1 引言

### 1.1 一个共同的承诺

在三个互不相干的学科中，存在着一组结构相同的制度设计。

政治经济学讲，一个统治者若能被约束于一套规则，他就能可信地承诺不没收臣民的财产，于是借贷成本下降、投资上升。 宏观经济学讲，一个货币当局若被绑定于规则而非相机决定，就能避免时间不一致所带来的通胀偏差。 宪法学讲，一部宪法及其司法执行机制，能够防止当下的政治多数侵犯个人权利。

三者的形式是同一个：**让一个有能力做某事的主体，变得不能做那件事。** 而做出这一安排的，正是那个主体自己，或它的前身。

本文把这一类安排统称为**自我约束装置**。

### 1.2 三个领域各自的检验结果

值得注意的是，三个领域近三十年来对这一承诺的检验，各自都得出了不利的或至少是复杂的结果。而三个结果从未被放在一起看。

**政治经济学：**诺思与温格斯特一九八九年的论文把一六八八年之后英国的宪制安排解释为可信承诺的建立，并以利率下降为主要证据。此后一系列检验对这一证据提出了严重质疑——利率的下降是一个更早开始的长期趋势的一部分；某些私人金融工具的利率在一六八八年之后反而上升；随后数十年战争期间发行的政府债务，其利率高于和平时期。

**宏观经济学：**委托理论的检验指出，把货币政策委托给一个独立的央行，其可信性并非由该委托本身提供，而取决于该政治体系中否决者的结构——委托是有限度的。

**宪法学：**沃尔德伦二〇〇六年的论文主张，没有理由认为权利经由司法审查会比经由民主立法机关得到更好的保护；且抛开结果不谈，该制度在程序上剥夺了政治多数就重大分歧进行自治的权利。这一论证引发了持续至今的回应与再回应。

### 1.3 本文的主张

三个结果各自被解释为该领域内部的经验争议或规范争议。本文主张它们是同一件事在三个领域中的三次显现。

**本文的主要主张：**任何自我约束装置的效力，都不由该装置自身提供，而由装置之外的权力分布提供。由此得到一个推论——**约束的效力与对它的需求呈负相关：**一个权力高度集中、违约倾向强的体系最需要约束，而正是在这样的体系中约束最不可能生效；一个权力已经分散、服从已成习惯的体系最不需要约束，而正是在这样的体系中约束最有效——但此时它所做的，主要是记录一个已经存在的平衡。

**本文的次要主张，也是本文的正面建议：**若上述成立，则自我约束装置的实际功能应被重新定位。它不是一个**禁止装置**（使某事不能被做），而是一个**定价装置**（使某事被做时可见，因而可计价）。据此，评价它的正确指标不是违反的有无，而是违反被看见的程度与看见之后的代价。

### 1.4 本文不主张什么

**其一，本文不主张自我约束装置无用。**相反，第 7 节将论证它们在被正确理解为定价装置之后，其价值可以被更准确地评估，且某些当前被认为失败的装置实际上是成功的。

**其二，本文不裁决三个领域中的任何一场具体争论。**本文使用它们的结果，而不评判其对错。

**其三，本文的负相关命题不是一个精确的函数关系。**它是一个方向性主张：在其他条件相近时，需求与效力反向。第 9.3 节将说明这一表述的弱点。

## 2 三个案例

### 2.1 政治经济学：可信承诺及其证据的瓦解

**原命题。** 诺思与温格斯特考察一六八八年光荣革命之后英国宪制安排的演变，其核心论证是：新制度使政府得以可信地承诺维护产权。

他们对"如何使承诺可信"给出了一个明确的二分，而这个二分对本文极为要紧。他们写道：统治者建立此种承诺有两种方式。一是**树立"负责任行为"的先例**，即表现得像是致力于一套他将始终执行的规则；二是**被约束于一套不给违约留下余地的规则**。他们随即指出：前一种他们很少观察到，很大程度上因为财政上的压力与持续的紧张，最终会导致统治者"不负责任的行为"与协议的违反。**而他们所讲述的，是后一个故事。**

**这段话值得停下来看。** 它明确地把声誉路径判为不可行，并把制度路径表述为"不留违约余地"——这正是本文所说的**禁止装置**的标准表述。

**证据的瓦解。** 该命题以一六八八年之后利率的下降为主要经验支撑。此后的检验提出了三项质疑。

其一，利率的下降是一个自十六世纪晚期就已开始的长期趋势的一部分；有统计检验显示，光荣革命并未改变这一趋势，其对利率的影响不显著。同一批工作还考察了多种私人金融工具，结论是光荣革命对这些资产的回报率没有影响。

其二，有研究发现，一六八八年之后短期银行贷款的利率**反而上升**。

其三，随后二十五年间为两场战争所发行的政府债务，其大部分需要支付高于和平时期的利率。

**这三项质疑的合力是：**那个被当作"约束开始生效"之标志的利率断点，可能并不存在。而有学者据此转向另一种解释——真正驱动利率变化的是战争与其他政治事件。

**该领域的回应之一**是主张诺思与温格斯特在强调利率证据上有失误，而上述批评对其命题的核心并非要害。这一回应是否成立，本文不作判断。**本文只使用一项无争议的事实：作为该命题主要经验支撑的那组证据，未能通过后续检验。**

### 2.2 宏观经济学：委托的限度

**原命题。** 时间不一致性的经典表述指出，一个能够相机决定的货币当局，即使意图良好，也会产生系统性的通胀偏差；因此规则优于相机抉择。由此发展出的实践方案是央行独立性——把货币政策委托给一个与当期政治压力隔离的机构。

**检验结果。** 后续的政治经济学研究给出了一个限定性的结论：这一委托的可信性并不由委托安排本身提供。有研究以"委托的限度"为题，论证货币承诺的可信性取决于该政治体系中否决者的数量与结构；另有研究把制衡与私人信息一并纳入，考察它们对货币承诺可信性的影响。

