arXiv:2601.12046v1 [econ.TH] 17 Jan 2026

不可逆失败逆转信息价值

Nicholas H. Kirk[1], [2*1] Markovian Group[2] Sa¨ıd 商学院，牛津大学

摘要

我们研究具有隐藏状态和吸收性失败的动态博弈，其中由信念驱动的行动可能引发不可逆崩溃。在这类环境中，能够维持活动的均衡通常处于可行性边界。我们表明，这种几何结构会内生地逆转信息价值：在任何有限时间范围内，更高的信息精度都会增加崩溃概率。我们通过极限可行性准则将这一机制形式化，并将不透明性建模为通过布莱克威尔劣化对信息结构进行的战略选择。当失败具有吸收性时，生存价值关于信念在局部呈凹性，这意味着透明度会破坏均衡可存续性，而足够的不透明性则能恢复它。在一个代理人事前选择信息结构的扩展博弈中，严格为正的不透明性是均衡存续的必要条件。

研究结果表明，在动态博弈中，产生对不透明性的内生需求的根本力量是不可逆失败，而非协调、模型设定错误或模糊性。

关键词： 不可逆失败、策略性不透明、武器化信息、极限可行性、布莱克威尔劣化、吸收状态、动态博弈。**JEL：** C73, D82, D83, G01, H12。

1 引言

人们普遍认为，信息能够稳定经济行为：更高的透明度有助于改善学习、约束行动并提升效率。然而，当失败不可逆时，这一假设便会失效。在许多动态环境中——例如破产、主权违约、政权崩溃或反叛——即使是微小的信念波动，也可能触发永久性摧毁系统的行动。在此类情形下，更高的信息精度反而可能增加灾难性结果发生的可能性，而非降低这种可能性。

本文为这一逆转提供了结构性解释。我们研究具有隐藏状态和吸收性失败的动态博弈，其中，在信念极为不利时采取行动，可能触发永久性崩溃。吸收性失败从根本上改变了延续价值的几何结构：它将原本平滑的收益函数转变为悬崖状的生存价值。关于不可逆失败下可持续行为的既有研究指出，维持活动的均衡通常处于可存续边界，因而没有可抵御崩溃的内部缓冲（Kirk, 2025）。这一机制对行为有直接影响。当信息以这种方式被武器化时，在可行性边界附近，更高的信息精度会严格降低预期生存率

*电子邮箱：sbsdfs26009@said.oxford.edu

。因此，代理人会最优地限制观测通道的信息性，并非因为信息遭到扭曲，而是因为不透明性会使生存价值凹化，并抑制那些会触发不可逆失败的信念实现。通过内生地对信息进行模糊化，代理人降低后验信念的离散程度，并抑制导致崩溃的信念实现发生的可能性。因此，策略性不透明并非被建模为虚报或虚假陈述，而是对信息结构（模糊化）的一种策略性选择；这种选择可能产生欺骗效果，但仍完全符合贝叶斯原则。

本文通过将不透明性建模为经由布莱克威尔劣化对信息结构作出的策略性选择，形式化了这一逻辑。我们采用极限可存续性准则评估结果：随着信息噪声趋于消失，在每个有限时间范围内，均衡都必须以任意高的概率避免崩溃。对于存在吸收性失败的环境，这一准则十分适用，因为在全支持噪声下，几乎必然生存无法实现，而期望收益标准可能掩盖脆弱性。

我们的主要结果表明，在武器化信息环境中，不透明性增强会弱扩大极限可存续结果的集合。我们证明，透明度可能引发全面崩溃：存在这样的环境，在最大透明度下不存在极限可存续均衡，而足够高的不透明性则能恢复均衡存在性。当代理人可以在事前选择不透明性时，任何满足极限可存续性的均衡都必须具有严格为正的不透明性。在标准贝叶斯说服中，信念下价值函数的曲率被视为原始给定条件；只有当收益外生地呈凹性时，信息才有害。

本文指出，在一大类具有重要经济意义的动态环境中，不可逆失败会内生地产生凹性。一旦认识到这一几何结构，劝说逻辑便随之而来，而非作为假设：不透明性是对生存价值进行最优凹化的结果。

核心机制可以简洁地概括如下。在存在吸收性失败的动态环境中，信念到延续价值的映射是不连续的。当均衡行动在失败阈值附近发生离散切换时，精确信息导致的后验集中会增加遭遇不可逆崩溃的风险。不透明性通过抑制那些越过生存悬崖的信念实现来缓解这一风险。本文的贡献并非提出一种新的均衡概念，而是给出一个结构性不可能结果，揭示了存在不可逆失败的动态博弈中，透明度与可行性之间的联系。

本文结构如下。第2节将本文的贡献置于信息设计、全球博弈和脆弱性相关文献的背景下。第3节介绍环境以及不透明性的建模。第5节提出极限可存续性准则。第6节对武器化信息进行形式化，第7节推导主要结果。第8节和第9节给出说明性应用。第10节总结全文。

