

协议约束博弈，利己为先不断生成新的博弈

网页全文 PDF 版

根据网站公开版导出 | 2026-10-17。保留网页正文、版本边界、参考文献与续读。本版是网页全文的独立下载版，不是另一个 DOCX 母版，也不表示原刊物已完成核验。下文关于源 PDF 未挂载的陈述为导出时网页原有说明，不指本网页导出版。

<https://www.civitas.top/documents/p03-surv-002/>

本文解决什么问题

协议仍然重要，但真正需要穿过时间的不是一份永恒协议，而是旧主体、旧均衡和旧规则失效以后仍能重新生成纠错与合作的能力。

版本与证据边界

当前版本：V1.1。本页提供上传源文档的在线文本转录；PDF 资产尚未通过本站二进制发布流程挂载，因此不提供虚构下载链接。表格与复杂版式以源文档为准。

继续阅读

- 看一个真实窗口：中美 AI 事件沟通渠道

<https://www.civitas.top/documents/p03-surv-004/index.html>

- 继续进入总体战略判断：唯一之路

<https://www.civitas.top/documents/p03-only-002/index.html>

- 共同存续怎样进入国家—国际结构

<https://www.civitas.top/documents/p03-strat-001/index.html>

在线全文

文明跃迁研究组 · 思想短篇

协议约束博弈，利己为先不断生成新的博弈

——为什么文明真正需要穿过时间的，不是一份永恒协议，而是能够重新生成和平与纠错的内生逻辑

思想短篇 V1.1 | 根判断与续读校准版

作者：子君赋 | 出品：文明跃迁研究组

协议可以约束已经发生的博弈，却无法阻止利己为先的逻辑不断重新生成新的博弈。

人类历史并不缺少协议。停战、联盟、军控、贸易安排、危机沟通、国际法与国内制度，都曾在不同时间发挥真实作用。问题也不能被简单写成“所有协议最终都会失败”。经验研究甚至发现，在特定历史样本中，大量联盟承诺得到履行。[1]

真正需要追问的是另一件事：为什么一份在签署时合理、有效、甚至被认真遵守的协议，不能仅凭“已经签过”就保证穿过几十年、几代人甚至更长时间？

一、协议是一时一地的均衡，不是穿越时间的自然常数

协议成立时，主体、利益、力量、技术、风险与对未来的预期处在某一种组合中。双方之所以愿意遵守，通常意味着在当时条件下，履约比违约更符合各自的判断。

但时间不会停在签约那一刻。政党会轮替，领导层会改变，社会可能发生重大变革，国家力量会此消彼长，新技术会改变原来的安全边界，新的主体与新的利益也会出现。

因此，真正履行协议的主体，可能已经不是当初形成协议时那个意义上的主体。Leeds 对 1816—1944 年联盟承诺的研究发现，力量变化以及政策制定过程的变化，是违约的重要预测因素之一。[1] Fearon 与 Powell 关于 commitment problem 的研究则进一步说明：即使今天存在双方都优于冲突的安排，当一方无法可信承诺未来仍会遵守，尤其当未来力量结构快速变化时，和平安排本身仍可能变得不稳定。[2][3]

协议不是被纸张腐蚀，而是被协议签署以后发生的世界重新定价。

二、真正的考验，不是遵守最有利的时候

一份协议在双方都明显受益时被遵守，并不能回答它最困难的问题。真正的考验发生在第一次出现这种情形的时候：

- 力量变化以后，破坏规则开始可能带来更大优势；
- 危机到来以后，隐瞒、延迟或抢先行动开始可能有利；
- 新的领导人与新的社会力量不再把旧承诺理解为自己的承诺；
- 技术跃迁以后，旧规则约束的对象已经发生变化；
- 原来共同面对的外部对手消失以后，原来的“我们”重新分化成不同的“我”。

如果一份协议的稳定完全依赖“所有参与者在所有未来条件下都继续认为遵守更有利”，那么它的长期稳定本身就依赖一个很强的前提：未来的主体、利益和力量结构不要发生足以改变计算的变化。

协议真正的考验，从来不是双方都有理由遵守它的时候；而是第一次有一方认为破坏它更有利的时候。

三、协议约束已经发生的博弈；利己为先不断生成新的博弈

这就是外部规则与底层逻辑之间的区别。

外部约束

主体先形成目的：我要赢、我要安全、我要领先、我要控制更多资源。协议、法律和制度随后规定哪些手段不能使用、哪些边界不能跨越。

这种结构可以非常有效。问题在于，只要目的生成的第一逻辑没有改变，规则始终承受一个内在压力：如果绕过规则更能实现原目的，为什么不绕过？

内生逻辑

更深的变化不是“我已经决定要做什么，只是被禁止这样做”，而是：当某个目的会摧毁自己与他者以后继续存在、纠错、退出和重新选择的条件时，这些后果在目的锁定以前就已经具有改变目的资格。

此时，合作、纠错与共同存续不再只是目的形成后的外部成本，也开始成为目的生成本身的一部分。

协议可以约束已经发生的博弈，却无法阻止利己为先的逻辑不断重新生成新的博弈。要改变博弈的长期命运，最终必须触及生成目的的一阶逻辑。

四、真正需要穿过时间的，不是同一份协议

没有理由要求一份今天的协议一百年以后仍然一字不变。主体会变，技术会变，风险会变，旧规则被新规则替代本身并不等于文明倒退。

真正重要的是：旧协议失效以前，是否还能形成新的协议；旧主体改变以后，新主体是否仍有办法重建共同安全；力量结构发生逆转以后，新的均衡是否不必经过不可逆冲突才能出现；一次严重判断错误以后，纠错通道是否仍然活着。