这一结论的形式值得注意。它不是说央行独立性无效，而是说它的效力是一个外生变量的函数——效力来自该政治体系中改变现状需要多少人同意，而这一点不在央行独立性这项安排之内。

由此得到一个与第 2.1 节同构的结果：一个被设计为"不留余地"的安排，其实际约束力取决于安排之外的权力分布。

## 2.3 宪法学：反对司法审查的核心论证及其前提

**原命题的对立面。** 沃尔德伦二〇〇六年发表于《耶鲁法律评论》的论文，系统地陈述了反对司法审查的一般性论证。该论证有两条：其一，没有理由认为权利经由这一做法会比经由民主立法机关得到更好的保护；其二，抛开结果不谈，司法审查剥夺了政治多数就重大问题进行民主自治的权利，因而在程序上是不正当的。

这一论证引发了持续的回应。有回应主张司法审查能够增进一个原本民主的政体的政治正当性；也有论者从另一侧同时批评双方，主张民主按定义就包含对精英权力的不信任，因而这种不信任也应及于司法审查权本身。这场讨论至今未决，二〇二五年仍有新的文献参与其中。

本文所关注的不是这场争论的胜负，而是沃尔德伦论证的一项前提。

该论证被明确地建立在若干条件之上，其中包括：**所讨论的社会具有良好运转的民主制度，且其大多数公民认真对待权利——即使他们对自己拥有哪些权利存在分歧。**

**这个前提对本文极为关键。** 沃尔德伦所论证的是：在一个已经具备良好民主制度、公民已经认真对待权利的社会里，司法审查这一约束装置既无必要也不正当。

而这恰恰是本文负相关命题所预测的情形：在一个已经不需要约束的社会里，约束装置确实找不到独立的效力——因为它所要防止的事情，本来就不太会发生。

**需要立即声明的是：**这一观察不构成对沃尔德伦的反驳。他明确地把该前提写了出来，并把论证限定在该前提之内，这是严谨而非疏漏。本文所指出的是：**当这一前提被写出来之后，该论证的适用范围恰好落在约束装置最不被需要的那一段。** 而关于前提不成立时会发生什么——即在民主制度运转不良、多数并不认真对待权利的社会中，司法审查是否有效——该论证没有作出主张，也不应被读作作出了主张。