2 相关工作

与全球博弈和危机脆弱性的关系。 大量文献表明，小幅噪声能够在协调环境中引发危机动态：它既可能通过选出唯一均衡来发挥作用（Morris 和 Shin，1998年），也可能在信息内生生成时（例如通过价格或政策）维持多重性和波动性（Angeletos等，2006年；Angeletos and Werning，2006年）。本文提出的机制则有所不同。不同于噪声通过协调效应引发恐慌的模型，本文指出存在一类环境，在这些环境中，更高的信息精度 本身就会造成不稳定，即使不存在多重性或战略不确定性也是如此。我们不研究

在外生噪声下由策略互补性驱动的均衡选择或多重性。相反，吸收性失败会内生地导致生存价值关于信念呈现局部凹性，因此，即使战略环境保持不变，更高的信息精度也可能增加受到崩溃冲击的风险。此外，本文关注的关键行为边际是对信息结构进行内生控制（通过布莱克威尔劣化实现不透明性），并依据极限可存续性准则进行评估，而非固定信息结构下均衡阈值的比较静态。

全球博弈与消失噪声。 本文与全球博弈文献相关。在该文献中，消失的私人噪声通过打破共同知识并消除多重性来细化均衡选择（Carlsson 和 van Damme，1999年；Morris 和 Shin，1998年）。在这类文献中，信号精度的提高具有稳定作用：随着噪声消失，行为平稳地收敛于选定的均衡。本文的机制则有所不同。在全球博弈中，噪声在多个均衡之间进行选择；而在本文中，精度会使均衡存在性消失。在存在吸收性失败的情况下，从信念到结果的映射是不连续的。随着信息噪声消失，后验集中反而可能激活触发失败的行动，导致在有限期内以正概率发生崩溃。透明度并非选择均衡，而是可能彻底破坏可行性。

信息披露、信号传递与信息设计。 不同于信号传递或信息披露模型，本文中的不透明性并不涉及虚报、说服或对信念的策略性扭曲（Grossman，1980年；Milgrom，1981年）。信念仍然是贝叶斯式的，且设定正确。不透明性被建模为对信息结构本身的内生选择——通过在行动之前对信号进行布莱克威尔劣化实现（Blackwell，1953年；Kamenica 和 Gentzkow，2011年）。其机制并非声誉操纵或说服，而是在具有不可逆失败的脆弱系统中，避免由信念驱动的可行性违约。

稳健控制与模糊厌恶。 这些结果不同于稳健控制和基于模糊性的研究方法；后者通过偏好、最坏情形信念或模型设定错误引入保守主义（Hansen 和 Sargent，2008；Epstein 和 Schneider，2003）。在本文中，代理人完全遵循贝叶斯方法，且模型设定正确。选择不透明性并非因为代理人厌恶不确定性，而是因为在信息被用作武器的环境中，透明度越高，遭遇不可逆失败的风险就越大。

可行性、可持续性与吸收状态。 对吸收性失败的关注与动态系统中关于可行性和可持续性的研究相关（Aubin，1991年；Altman，1999年）。吸收状态和不可逆转变也是违约、政权崩溃和制度瓦解模型中的核心概念（Acemoglu 等人，2005年）。本文的贡献在于表明，在不可逆崩溃的情况下，信息结构本身会成为一种策略空间：随着信息摩擦消失，代理人可能会选择最优地粗化信息，以维持存续。

与学习和模型设定错误的关系。越来越多的文献通过学习动态、模型设定错误或稳健性顾虑来解释脆弱性：在这些情形下，代理人因信念扭曲或引入最坏情形推理而采取保守行动（Epstein 和 Schneider，2003；Hansen 和 Sargent，2008）。有关罕见灾难和尾部风险的相关研究强调，在低频但具有灾难性的事件发生时，信念更新如何发挥作用，并表明悲观学习能够

产生持久的宏观经济效应（Kozlowski 等，2020）。近来，基于模型设定错误的均衡概念将代理人在其主观模型存在系统性错误时的行为形式化（Esponda 和 Pouzo，2016）。

本文的机制有所不同。代理人完全遵循贝叶斯原则，模型设定正确，并且对环境具有共同知识。本文不假设模糊偏好（Epstein 和 Schnei- der，2003年）、最坏情况扭曲（Hansen 和 Sargent，2008年）或错误设定的学习动态（Esponda 和 Pouzo，2016年）。即使在消失噪声下信念趋于收敛，信息价值也会发生逆转：更高的信息精度会通过激活存续边界附近由信念触发的行动，增加暴露于不可逆失败的风险。

因此，策略性不透明并非对认知不确定性或悲观学习（Kozlowski 等，2020年）的回应，而是在正确的信念本身可能具有破坏性的环境中，对信息结构进行的理性调整。这表明，驱动对不透明性的内生需求的根本力量是不可逆性，而非学习或模型设定错误。

与存续性选择相关的研究。存续性选择考察的是：当各种性质的扰动为外生给定时，哪些均衡能够存续；而策略性不透明考察的是：一旦代理人认识到这种脆弱性，他们如何最优地调整观测通道。二者相辅相成：选择在给定信息环境下筛选行为，而不透明性则改变信息环境。

行为解释。 尽管这一机制涉及信息，但其贡献在于行为层面，而非认知层面。代理人不会误解、扭曲或操纵信念。相反，他们会将选择处理或披露多少信息作为自身策略行为的一部分。在存在不可逆失败的环境中，限制信息暴露成为一种理性的自我保护方式。