所以，文明真正需要保存的不是 Agreement Persistence——协议文本永远持续，而是 Correction Regeneration——纠错结构能够重新生成。

协议可以失效，纠错不能死。

这也意味着，和平的长期稳定不能仅理解为“把今天的冲突压住”。真正更难的问题是：新的主体、新的利益和新的时代不断出现以后，和平与纠错能否一次又一次重新生成。

五、“内生”不是道德要求，而是目的生成位置的变化

利己为先，不排斥合作，但内生博弈。

利他为先的利己，不排斥竞争，但内生合作。

“内生”描述的是目的生成的第一阶关系结构，不是最终行为的绝对预测。利己主体可以合作，利他为先的主体也可以竞争；真正不同的是，合作与竞争分别在什么位置进入目的形成。

这里所说的内生，不是要求主体放弃自身利益，也不是要求永久和平、单边退让或取消防御。

它指向的是目的生成顺序：在自己的目的被锁定以前，相关他者、共同后果与未来纠错条件是否已经能够进入，并真实改变目的。

在“利己为先”的结构中，合作可以出现，但往往需要首先被证明对我有利；一旦回报、威慑或反制消失，合作理由可能随之消失。

在“利他为先的利己”中，他者并不需要先成为“我”、先证明可以回报我、或先证明能够伤害我，才取得改变目的的因果资格。自身利益仍然存在，防御仍然存在，但自身利益不再天然占有目的生成的第一先手位置。

因此，这不是把一份新协议放在旧逻辑之上，而是尝试改变未来每一代主体在形成目的时，为什么会再次生成合作、纠错与共同存续条件。

六、AI、AGI 让这个问题第一次变得更急

在人类能力有限的时代，即使目的有问题，距离、组织成本、信息不足、执行速度与资源限制仍会给规则追赶留下时间。AI、AGI 及更强智能可能同时削弱这些天然阻力。

如果未来智能只继承“完成目标以后再由规则纠偏”的结构，那么能力越强，下游规则越需要不断追赶更快、更复杂的目的执行。

更深的选择是：不仅给未来智能更多规则，还让“新后果是否应当重开原目的”“谁在目的锁定以前就应当被看见”“哪些行动会取消未来纠错条件”进入目的生成本身。

文明真正需要传给未来智能的，未必是一套永远不变的答案，而可能是产生答案的底层方式。

七、开放边界：我们的任务不是替世界找完所有道路

这篇文章不承担证明“宇宙中绝不存在第二条路”的任务，更不承担证明第二条路存在的任务。

如果有人认为外部协议、机制设计、威慑、分布式治理或其他完全不同的道路，能够在不触及这里所说的底层目的生成逻辑的情况下，长期穿过主体更替、力量变化、技术跃迁与危机，并持续保留纠错与共同存续条件，可以完整提出。

我们允许，也邀请这样的探索。真正不同的道路如果被提出，可以进入公开讨论与检验；必要时也可以协助把它说清楚、展开因果和失败条件。

但求证这些替代路线、替世界穷尽所有可能道路，不是本文的任务。本文继续做的是：把已经看到的问题和因果说清，把已经形成的方向推进到现实。

我们不替世界寻找完所有道路，但欢迎所有道路出现。

八、真正重要的，不是更多人接受同一个答案

开放讨论的目的也不是把“唯一之路”变成一次赞成票统计。越多人参与，真正增加的应该是不同主体的独立推演：不同经验、不同学科、不同利益位置、不同文化与不同反例，共同进入同一个问题。

有人重新走一遍因果链，得到与本文相同的结果，理解因此更深；有人指出逻辑断点，判断因此修正；有人提出更好的道路，问题因此得到解决；有人把一个尚未解决的问题变成现实行动，思想因此进入世界。

真正重要的，不是更多人相信同一个答案，而是更多人开始真正思考同一个问题。

九、结语：文明需要的是能够重新生成和平的逻辑

人类历史一次次建立协议，又一次次面对新的主体、新的利益和新的力量。协议因此仍然必要，但协议本身不是稳定的根。

如果文明只在冲突以后重新加规则，就永远需要下游制度去追赶上游目的；如果目的生成本身开始把他者、共同后果与未来纠错提前纳入，那么规则与协议才可能越来越成为这种内生逻辑的现实表达，而不是永远与它对抗的外部刹车。

真正稳定的和平，不是被规则压住的博弈，而是新的主体出现以后，仍能重新生成不把彼此推向不可逆的理由。

这不是一份关于“永久和平”的承诺。它只是把问题推进到更靠前的位置：我们真正要改变的，也许从来不是某一份协议，而是未来的主体为什么还会重新生成同一种博弈。

参考文献与邻近研究

[1] Leeds, B. A. Alliance Reliability in Times of War: Explaining State Decisions to Violate Treaties. *International Organization*, 57(4), 801–827, 2003. DOI: 10.1017/S0020818303574057.

[2] Fearon, J. D. Rationalist Explanations for War. *International Organization*, 49(3), 379–414, 1995.

[3] Powell, R. War as a Commitment Problem. *International Organization*, 60(1), 169–203, 2006. DOI: 10.1017/S0020818306060061.

续读

- 上游现实结构：《危机涌现：AGI时代国家间博弈的结构性新形态》v4.0。
- 总体战略判断：《唯一之路·人工智能时代的国家长远利益与文明跃迁》v6.21。
- 现实治理结构：《人工智能治理阵营化与人类整体安全》v2.0。
- 制度展开与存续主轴：《人类命运共同体的实现、演变与跃迁》V3.9.3-P1；《AI、AGI、ASI时代，文明何以继续？》V1.1.1；REAL-ACT 01《中美AI事件沟通渠道》V0.2.1。

作者：子君赋 | 出品：文明跃迁研究组