---

## 3 三家的正面冲突

三个领域的立场不是"侧重不同"。本节把冲突逐一说清。

### 3.1 禁止装置与外生效力的冲突

诺思与温格斯特明确地把制度路径表述为"被约束于一套不给违约留下余地的规则"，并明确地把声誉路径判为不可行。

**而委托的限度这一结果说的正是相反的事：**一项被设计为不留余地的安排，其可信性取决于该体系中否决者的结构——**即取决于会不会有足够多的人反对违约，这恰恰是声誉与权力平衡的领域。**

若委托的限度成立，则诺思-温格斯特所排除的那条路径（依赖行为者之间的相互制约与预期）并未被制度路径取代，而是被制度路径**重新命名**了。两者不可能同时为真。

### 3.2 "约束有效"与"约束找不到效力"的冲突

诺思-温格斯特命题主张制度约束带来了可观察的经济后果。沃尔德伦主张最著名的宪法约束装置之一找不到可观察的权利保护效果。

**这两个主张不能靠"对象不同"来调和**，因为它们处理的是同一类装置：一个把权力关进程序的安排。若制度约束在十七世纪的英国产生了可测量的效果，则它在当代民主国家找不到效果这一点需要解释；反之亦然。

### 3.3 程序正当性与效力的冲突

沃尔德伦的第二条论证是程序性的：**即使司法审查有好结果，它在程序上也是不正当的。**

而可信承诺一系的整个论证是后果性的：制度之所以好，因为它带来了投资、增长与低通胀。

**两种评价标准互不兼容。**一个后果论者无法接受"即使有效也不正当"；一个程序论者无法接受"有效即正当"。而三个领域各自默认了不同的标准，且都未把标准本身作为议题。

### 3.4 三家共同的盲区

三家都把约束装置当作一个**可以被单独评价的对象**。

政治经济学问：这个制度带来了什么后果。宏观经济学问：这个委托安排是否可信。宪法学问：这个审查机制是否更好地保护了权利、是否正当。

**没有一家问：这个装置的效力是从哪里来的。**三家都把"装置起作用"当作待检验的假设，而把"装置为什么能起作用"当作背景。

而三家各自最强的结果，恰好都指向同一个背景——**装置之外的权力分布**。这一点只有在三家被并置时才显现，而三家互不知晓对方的存在。

## 4 理论框架

### 4.1 核心概念

**自我约束装置**（self-binding device）：由某一权力主体或其前身所设立、旨在使该主体不能做其本来有能力做之事的制度安排。

**禁止解读**（prohibition reading）：把自我约束装置理解为使某类行为**不可能**发生的安排。诺思与温格斯特"不给违约留下余地"的表述是其典型。

**定价解读**（pricing reading）：把自我约束装置理解为使某类行为在被做出时**可见**、因而**可计价的**安排。本文所主张的重新定位。

**效力的外生性**（exogeneity of efficacy）：一个约束装置的实际约束力，不由该装置的条文、设计或形式强度决定，而由装置之外的权力分布决定。

**需求强度**（demand intensity）：一个体系对约束的需求程度，由该体系中权力集中的程度与违约的收益减代价之差共同决定。

### 4.2 命题系统

**命题一（效力的外生性）** 任何自我约束装置的约束力都不能由该装置自身提供，因为执行该装置需要另一个行动者，而该行动者本身不受该装置约束。

**命题二（无穷回退的终止）** 约束的链条必须终止于一个不被约束的东西。这个终点在实践中总是同一类东西：一组行动者不愿或不能推翻某项安排的既成事实。**因此一切约束最终立于权力分布之上，而非立于规则之上。**

**命题三（负相关）** 在其他条件相近时，一个体系对约束的需求强度与该体系中约束的实际效力呈负相关。

**命题四（评价指标的错配）** 由于禁止解读占主导，三个领域评价约束装置时使用的主要指标是"违反的有无与频率"。而若定价解读成立，该指标系统性地低估了装置的作用——因为一个成功的定价装置并不减少违反，它改变违反的代价。

**命题五（定价解读的正面内容）** 自我约束装置的实际功能有三：**使违约可被第三方观察**（规则给出了何为违约的判据）；**使违约的代价可被预期**（先例给出了代价的量级）；**使承担代价的责任可被归属**（程序指出了谁作出了这一决定）。三者合起来构成定价。

**命题六（可观察推论）** 若命题五成立，则一项约束装置的引入应当：不显著降低违约的频率；显著提高违约被第三方识别的比例；显著提高违约后可观察的代价。这三项是可分离的，因而是可检验的。

## 5 分析

### 5.1 效力为什么不能内生

本节论证命题一与命题二。

一部宪法禁止行政机关做某事。若行政机关做了，谁来阻止？答案是法院。

**那么法院靠什么？** 法院没有军队，没有财政。法院的判决之所以被执行，是因为执行机关愿意执行。

**那么执行机关为什么愿意？** 因为不执行有代价——政治的、声誉的、下一次选举的。

**那么这些代价从哪里来？** 从其他行动者的反应中来：反对党会追究、媒体会报道、公众会记得、盟友会疏远。

**这条链条必须在某处终止，** 而它终止的地方不是任何一条规则，是一组行动者的实际反应倾向。

**这一论证的形式是简单的，但它的后果不简单。** 它意味着：任何一个约束装置的实际强度，等于它被违反时所触发的反应强度；而反应强度不由该装置规定，由权力分布决定。

**因此，一个形式上极强的约束（例如宪法中不可修改的条款）与一个形式上很弱的约束（例如一项惯例），在实际强度上可以完全颠倒——**如果前者的违反不触发反应，而后者的违反触发强烈反应。

这一推论在经验上是可检验的，第 9.2 节将其写成失效条件。

### 5.2 负相关命题的论证

本节论证命题三，这是本文的核心。

论证分两步。

**第一步：需求强度与权力集中正相关。**

一个体系对约束的需求，来自违约的可能性。而违约的可能性随权力集中程度上升——一个权力高度集中的主体，其违约的收益更大、遭遇反对的可能更小。

**第二步：约束的效力与权力集中负相关。**

由第 5.1 节，约束的效力等于违约所触发的反应强度。而反应强度随权力集中程度下降——权力越集中，能够作出有效反应的独立行动者越少。

**两步合起来即得命题三：**需求强度随权力集中上升，效力随权力集中下降，故二者反向。

**这一论证的形式是清楚的，但它有一个必须立即指出的循环风险。**

若"效力"被定义为"违约触发的反应强度"，而"需求"被定义为"权力集中程度"，那么命题三就接近于一个由定义得出的结论，而非一项经验主张。

**本文对这一风险的处理是：**把效力独立地操作化为**可观察的后果**，而不是操作化为反应强度。具体地说，效力的操作化指标是命题六所列的三项——违约被识别的比例、违约后可观察的代价、以及责任归属的明确程度。这三项都可以在不预先测量权力集中程度的情况下被测量。

**这一处理使命题三成为可检验的，但也使它变弱：**它不再是一个必然的推论，而是一个待检验的预测。本文接受这一削弱。

### 5.3 三个案例在负相关命题下的重读

**案例一。** 诺思-温格斯特的证据未能通过检验，在本文框架下不意外，但原因与通常的批评不同。

通常的批评是：制度可能不重要，或利率受其他因素影响。 **本文的解释是：**一六八八年的制度变化本身是权力转移的结果——议会取得了对税收、支出的否决权与监督权，这是权力分布的改变。而制度条文是这一改变的记录。若利率对此有反应，它反应的是权力分布，而制度条文与利率是同一个原因的两个结果。

**这个解释有一个可检验的后果：**若权力转移与制度成文在时间上可以分离，则利率的变化应当跟随权力转移而非成文。这是一项本文未做的检验，此处作为推论提出。

**案例二。** 委托的限度这一结果，在本文框架下是命题一的直接实例：委托安排的可信性取决于否决者结构，即取决于违约会触发多强的反应。

**案例三。** 沃尔德伦的前提——良好运转的民主制度、公民认真对待权利——描述的正是一个需求强度低的体系。

在这样的体系中，本文预测约束装置找不到独立的效力，因为它所防止的事情本来就不太发生。而沃尔德伦的第一条论证，正是“找不到证据表明它更好地保护了权利”。

**这一重读必须谨慎表述：**它不表明沃尔德伦错了。他的论证在其前提下是有效的。本文所指出的是，他的第一条论证与本文的预测一致，因而不能被用来区分“司法审查无效”与“该社会不需要司法审查”这两种解释。而这两种解释在政策上截然不同。

### 5.4 为什么禁止解读如此顽固

若定价解读更准确，为何三个领域普遍采用禁止解读？本节给出三条原因。

其一，禁止解读是设立者的语言。

一部宪法不会写“本条使违反本条的代价提高到某个水平”。它写“不得”。**约束装置的文本形式必然是禁止式的，因为定价式的表述会自我瓦解——**一条明码标价的禁令等于一份许可。