本文也与早期关于存在不可逆失败的动态博弈中的可持续性和均衡可行性的研究相关。Kirk，2025 对可持续开发均衡（SEE）的研究刻画了在精确或稳定信念下维持可行性的均衡行为。本文探讨的是一个不同的问题。本文将可持续均衡处于边界这一特征视为既定条件，研究当可行性容易受到信息精度影响时，代理人如何最优地塑造信息结构本身。从这个意义上说，SEE 分析的是以信念为条件的可行性，而策略性不透明分析的是：当信念分散内生地威胁可行性时，代理人围绕信息结构所采取的行为。

因此，本文通过指出一种旨在降低信息精度的信息设计动机，补充了危机/全球博弈文献；这种动机并不依赖多重性，而是源于不可逆失败。与全球博弈、说服或稳健控制模型不同，本文表明，即使是信念正确的完全贝叶斯代理人，在失败不可逆时，为了确保均衡存在性，也可能需要更少的信息。

表1概述了这一机制与全球博弈和贝叶斯说服的显著差异：此处的信息不会选择均衡或重新分配行为，而是决定均衡存续是否可行。

3 环境

时间是离散的，索引为 $t = 0$、1、2、.....。玩家集合 $N = \{1, \ldots, n\}$是有限的。收益相关状态 $s_t \in S$ 按照紧致度量空间 S上的受控马尔可夫核 $Q(\cdot \mid s_t$、 $a_t)$演化，其中 $a_t = (a_{1t}$、.....、 $a_{nt})$是日期 t的行动组合。

子集 $F \subset S$ 是吸收态：一旦 $s_t \in F$，就会永久崩溃，且所有延续收益均为零。

每位玩家 i 都有一个紧致的行动空间 A_i，并在每个时期选择行动 $a_{it} \in A_i$ ，直到发生失败。流收益由可测函数给出

$$u_i : S \times A \to \mathbb{R}, \quad A := \prod_{i \in N} A_i,$$

并作如下归一化：对于所有 $s \in F$， $u_i(s, a) = 0$。

玩家无法直接观察状态 s_t 。相反，在给定 s_t 的条件下，他们观察到由观测通道 $\kappa_\varepsilon(\cdot \mid s_t)$ 生成的信号，其中 $\varepsilon > 0$ 表示信息噪声，且$\varepsilon \downarrow 0$。信号满足单调似然比性质。

策略是从私人历史到行动的可测映射。均衡是所诱导的动态博弈的完美贝叶斯均衡（PBE）[1] ，并在可能的情况下根据贝叶斯法则更新信念。

令 $\Delta(S)$ 表示 S 上的博雷尔概率测度集合。令 $b_t \in \Delta(S)$ 表示在均衡策略组合和公共信息结构下，通过贝叶斯更新对 s_t 诱导出的共同后验信念。共同后验意味着，条件于公共历史和观测通道，玩家关于 s_t 的信念几乎处处一致。

对于任意 策略组合 σ 和任意有限期 $T < \infty$，定义有限期存续值 lue

$$V_T(b; \sigma) \equiv 1 - \Pr_\sigma(\tau_F \leq T \mid b_0 = b),$$

也就是说， 从信念 b 开始，直至时间 T 避免吸收失败集的概率。

以下所有结果均不受玩家数量影响，也不依赖于对称性或特定的收益函数形式。

均衡与信息结构

我们施加以下正则性条件：

假设 1 （正则性）。 ***(i)*** 马尔可夫核 $Q(\cdot \mid s, a)$ 关于 (s, a) 连续，并满足费勒性质。

(ii) 收益函数 u_i： $S \times A \to \mathbb{R}$ 连续且有界。

(iii) 观测通道 $\kappa^\varepsilon(\cdot \mid s)$ 相对于某个支配测度具有密度，在全局范围内满足单调似然比性质，并且当 $\varepsilon \downarrow 0$ 时弱收敛至 s 处的退化分布。

(iv) 失败集 $F \subset S$ 是闭集。

[1]对于以下结果，采用完美贝叶斯均衡并不失一般性；所有论证仅依赖于均衡路径上的序贯理性和贝叶斯更新。

共同后验。 在具有公共观测通道 κ^{ε}的完美贝叶斯均衡下，所有玩家观测到的信号均来自同一条件分布（但存在方差趋于零的独立特质噪声）。因此，随着私人噪声趋于零，以公共历史为条件的后验信念 b_t 依概率收敛到一个共同极限。我们将此极限称为共同后验。所有关于信念和生存价值的陈述，均指沿着这条共同后验路径成立。

有限期限收益。 对于任意有限期限 $T < \infty$ 和初始信念 $b \in \Delta(S)$，玩家 i在策略组合 σ 下的预期折现收益为

$$U_i^T(b;\sigma) = \mathbb{E}^{\sigma}\left[\sum_{t=0}^{T} \delta^t u_i(s_t, a_t)\mathbf{1}_{\{t<\tau_F\}} \,\middle|\, b_0 = b\right], \tag{1}$$

其中， $\delta \in (0,1)$ 是共同折现因子。

4 内生不透明性

代理人在开始行动前选择观测通道的信息性。

假设 2 (不透明性即布莱克威尔劣化)。 对于 $\lambda' \geq \lambda$，通道 $\kappa^{\varepsilon,\,\lambda'}$ 是 $\kappa^{\varepsilon,\,\lambda}$的布莱克威尔劣化。不透明性更高的通道在布莱克威尔序下的信息性至多相同。