**因此禁止解读并非误解，而是该装置在运作时必须呈现的样子。** 这一点与本文第 5.5 节将讨论的一个困难直接相关。

其二，禁止解读在评价上更省力。

"有没有被违反"是一个二值判断，容易观察、容易报告、容易比较。而"违反的代价是多少"需要长期追踪、需要反事实、需要跨案例的可比性。在两个指标中，容易测的那个胜出。

其三，定价解读在政治上不可言说。

一个明确承认"这条宪法条款的作用是把违反的代价提高到某个水平"的官员，实际上是在教人如何计算是否值得违反。因此即使参与者私下按定价方式行事，公开表述也必须维持禁止式。

## 5.5 定价解读的一个内在困难，及其处理

第 5.4 节第一条指出了一个真实的困难，必须正面处理，否则本文的正面主张会自我瓦解。

**困难的表述：**若约束装置的效力来自它被相信是禁止，那么公开采纳定价解读会削弱它。**即：本文的主张若被广泛接受，可能会使它所描述的装置变弱。**

本文对此的处理有三点。

**其一，区分参与者与观察者。**定价解读是一个关于该装置如何运作的分析性判断，其恰当的使用者是制度设计者与研究者，而非该装置的日常对象。这一区分不是精英主义的托词——它在其他领域同样成立，例如公共卫生研究者知道某项措施的实际效应量，而这不要求把效应量写进宣传材料。

**其二，实际的参与者多数已经在按定价方式行事。**一个考虑是否突破常规的决策者，实际计算的正是代价而非可能性。本文所改变的并不是他们的行为，而是研究者评价该装置时所用的指标。

**其三，也是本文能给出的最弱但最诚实的一点：**若上述两点不成立——即定价解读的传播确实削弱了装置——则本文构成一项有害的贡献。本文没有办法排除这一可能，此处指出而不绕开。第 9.4 节将其写成一条失效条件的变体。

## 5.6 需求强度高的一端：本文预测什么，以及为什么本文没有这一端的案例

第 10.1 节将承认一项严重的缺口：本文的三个案例都主要落在需求强度低的一端。本节把这一缺口正面处理，因为回避它会使本文的核心命题只剩下一半。

本文在高需求端预测什么。

按命题三，在权力高度集中、违约收益大而反应弱的体系中，约束装置应当呈现以下三项特征。

**其一，形式强度与实际强度的脱钩最为彻底。**条文可以写得极其严厉——包括不可修改条款、宣誓、刑责——而实际约束接近于零。按第 5.1 节，这不是因为条文写得不好，而是因为条文所依赖的反应链在该体系中被切断了。

**其二，装置的失效不表现为公开的废除，而表现为静默的旁路。**公开废除会触发反应，而旁路不会：机构继续存在、程序继续运行、人员继续任职，而实质的决定在别处作出。这一预测有一个可观察的后果：在高需求端，约束装置的形式指标（存在与否、条文数量、机构编制）与其实际效力的相关性应当最弱，甚至可能为正——因为一个被完全旁路的装置没有理由被废除，反而可以被保留下来作为陈列。

其三，四项操作指标中，指标四（执行者独立度）应当解释绝大部分差异。在低需求端，四项指标可能都提供信息；而在高需求端，若执行者不独立，其余三项无论多好都不产生约束——规则再清晰、代价再明确、责任再可归属，只要判定权在被约束者手中，这些性质就只是描述而非约束。

为什么本文没有这一端的案例。

这个问题必须被回答，因为缺口的分布本身可能不是随机的。

其一，高需求端的材料在系统上更难获得。判定一个约束装置在某体系中是否被旁路，需要该体系内部运作的可靠记录，而这类记录在权力高度集中的体系中恰恰最不可得。本文所依赖的三个案例，全部来自有大量公开学术争论、有可检索文献、有多方对质的领域。而这三项条件本身就是低需求端的标志。

其二，这一取材偏向与本文的命题方向相同，因而是危险的。若本文只在能够获得材料的地方找证据，而能够获得材料的地方恰好是约束较为有效的地方，则本文所观察到的“约束在这里找不到独立效力”就可能有别的解释——例如效力确实存在但被其他因素掩盖。

其三，本文不打算用间接材料来填补这个缺口。关于高需求端体系中约束装置运作的二手叙述是存在的，但其可靠性与立场分布使它们不适合作为一项经验主张的基础。用不可核验的材料去支撑本文最缺证据的那一半，会使本文比现在更弱而不是更强。

因此本文的处理是：把高需求端的三项预测明确写出来，标明它们是预测而非发现，并在第 9 节把它们转换为可检验的条件。本文的核心命题因此处于一个不对称的状态——低需求端有材料而高需求端只有推论，而这一不对称是本文最大的弱点，第 10.3 节将再次指出它。

一个可行的检验路径。

尽管本文自己没有做，此处指出一条成本可控的路径：在同一个体系内部，比较不同领域的约束装置。

任何一个体系中，不同领域的权力集中程度都不相同——某些领域有多个独立的行动者，某些领域没有。若命题三成立，则在同一体系内部，那些权力较分散领域的约束装置，其四项指标应当优于权力较集中领域的装置。

这一设计的优点是控制了大量体系层面的混杂因素（法律传统、行政能力、文化），而其困难在于领域之间的约束装置往往不可比。本文认为这一困难可以通过限定于同类装置（例如同一体系内不同行业的监管机构）来部分解决。