较高的 λ 对应更高的不透明性。

用信息设计的语言来说（Blackwell，1953年；Kamenica 和 Gentzkow，2011年），选择 λ 相当于选择一个可行的后验分布（其均值固定为先验）。更不透明的通道会产生在凸序意义下离散程度较低的后验分布；当生存价值关于信念呈凹性时，这正是相关的比较静态（命题1）。

注释1 （不透明性与造假）。 本文将不透明性建模为对信息性的选择（布莱克威尔劣化）。这不同于策略性误报或撒谎，不过在实际应用中，对测量和信息披露的制度性控制可能服务于类似的目标。

对信息结构的控制。 在全文中，不透明性是对公开可用的观测通道所作的选择。个体代理人无法私下且无成本地偏离信号精度。这对应于信息获取集中化、受技术限制或受制度管理的环境（例如会计准则、统计机构、殖民地行政机构或披露制度）。因此，均衡问题在于代理人是否会最优地选择他们共同面对的信息结构，而不在于他们是否会秘密获取额外的私人信号。

5 消失噪声下的可行性

在全支持噪声下，几乎必然生存无法实现。因此，我们采用适用于不可逆失败的极限可存续性准则。

定义 1 （极限可存续性）。 策略组合 σ 在不透明性 λ 下是极限可存续的，当且仅当对每个有限时间范围 $T < \infty$,

$$\lim_{\varepsilon \downarrow 0} \Pr_{\sigma,\kappa^{\varepsilon,\lambda}}(\tau_F \leq T) = 0.$$

这要求随着信息噪声消失，在每个有限时间范围内崩溃发生的可能性都能任意趋近于零。

6 武器化信息

在静态说服问题中，价值函数关于信念的凹性是一项有关偏好的假设。在存在吸收性失败的动态环境中，凹性则源于信念更新与不可逆状态转移之间的相互作用。下一个定义将这种涌现的几何结构形式化。

定义 2（武器化信息）。 固定策略组合 σ 和有限期 $T < \infty$。若环境在 $(\sigma, T, b^\dagger)$ 下呈现武器化信息，则存在 $b^\dagger \in \Delta(S)$ 和 $\delta > 0$，使得：

(i) 离散行动切换： 至少对于一名玩家 i，均衡行动对应$a_i(b;\sigma)$ 在 $b^\dagger$处导致失败风险率发生不连续变化：存在 $p > 0$ 使得

$$\begin{aligned} &For\ b < b^\dagger\ with\ \|b - b^\dagger\| < \delta: \quad \overset{\sigma}{\Pr}(s_{t+1} \in F \mid s_t, a(b;\sigma)) \geq p, \\ &For\ b > b^\dagger\ with\ \|b - b^\dagger\| < \delta: \quad \overset{\sigma}{\Pr}(s_{t+1} \in F \mid s_t, a(b;\sigma)) = 0. \end{aligned}$$

(ii) 吸收性失败： F 是吸收性的（$s_t \in F$ 意味着对所有 $k \geq 0$， $s_{t+k} \in F$ 成立），且对于所有 $s \in F$ 和 $a \in A$， $u_i(s, a) = 0$。

注释 2（噪声导致的脆弱性与精确度导致的脆弱性）。 在危机的全局博弈模型中，小噪声可能通过协调激励和高阶信念引发挤兑或攻击；其比较静态通常是在固定信息结构下，以均衡选择或多重性的形式来表述。这里，脆弱性通过另一种渠道发挥作用：吸收性失败会在生存价值中造成悬崖，使价值函数在信念上呈现局部凹性，从而逆转信息价值。由此产生的对不透明性的偏好，源自不可逆性所导致的几何结构，而非标准危机模型中的均衡多重性逻辑。

注释 3（悬崖的几何结构）。 吸收性失败会产生“悬崖式”的生存价值：对于将足够概率质量赋予安全状态的信念，存活概率接近 *1*；而对于跨越触发区域的信念，存活概率则会急剧下降。这种悬崖式几何结构意味着，在下降处附近，$V_T(\cdot;\ \sigma)$ 呈现局部凹性，从而逆转信息通常具有的正向价值。

命题1（局部凹性下的信息价值为负）。 固定（σ、T），并假设定义 *2* 成立。设 π 是均值为 b 的任意后验分布（即 $\int \mu \, d\pi$ （ μ） $= b$）。那么对所有 $b \in \mathcal{N}$，

$$\int V_T(\mu;\sigma)\, d\pi(\mu) \ \leq \ V_T(b;\sigma),$$

对于支撑于 $\mathcal{N}$ 上的非退化 π，该不等式为严格不等式。

*Ske*等等。 这是将詹森不等式应用于在 $\mathcal{N}$ 上严格凹的函数 $V_T(\cdot;\ \sigma)$. □

表1：机制比较

特征	全球博弈（Morris–Shin）	贝叶斯说服 (Kamenica–Gentzkow)	本文
信息作用	选择均衡	塑造行为	使存续成为可能
凹性来源	协调收益	外生偏好	内生（吸收性失败）
噪声效应	细化多重性	不适用	破坏可行性
最优精度	更高精度稳定	取决于曲率	较低精度必要性-sary
关键边际	阈值选择	后验分布	信息结构
机制	策略互补性	说服	不可逆性