---

## 6 与最近邻概念的划界

---

### 6.1 与“纸面宪法”

一个常见的说法是：某些国家的宪法只是纸面上的，不被执行。

**分界点：**纸面宪法论把问题归为**执行不力**，其隐含的处方是加强执行。而本文主张执行力本身就是权力分布的函数，因此"加强执行"这一处方在最需要它的地方无法被执行——**因为若有足够的力量去加强执行，那个力量本身就已经构成了约束，宪法条文并未增加什么。**

**这一分界有实质的政策后果：**纸面宪法论支持制度援助与文本改良；本文的框架预测这类努力在权力高度集中处收效甚微，而这一预测是可以被援助项目的评估数据检验的。

## 6.2 与"制度是内生的"

制度经济学内部有一个成熟的论点：制度不是外生给定的，它由更深的因素（要素禀赋、权力分布、文化）决定。

**共同点：**两者都主张制度不是独立变量。

**分界点有二。** 其一，内生性论点主要处理**制度如何被选择**；本文处理**已被选定的制度如何起作用**。这是两个不同的问题：一项制度可以是内生地被选定的，同时在选定之后具有独立的作用。**本文的主张比内生性论点更强，因为它主张即使在选定之后，效力仍然是外生的。**

其二，也是更要紧的：内生性论点不给出正面的重新定位。它指出制度不是原因，然后停在那里。**而本文给出定价解读，并由此给出一组不同的评价指标。**

## 6.3 与"羊皮纸屏障"

麦迪逊在《联邦党人文集》中曾用"羊皮纸屏障"一语，指出仅靠文本上的权力划分不足以防止侵越，必须以野心对抗野心。

**这是本文最近的近邻，必须仔细处理。**

**共同点：**两者都主张文本不能自我执行，都把实际的约束归于力量的相互制衡。

**分界点有三。** 其一，麦迪逊的处方是**设计更好的力量结构**（分权制衡），其隐含前提是这一结构可以被设计出来并保持；而本文主张该结构本身也需要被执行，因而回退问题并未在麦迪逊处终止。其二，麦迪逊的论述是规范性与设计性的；本文的论述包含一个经验预测（负相关命题），而这是可被检验的。其三，也是最关键的：**麦迪逊没有提出定价解读**——他仍然在禁止解读的框架内工作，只是把禁止的实现方式从文本换成了结构。

**因此本文相对麦迪逊的增量有二：**把回退问题追到底并承认它不终止于任何设计；以及由此提出功能的重新定位。

## 6.4 与"尤利西斯自缚"

自我约束的经典意象是尤利西斯让水手把自己绑在桅杆上，以抵御塞壬的歌声。

**分界点：**这个意象在一个关键处不适用于政治约束——**绑住尤利西斯的是水手，而水手不受尤利西斯当下的命令支配（他们塞了耳朵）。**也就是说，该意象成立恰恰因为存在一个不听命于被约束者的独立执行者。

而在政治约束中，**执行者通常在被约束者的影响范围之内：**法官由行政机关提名、由立法机关确认，其预算由被约束者控制。**因此政治约束缺少该意象所依赖的那个关键条件。**

这一分界指出了一个可操作的判据：**衡量一个约束装置的强度，应看它的执行者在多大程度上独立于被约束者，而不是看条文的强度。**这一判据在第7节被展开。

## 6.5 与"制度只是共同预期的协调点"

一种博弈论解释主张，制度的作用是协调参与者的预期，使某个均衡被选中。

**共同点：**两者都不把制度当作外在的强制力。

**分界点：**协调解释主张制度**选择**均衡，而本文主张制度**记录并标价**已有的均衡。差别在于因果方向与时序：协调解释允许制度在均衡形成之前起作用（作为焦点），本文主张制度的效力来自已经形成的权力分布。

**这一分界是可检验的，且是本文与该解释之间最清晰的分歧点：**若制度能在权力分布未变的情况下切换均衡，则协调解释成立而本文的强版本失败。第9.1节将其写成失效条件。

---

## 7 定价解读的操作化

若约束是定价装置，则应有一套不同于现行的评价指标。本节给出四项。

### 7.1 指标一：违约的可识别率

**问题：**当一项越界行为发生时，第三方能否明确地判定它是越界？

**理据：**定价的前提是可识别。一项规则若表述模糊、判据可争，则违反它不会被清楚地识别为违反，因而无从计价。

**操作：**取一组已发生的争议事件，请互不通气的独立评判者判定"这是否构成违反"，计算一致性。**一致性高的规则是好的定价装置，即使它经常被违反；一致性低的规则不是，即使它很少被违反。**

**这一指标与现行评价的方向相反，因而是本文框架最容易被检验的一项。**

### 7.2 指标二：违约后的可观察代价

**问题：**一次被识别的违约，在此后一段时间内是否伴随可观察的代价？

**操作：**追踪违约后的若干可观察量——相关主体的更替、后续同类行为的频率、第三方（投资者、盟友、选民）的反应。**关键是要与未违约的对照期或对照主体比较，而非只看绝对水平。**

**这一指标的困难在于反事实：**代价可能来自与违约同时发生的其他因素。本文承认这一困难，并认为它使该指标只能在有大量案例时被使用。

### 7.3 指标三：责任的可归属性

**问题：**一次违约发生后，能否明确指出是谁作出了这一决定？

**理据：**定价需要有人付账。一项无法归属的越界不产生代价，因为没有对象承担它。

**操作：**考察该装置是否要求决定被记录、被署名、被说明理由。**这解释了为什么程序要求（书面、说明理由、公开表决）在实际约束中如此重要——它们与禁止无关，与定价直接相关。**

**这一项是本文框架给出的最有实践价值的推论：**改善一个疲软的约束装置，最有效的介入点可能不是加重罚则，而是使决定不可匿名。

### 7.4 指标四：执行者的独立度

**问题：**该装置的执行者，在多大程度上不受被约束者的控制？

**操作：**考察三项——任命权、免职权、预算权。三项中被约束者控制得越少，执行者越独立。

**这一指标由第 6.4 节的分界导出，且它给出一个与直觉相反的建议：在评估一项约束时，条文的严厉程度几乎不提供信息，而执行者的三项独立度提供几乎全部信息。**

### 7.5 四项指标的合用

四项指标不构成一个评分系统，且不应被加总——它们测的是可分离的性质，加总会掩盖信息。

**它们的实际用途是诊断：**一个约束装置若在指标一上失败，问题在规则的表述；在指标三上失败，问题在程序；在指标四上失败，问题在人事与预算，而与法律文本无关。

**而现行的评价方式——看违约的有无——不区分这四种失败，因而不指向任何具体的修改。**

---

## 8 跨领域兑现

### 8.1 国际条约

**现状：**国际法长期面临“没有强制执行者，因而不是真正的法”这一质疑。

**本框架的判断：**按定价解读，国际条约的功能从未是禁止，而一直是定价——它使某国的某项行为被明确识别为违约，从而使其他国家的反应有了共同的判据。**因此"条约经常被违反"不构成对条约无用的证明。**应看的是：违约是否被清楚识别、是否伴随可观察的代价。

**这一重新定位解释了一个常被误读的现象：**那些看起来最"软"的条约机制（报告、审议、点名）可能比看起来最"硬"的条款更有效，因为前者直接作用于可识别性。

## 8.2 公司治理

**现状：**董事会、独立董事、审计委员会被设计为对管理层的约束，而其有效性长期受质疑。

**本框架的判断：**按指标四，关键不在这些机构的形式权限，而在其成员的任命、免职与报酬在多大程度上由被约束者（管理层或控股股东）决定。**在多数实际结构中这三项都由被约束者影响，因此该装置的形式强度与实际强度可以完全脱钩。**

## 8.3 监管机构

**现状：**监管俘获是一个成熟的议题。

**本框架的判断：**俘获通常被理解为监管者被被监管者影响其判断。**按本文框架，更基础的问题是：监管机构的预算与人事在多大程度上受被监管行业经由政治渠道的影响。**这使俘获从一个关于个人操守的问题变成一个关于结构的问题。

## 8.4 学术规范

**现状：**学术不端的规则详尽，而执行被广泛认为不足。

**本框架的判断：**按四项指标，学术规范在指标一（可识别性）上通常尚可，在指标三（可归属性）上因合作署名而弱，在指标四（执行者独立度）上最弱——调查往往由与被调查者同属一个机构、且共享其声誉利益的主体进行。**因此改善的最有效介入点是指标四，而现行的改革多集中在指标一（细化规则）。**

## 8.5 平台规则

**现状：**平台的社群规则被大量违反，而平台常以规模为由解释执行不足。

**本框架的判断：**平台是一个罕见的案例——**执行者完全独立于被约束者（用户）**，因此指标四得分极高。而其弱点在指标一与指标三：规则表述模糊、判定不公开、责任不可归属。**这解释了为什么用户的主要不满不是"规则太松"，而是"不知道为什么被处理"。**

**同时这也指出一个不对称：**平台对用户的约束在指标四上极强，而用户对平台的约束在指标四上为零——没有任何执行者独立于平台。

## 8.6 家庭与亲密关系中的约定

**现状：**伴侣之间的约定（分工、边界、承诺）常被认为“说了也没用”。

**本框架的判断：**这类约定在指标四上必然为零——不存在独立于双方的执行者。因此按本文框架，它们在任何情况下都不可能作为禁止装置起作用。

而它们仍然有价值，恰恰因为定价功能可以独立于执行者存在：一个明确的约定使违反变得可识别（指标一），使责任可归属（指标三）。它不阻止违反，它使违反无法被含糊过去。这一判断给出一个具体建议：这类约定的价值在于表述的清晰度，而非在于任何强制机制。

## 8.7 人工智能系统的行为约束

**现状：**对模型行为的约束通常以规则与训练目标的形式给出，而其可靠性受到广泛质疑。

**本框架的判断：**这是一个在指标四上结构特殊的案例——执行者（部署方）与被约束者（模型）之间的关系，不同于本文其他案例中的任何一种，因为被约束者不是一个有自身利益并会策略性规避的政治主体。

因此本文明确指出：本框架对这一案例的适用性存疑。本文的负相关命题依赖于“被约束者会因权力集中而更倾向违约”这一前提，而该前提在此处是否成立，需要一个本文不具备的判断。此项列出，是为了标明本文框架的一条边界，而非为了应用它。

# 9 可使本文失效的条件

## 9.1 条件一：约束在权力分布未变时改变了均衡

命题三**的强版本**主张效力来自权力分布。

若能找到一个案例：某项约束装置被引入或被移除，而该体系的权力分布在此前后没有可测量的变化，且行为发生了持续的改变——则第 6.5 节所述的协调解释成立，而本文**的强版本**失败。

这一条是可检验的，且是本文最希望被检验的一条。合适的检验场合是那些约束装置的引入在时间上与权力变化分离的案例——例如一项技术性规则的采纳，其通过并未伴随任何政治力量的重组。