引理 1（吸收性失败会产生局部凹性）。 考虑一个失败具有吸收性且延续价值为零的动态环境。假设存在一个信念区域，其中当信念跨越某个阈值时，均衡行动会从不诱发失败的行为离散地切换为诱发失败的行为。那么，对于任意有限时间范围 T，在相应信念阈值附近，生存价值 V_T（·；σ）局部凹。

概述。 吸收性失败会使延续收益发生不连续截断。当均衡行动在信念阈值处发生离散变化时，随着信念越过该阈值，存续至时间范围 T 的概率会急剧下降。这在 V_T中形成悬崖，从而意味着局部凹性：信念的均值保持扩散会将概率质量转移到失败区域，同时不会使安全一侧的存续概率超过 1。将这一论证形式化即可得到该结论。 □

备注4（无静态对应物）。 引理 *1* 中识别出的凹性，在没有吸收状态的静态贝叶斯劝说问题中并无对应物：它源于不可逆失败与信念触发行动之间的内生交互，而非收益的原始曲率。

假设 3（不透明性会产生后验的均值保持扩散）。 对于 $\lambda' \geq \lambda$，在 $\kappa^{\varepsilon,\,\lambda'}$ 下诱导的后验，是在 $\kappa^{\varepsilon,\,\lambda}$ 下诱导后验的均值保持收缩（等价地，在相同先验条件下，较高不透明性对应的后验分布，是较低不透明性下后验分布的均值保持扩散）。

备注5（凹化与最优不透明性）。 在贝叶斯劝服框架中（*Blackwell，1953*年；*Kamenica* 和 *Gentzkow，2011*年），选择信息结构的参与者实际上是在选择一个均值固定的后验分布。当 V_T（·；σ）局部为凹时，完全透明并非最优：最优政策通过使后验远离悬崖，对生存价值进行凹化。因此，策略性不透明是对生存函数进行凹化的最优方式。

充分原始条件（触发机制）。局部凹性的一个便捷充分条件是存在触发结构：在某个后验区域内，均衡行动会离散地切换为导致失败的行动（例如撤出、攻击、拒绝续借），因此后验的小幅变动可能导致失败风险率的大幅变化。在应用中，我们将这种触发机制称为“信息武器化”。

7 主要结果

定理 1（不透明性作为可存续性细化）。假设环境中存在武器化信息。给定任意有限期 $T < \infty$。那么，在所有 *Blackwell* 可比较的信息结构中，最大透明度会使有限期存续最小化，而任何足够不透明的信息结构在极限可存续性序中都弱支配透明度。因此，任何极限可存续均衡都必须以严格正的不透明性为支撑。

定理 2（动态博弈中的透明度诱发崩溃）。存在一些具有吸收性失败和内生不透明性的动态博弈，使得：*(i)* 在最大透明度下，不存在极限可存续均衡；但 *(ii)* 在足够高的不透明性下，至少存在一个极限可存续均衡。此外，对于先验和收益参数的开集而言，透明度下极限可行性的失败具有一般性。

定理 3（最优不透明性）。考虑如下扩展博弈：代理人共同选择 λ，随后在 $\kappa^{\varepsilon,\lambda}$ 下进行诱导动态博弈的均衡。假设 *(a)* 流收益弱非负，且至少有一名代理人在某些避免失败的历史中获得严格正的流收益；*(b)* 在最大透明度 $\lambda = \bar{\lambda}$ 下，每个均衡都不满足极限可行性，即存在 $T < \infty$ 和 $\delta > 0$，使得对于每个均衡 σ，

$$\liminf_{\varepsilon \downarrow 0} \Pr_{\sigma,\kappa^{\varepsilon,\lambda}}(\tau_F \le T) \ge \delta.$$

若存在 $\lambda > \underline{\lambda}$，以及在 $\lambda = \bar{\lambda}$ 下满足极限可行性的均衡 $\bar{\sigma}$，则在扩展博弈的任何均衡中，所选择的不透明性均满足 $\lambda^* > \underline{\lambda}$。

注释 6. 尽管不透明性在完全信息下能够提升配置效率，人们仍会选择不透明性。这种权衡完全由不可逆失败所驱动。

8 示例 性基准：吸收性失败下的协调

本节将按照全球博弈的思路，给出一个完整推导的基准模型，并加入吸收性违约和内生不透明性。该基准旨在明确展示计算过程，并以构造性方式说明定理 1–3。

8.1 模型设定

有两名玩家 $i \in \{1, 2\}$。隐藏基本面 $\theta \in \mathbb{R}$ 决定偿付能力。每一期，每名玩家选择 $a_i \in \{0, 1\}$，分别表示继续 $(a_i = 1)$或退出 $(a_i = 0)$。如果撤出导致系统跌破阈值 θ，就会发生吸收性违约。

玩家观察私人信号

$$x_i = \theta + \sqrt{\varepsilon}\,\eta_i, \qquad \eta_i \sim \mathcal{N}(0, 1),$$