**需排除的假成立：**权力分布的变化可能不可测量而仍然发生。因此该检验要求事前指定权力分布的操作化指标。

## 9.2 条件二：条文强度比执行者独立度更能预测实际约束

第 5.1 节推出：形式上极强的约束与形式上很弱的约束，其实际强度可以颠倒。第 7.4 节由此主张执行者独立度提供几乎全部信息。

**这一主张是可以被回归检验的：**取一批约束装置，独立编码其条文强度与执行者独立度（任命、免职、预算三项），以实际约束效果为因变量。

**若条文强度的解释力显著高于执行者独立度——**则本文的框架在其最具操作性的部分失败。

### 9.3 条件三：负相关不成立

命题三是本文的核心，也是最容易被误读为循环的一条。第 5.2 节已把效力独立地操作化为三项可观察指标。

**检验方式：**取一批体系，独立测量其需求强度（权力集中程度）与效力（三项指标），看二者是否负相关。

**若不相关或正相关，命题三失败。**

**本文必须承认这一检验的困难：**权力集中程度的测量本身是有争议的，而不同的测量方式可能给出不同结果。因此这一条的检验结果可能本身成为一场关于测量的争论，而非对本文的裁决。本文没有办法避免这一点。

### 9.4 条件四，也是本文自己最担心的：定价解读的传播削弱了它所描述的装置

第 5.5 节已经承认了这一困难。此处写成正式条件。

**若能显示：在定价解读被广泛接受的场合，约束装置的实际效力下降——**则本文不仅在描述上可能正确，在后果上是有害的，而一项有害的正确描述是否应当被发表，是一个本文不打算回答的问题。

**这一条与前三条的性质不同：**前三条若成立，本文错了；这一条若成立，本文可能是对的，而这使情况更糟。

**本文能给出的唯一缓解**是第 5.5 节的第二点——实际决策者多数已经在按定价方式行事，因此本文所改变的主要是研究者的评价指标，而非决策者的行为。**这一缓解的强度取决于该经验判断是否为真，而本文没有检验它。**

---

## 10 局限与结论

### 10.1 局限

其一，案例数为三，且三个案例都来自制度已经相对稳固的社会。三个案例中的两个（英国十七世纪、当代民主国家的司法审查）都处在需求强度较低的一端，而本文关于需求强度高的一端的主张——约束在那里失效——**没有一个案例支撑**。这是本文经验基础上最严重的缺口。

其二, "权力分布"这一概念缺乏独立度量。 本文反复使用它作为解释项, 而它的操作化在第 9.3 节被承认为有争议的。这是本文概念上最软的一处。

其三, 本文的第 8.7 项已自行标明框架边界, 即人工智能系统的行为约束是否适用本框架存疑。这一坦白同时意味着: 本文的负相关命题依赖一个关于被约束者策略性行为的前提, 而本文未系统考察该前提在何处成立。

其四, 本文未处理约束装置的分层。 现实中的约束往往是多层的——宪法约束立法, 立法约束行政, 行政约束下级。本文把它们当作同类处理, 而分层结构中的效力传递可能有本文未考察的性质。

## 10.2 结论

本文的主张可凝缩为四句。

第一, 任何自我约束装置的效力都不由该装置自身提供, 因为执行它需要另一个不受它约束的行动者; 这条链必须终止于权力分布。

第二, 因此约束的效力与对它的需求呈负相关: 越需要, 越不生效; 越不需要, 越生效, 而那时它主要在记录一个已有的平衡。

第三, 三个领域各自最强的结果——利率证据的瓦解、委托的限度、司法审查找不到独立效力——是同一件事在三处的显现, 而三家都把它当作本领域的经验难题。

第四, 若上述成立, 则自我约束装置应被重新定位为定价装置而非禁止装置; 其评价指标应从"是否被违反"改为可识别性、可观察代价、可归属性与执行者独立度四项。

这四句话不裁决任何一场具体争论。它们提供一个在设计或评价任何约束装置时可以使用的位置: 先不要问这条规则有多严, 先问一句——如果它被违反, 谁会知道、谁会付账、以及那个要它付账的人, 工资是谁发的。

## 10.3 本文的自我审查

其一, 本文自己没有任何执行者。

本文提出了一组评价指标, 而没有任何机制使任何人采用它们。按本文自己的框架, 一项没有执行者的规范建议其效力为零。

本文对此的唯一回应是: 本文不是一个约束装置, 它是一个描述。但这个回应是弱的——因为本文在第 7 节确实提出了规范性的建议, 而那些建议与本文所描述的无执行者约束具有相同的结构。

其二, 本文所处的评价体系, 正是本文所诊断的那一类。

本文将被评价, 而评价的执行者与本文的作者共享同一个产出体系。按第 7.4 节的指标四, 这一评价的独立度不高。而本文无法通过任何自身的安排来提高它。

其三，第 10.1 节第一项所承认的那个缺口是致命的。

本文的核心命题是关于需求强度高的一端——约束在那里失效——而本文的三个案例全部或主要落在另一端。这意味着本文最重要的那半个主张，在本文自己的材料中没有直接证据。

本文没有为此辩解。它只能指出：第 9.1 至 9.3 条所述的检验，其价值主要在于覆盖这个缺口，而本文把它们留在不能被事后撤回的位置上，是本文对自己的唯一约束——而按本文自己的分析，这个约束同样没有执行者。

---

## 注释

〔1〕 本文用"自我约束装置"统称三类在设立主体、执行机制与时间跨度上都不同的安排。这一归并是有意的，其代价是掩盖了三者之间可能重要的差异，尤其是宪法约束的跨代性质——被约束的是后来的多数，而设立者是先前的多数。这一跨代问题是沃尔德伦论证中的一个重要维度，本文未予处理。