其中 $\varepsilon > 0$ 表示消失噪声。不透明性体现为对信号方差的选择：λ 将噪声缩放至 $\sqrt{\varepsilon}\lambda$，其中 $\lambda \in [\underline{\lambda}, \bar{\lambda}]$。因此，较大的 λ 是对较小 λ进行的布莱克威尔劣化。

8.2 均衡刻画

阈值策略。 在对称单调均衡中，每个玩家 i 当且仅当其信号 x_i 低于阈值 $x^*(\varepsilon,\ \lambda)$时退出。这类均衡的存在性和唯一性遵循标准的全球博弈论证（Morris 和 Shin，2003年）。

阈值 $x^*(\varepsilon,\ \lambda)$ 由无差异条件决定：收到信号x^* 的玩家在继续和退出之间无差异。给定二元状态 $\theta \in \{0,1\}$ 和高斯信号$x_i = \theta + \sqrt{\varepsilon\lambda}\eta_i$，其中 $\eta_i \sim N(0,\ 1)$，后验信念为

$$\Pr(\theta = 1 \mid x_i) = \frac{q \cdot \phi\left(\frac{x_i - 1}{\sqrt{\varepsilon\lambda}}\right)}{q \cdot \phi\left(\frac{x_i - 1}{\sqrt{\varepsilon\lambda}}\right) + (1-q) \cdot \phi\left(\frac{x_i}{\sqrt{\varepsilon\lambda}}\right)}, \tag{2}$$

其中 ϕ 表示标准正态密度，$\Pr(\theta = 1)$ 中的 $q =$ 是先验。

在无差异阈值处，继续的期望收益等于退出的收益（标准化为零）：

$$\delta R \cdot \Pr(\theta = 1 \mid x_i = x^*) \cdot \Pr(\text{other player continues} \mid x_i = x^*) = 0. \tag{3}$$

在对称均衡中，当且仅当 $x_j \geq x^*$ 时，另一位玩家继续。已知$x_j = \theta + \sqrt{\varepsilon\lambda}\eta_j$与 $\eta_j \sim N(0,\ 1)$ 独立于 η_i，我们有

$$\Pr(x_j \geq x^* \mid x_i = x^*, \theta) = 1 - \Phi\left(\frac{x^* - \theta}{\sqrt{\varepsilon\lambda}}\right), \tag{4}$$

其中 Φ 是标准正态累积分布函数。

后验集中。 随着 $\varepsilon \downarrow 0$，后验信念逐渐集中：

- 若 $\theta = 1$：$\Pr(x_i \approx 1) \to 1$，因此后验收敛到 $\Pr(\theta = 1 \mid x_i) \to 1$。
- 若 $\theta = 0$：$\Pr(x_i \approx 0) \to 1$，因此后验收敛到 $\Pr(\theta = 1 \mid x_i) \to 0$。

透明度下的失败。 在最小信息模糊化 $\lambda = \lambda$ 下，假设真实状态为 $\theta = 1^-$，且基本面任意接近失败阈值 θ。随着 $\varepsilon \downarrow 0$，x_i 的分布集中在 1 附近。然而，由于高斯噪声 η_i 的尾部影响，一个或两个玩家收到信号 $x_i < x^*(\varepsilon,\ \lambda)$的概率有一个严格大于零的下界。

具体而言，对于任意 $\varepsilon > 0$，

$$\Pr(x_i < x^* \mid \theta = 1) = \Phi\left(\frac{x^* - 1}{\sqrt{\varepsilon\lambda}}\right) > 0. \tag{5}$$

由于失败由单方面退出触发（至少一名玩家选择 W），在 $t = 1$ 时失败的概率为

$$\Pr(\tau_F = 1 \mid \theta = 1) = 1 - \left[1 - \Phi\left(\frac{x^* - 1}{\sqrt{\varepsilon\lambda}}\right)\right]^2. \tag{6}$$

当 $\varepsilon \downarrow 0$ 时，均衡阈值 $x^*(\varepsilon,\ \lambda)$ 收敛到一个有限极限 $x^*_\infty < 1$（由协调博弈的优势区域决定）。因此，

$$\liminf_{\varepsilon \downarrow 0} \overset{\sigma,\kappa^{\varepsilon,\lambda}}{\Pr}(\tau_F \leq 1) \geq 1 - \left[1 - \Phi\left(\frac{x^*_\infty - 1}{\sqrt{0^+}}\right)\right]^2 = 1 - 0 = 1 > 0, \tag{7}$$

其中，该极限利用了 $\Phi(z)$ 在 $z \to +\infty$ 时 $\to 1$ 。因此，完全透明条件下的均衡行为不符合极限可存续性准则。

不透明性下的恢复。 现在考虑最大不透明性 $\lambda = \bar{\lambda}$ ，且 $\lambda \to \infty$。信号变得毫无信息：对于任意实现值 x_i，后验都收敛于先验，

$$\Pr(\theta = 1 \mid x_i) \to q \quad \text{for all } x_i, \tag{8}$$

当 $\lambda \to \infty$ （噪声方差 $\varepsilon\lambda \to \infty$ 支配信号）时。

鉴于继续会带来期望收益 $qR > 0$ （根据参数限制 $qR > 0$），而退出的收益为零，两位玩家无论信号实现为何，都会最优地选择 C。两位玩家始终继续的策略组合构成一个均衡。由于退出从未发生，失败必然得以避免：

$$\Pr^{\sigma,\kappa^{\varepsilon,\bar{\lambda}}}(\tau_F \le T) = 0 \quad \text{for all } T < \infty \text{ and all } \varepsilon > 0. \tag{9}$$

因此，该均衡具有极限可行性。

行为解释。 关键在于，不透明性是一种有意的选择。代理人预见到，透明性会通过信念驱动的退出，以正概率诱发崩溃，因此会最优地选择更高的不透明性$\lambda > \bar{\lambda}$ ，以抑制触发事件的实现并维持极限可行性。在存续条件下，不透明性并不会提高配置效率；它改变了信息环境，使存续本身变得可行。

9 将殖民掠夺与起义风险重构为一个说服问题

本节将殖民环境中的策略性不透明问题，重新表述为一个存在吸收性失败的信息设计问题。这个例子明确说明了起义风险如何导致生存价值呈现局部凹性，以及为什么当信息结构由委托人控制时，不透明性是最优选择：它展示了武器化信息的一般逻辑如何适用于由委托人控制的信息环境，在这种环境中，不透明性是单方面选择的，而非共同选择的。所有结果都源自第 6 节和第 7 节所揭示的同一几何结构。

9.1 环境

一个殖民强国（宗主国）在离散时间内统治一个殖民地。殖民地潜在的政治稳定性由状态 $s_t \in [0,1]$概括。存在一个崩溃阈值 $s > 0$ ，若$s_t \le s$，便会发生起义，殖民统治也将不可逆转地丧失。失败是吸收性的。

宗主国是一个统一的决策者，同时控制 (i) 财政政策和 (ii) 用于评估政治稳定性的信息结构。信息获取不存在私下且无成本的偏离。

每个时期，宗主国都会选择一个掠夺水平 $x_t \in [0, \bar{x}]$ （征税、征用、强迫劳动）。资源攫取会带来当期收入，但会增加起义风险率。殖民地本身不会采取策略性行动。

状态动态如下：

$$s_{t+1} = s_t - \alpha x_t + \varepsilon_t,$$

其中 $\alpha > 0$ 且 ε_t 是一个方差较小的独立同分布冲击。一旦 $s_t \le s$, 该过程便会被吸收 d.

宗主国无法直接观测到 s_t 。相反，它会事前承诺采用一种观测通道

$$y_t = s_t + \eta_t, \qquad \eta_t \sim \mathcal{N}(0, \sigma^2(\lambda)),$$

其中， λ 表示不透明性： λ 越高，信号越嘈杂。 λ 的选择是公开的，并在博弈开始前固定。

与基准模型的对应关系。状态 s_t 对应第3节中的收益相关状态 $s_t \in S$ ，起义是吸收失败集 F，资源攫取 x_t 是行动 a_t，而观测通道 $\sigma^2(\lambda)$是对不透明性的选择。信念 b_t 是最优政策的充分统计量。

9.2 收益与目标

宗主国从资源攫取中获得的每期收益为 $\pi(x_t)$，其中 $\pi' > 0$ 且 $\pi'' \leq 0$。起义后没有延续收益。

对于有限时间范围 T，相关的价值函数是按存续概率加权的资源攫取收益

$$V_T(b) \equiv \mathbb{E}\left[\sum_{t=0}^{T} \pi(x_t)\mathbf{1}\{t < \tau_F\} \;\middle|\; b_0 = b\right],$$

其中， b_t 表示由所选信息结构诱导的关于 s_t 的后验信念， τ_F 为起义发生的（吸收）时点。

9.3 最优政策与悬崖几何结构

固定给定信息结构下的最优资源攫取政策 $x(b)$。由于起义具有吸收性，且资源攫取会提高越过 s 的风险，最优政策呈现触发结构：存在 $b^\dagger$ ，当信念越过 $b^\dagger$时，资源攫取会从正值离散地切换为零。这正是定义 2 中的触发条件。

在这一政策下， $V_T(b)$的生存部分呈现悬崖特征：当信念稳稳高于 $b^\dagger$时，存续至时间范围 T 几乎是确定的；而当信念越过触发区域时，即使只激进攫取一个时期，也会以很高概率引发起义。因此，在 $b^\dagger$附近，生存价值关于信念局部凹。

9.4 不透明性作为最优凹化

在完全透明（ λ较低）的情况下，随着信号噪声趋于消失，后验信念会迅速集中。当真实稳定性 s_t 接近起义阈值时，即使微小的信息波动也会产生越过触发区域的后验实现，从而触发激进攫取，并在有限期内以正概率导致起义。

通过提高不透明性（更高的 λ），宗主国诱导后验信念发生均值保持收缩。这使信念远离悬崖区域，消除进入触发失败的后验的偏离，并恢复极限可行性。

不同于标准的劝服问题（在这类问题中，信息设计会在参与的条件下塑造行为），这里的不透明性通过阻止触发失败的信念实现，使参与（存续）本身成为可行的。宗主国并不是为了改变存续条件下的行为而操纵信念；相反，它改变信息结构，使存续从一开始就成为可能。