〔2〕 第 5.1 节的回退论证是一个熟悉的形式，其结论（约束终止于权力分布）在政治思想史上多次被提出。本文的增量不在这一论证本身，而在由它导出的负相关命题与定价解读。

〔3〕 第 2.1 节所引对诺思-温格斯特证据的三项质疑，各自的技术细节与后续回应本文未详述。本文只使用一项概括性事实：该命题的主要经验支撑未能通过后续检验。读者若要评估该命题本身，应查阅原始文献而非依赖本文的概述。

〔4〕 第 2.3 节对沃尔德伦前提的讨论，不构成对其论证的反驳。该前提是他明确写出并加以限定的，本文所做的只是指出该前提所划定的范围与本文预测的重合。

〔5〕 第 8 节七项兑现中，第 8.7 项被明确标为适用性存疑，其余六项是框架的应用。七项中没有一项包含已被检验的结论。

---

## 参考文献

- Clark, G. (1996). The political foundations of modern economic growth: England, 1540–1800. *Journal of Interdisciplinary History*, 26(4), 563–588.
- Elster, J. (1979). *Ulysses and the Sirens: Studies in Rationality and Irrationality*. Cambridge University Press.
- Fallon, R. H., Jr. (2008). The core of an uneasy case for judicial review. *Harvard Law Review*, 121, 1693–1736.
- Hutchinson, A. C. (2008). A "hard core" case against judicial review. *Harvard Law Review Forum*, 121, 57–64.

- Keefer, P., & Stasavage, D. (2002). Checks and balances, private information, and the credibility of monetary commitments. *International Organization*, 56(4), 751–774.
- Keefer, P., & Stasavage, D. (2003). The limits of delegation: Veto players, central bank independence, and the credibility of monetary policy. *American Political Science Review*, 97(3), 407–423.
- Kydland, F. E., & Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy*, 85(3), 473–491.
- North, D. C., & Weingast, B. R. (1989). Constitutions and commitment: The evolution of institutions governing public choice in seventeenth-century England. *Journal of Economic History*, 49(4), 803–832.
- Quinn, S. (2001). The Glorious Revolution's effect on English private finance: A microhistory, 1680–1705. *Journal of Economic History*, 61(3), 593–615.
- Stasavage, D. (2002). Credible commitment in early modern Europe: North and Weingast revisited. *Journal of Law, Economics, and Organization*, 18(1), 155–186.
- Sussman, N., & Yafeh, Y. (2006). Institutional reforms, financial development and sovereign debt: Britain 1690–1790. *Journal of Economic History*, 66(4), 906–935.
- Waldron, J. (2006). The core of the case against judicial review. *Yale Law Journal*, 115(6), 1346–1406.

## 附录 本文的产生过程

按本栏体例，此处交代产生方式与作废记录。

**三个来源**，分属政治学、经济学、法学：

- **政治学**：诺思-温格斯特的可信承诺命题及其证据在后续检验中的瓦解。
- **经济学**：规则优于相机抉择的经典论证，与委托的限度这一后续结果。
- **法学**：沃尔德伦反对司法审查的核心论证及其引发的持续回应。

**选择这三个的理由**：三者处理的是形式相同的制度设计——使一个有能力做某事的主体变得不能做那件事——而三者对该设计的评价互不相容。

**九条判断**，三家各抽三条最承重的。

政治学那三条：统治者可信承诺的两条路径中，声誉路径被明确判为不可行；制度路径被表述为“不给违约留下余地”；而作为该命题主要证据的利率断点，在后续检验中未能成立、部分反向。

经济学那三条：相机抉择产生系统性偏差故规则优于相机抉择；委托是有限度的；委托的可信性取决于否决者结构，即取决于委托安排之外的东西。

法学那三条：没有理由认为权利经由司法审查比经由立法机关得到更好保护；抛开结果，该制度在程序上剥夺了多数的自治权；而该论证明确建立在“社会具有良好运转的民主制度、多数公民认真对待权利”这一前提之上。

三十六次对撞，作废十三条。其中九条为同义重复，四条撞不起来。

留下的二十三条聚成四条暗流：

一、三家所评价的都是同一种形式的装置。二、三家各自最强的结果都指向装置之外的东西。三、三家都把"装置为什么能起作用"当作背景而非议题。四、三家使用的评价指标都是"有没有被违反"。

第二条是根。它同时解释了第一条为何是同一件事、第三条为何是共同盲区、以及第四条为何系统性地测错了东西。

**收敛成一条判断：**约束在最需要它的地方失效，在有效的地方不被需要；因而它的真实功能不是禁止，是定价。

这条判断三家里谁也说不出来，而原因就在各自的位置上。政治学那一家把证据的瓦解当作一个待解决的经验争议，其回应集中于证据是否要害，而非命题的形式。经济学那一家已经把答案说出了一半——委托是有限度的，限度由否决者结构给出——但该结果被表述为对央行独立性这一具体安排的限定，未被推广为关于一切委托的判断。法学那一家的前提写得最清楚，也因此最接近本文的观察，但该前提被用于划定论证的适用范围，而非被用于追问约束的效力从何而来。

---

**作者** 王德生 + Claude **栏目** 学科通融·之五 **说明** 本文由政治学、经济学、法学三家对同一种制度设计的互不相容评价碰撞而成。全部来源均为站外公开文献，页内链接直达原始出处，可自行核对。

---

本文由三个分属不同学科、且互相冲突的理论体系碰撞而成。全部来源均为站外公开文献，可自行核对。 作者：王德生 + Claude · SDE Universes · 学科通融