直观地说，当生存价值关于信念局部呈凹性时，委托人更偏好信念的期望值，而不是信念的实现值：不透明性优于透明度，因为它抑制了跌过生存悬崖的信念实现。

10条结论

本文表明，当失败不可逆时，信息价值可能发生逆转。这种逆转源于：在不可逆失败下维持活动的均衡通常运行在可存续边界上，因而没有抵御崩溃的内部缓冲。在具有吸收性失败的动态环境中，更高的信息精度可能通过放大触发崩溃的信念驱动的行动，增加灾难性结果发生的可能性。不可逆性在内生上导致生存价值关于信念呈现局部凹性，从而打破了信息越多总是越有利这一标准假设。

殖民掠夺的例子表明，无论不透明性是由各方共同选择（第8节），还是由委托人单方面选择（第9节），透明度与可行性之间都会产生同样的不相容性，这凸显出该机制是结构性的，而非战略性的。

其核心含义体现在行为层面。代理人并不会通过扭曲信念或采取非理性行动来应对脆弱性；他们会调整允许信息在多大程度上影响自身行动。当信息被武器化时，策略性不透明——通过 Blackwell 加扰建模为对信息结构的选择——便成为一种理性的生存应对。不透明性并非为了在存续条件下提高配置效率，而是为了使存续本身成为可能。

从概念上说，本文的贡献在于识别出一类广泛且具有经济核心意义的动态环境：在这类环境中，价值函数的凹性并非偏好的原始特征，而是不可逆失败的内生结果。一旦认识到这一几何结构，信息设计的逻辑便随之成为一种推论，而非假设：不透明性之所以最优，是因为它通过抑制失败阈值附近的信念分散，使生存价值函数凹化。

本文的分析提出了一个更为一般性的启示。只要经济系统表现出吸收性失败和信念触发行动——例如违约、挤兑、起义或制度崩溃——信息结构本身就会成为一项一阶策略空间。要理解透明度何时能够稳定行为、何时反而会使行为失稳，就必须考虑不可逆性所诱导的几何结构。这一视角为研究信息设计、脆弱性以及失败无法逆转环境中的政策开辟了新的方向。分析的一个含义是一项新的不可能性结果：在具有吸收性失败的动态环境中，一旦信念本身能够触发不可逆崩溃，透明度与可行性通常便不相容。

参考文献

Acemoglu, D., Johnson, S., and Robinson, J. A. (2005年). 制度是长期增长的根本原因. 《经济增长手册》, 1:385–472. Altman, E. (1999年). 《受约束马尔可夫决策过程》. Chapman and Hall/CRC. Angeletos, G.-M., Hellwig, C., and Pavan, A. (2006). 全球博弈中的信号传递：协调与政策陷阱. 《政治经济学杂志》, 114(3):452–484.

Angeletos, G.-M. and Werning, I. (2006). 危机与价格：信息聚合、多重性与波动性. 《美国经济评论论文与会议录》, 96(2):172–177. Aubin, J.-P. (1991年). 可行性理论. Birkh¨auser. Blackwell, D. (1953年). 实验的等价比较. 《数学统计年刊》, 24(2):265–272. Carlsson, H. and van Damme, E. (1999年). 协调与公共信息的作用. 《博弈与经济行为》, 26(1):1–19. Epstein, L. G. and Schneider, M. (2003年). 递归多先验. 《经济理论杂志》, 113(1):1–31. Esponda, I. and Pouzo, D. (2016). Berk–nash均衡：一种用于建模具有设定错误模型的代理人的框架. 《计量经济学》, 84(3):1093–1130. Grossman, S. J. (1980年). 企业的信息披露. 《金融学杂志》, 35(2):323–337. Hansen, L. P. and Sargent, T. J. (2008年). 稳健性. 普林斯顿大学出版社. Kamenica, E. and Gentzkow, M. (2011年). 贝叶斯说服. 《美国经济评论》, 101(6):2590–2615. Kirk, N. H. (2025). 动态博弈的可持续剥削均衡. $arXiv$. arXiv:2512.07629[econ.TH],审稿中. Kozlowski, J., Veldkamp, L., and Venkateswaran, V. (2020). 撬动经济的尾部：信念驱动的商业周期与持续性停滞. 《政治经济学杂志》, 128(8):2839–2879. Milgrom, P. (1981年). 好消息与坏消息：表示定理及其应用. 《贝尔经济学杂志》, 12(2):380–391. Morris, S. and Shin, H. S. (1998年). 协调与货币危机. 《欧洲经济学会杂志》, 6(1):1–23. Morris, S. and Shin, H. S. (2003年). 全球博弈：理论与应用. 《经济学与计量经济学进展》, 1:56–114.

A 附录：证明

A.1 从布莱克威尔劣化到后验的凸序

主要结果采用如下比较静态结论：信息量更大的信号会产生在凸序上更分散的后验。在一般的状态空间中，这可以用鞅耦合来表述；为使论证清晰明了，我们针对二元状态陈述该结果。这已足以支持下文的构造，并且体现了说服几何。

引理 2 (布莱克威尔序蕴含二元情形下后验的凸序扩散**)**。 设$S = \{0, 1\}$，并令 $b \in [0, 1]$ 表示状态1的后验概率。固定一个先验 $b_0 \in (0,\ 1)$。设κ 和 κ' 为两个信号结构，且 κ' 是 κ的布莱克威尔劣化。令 B 和 B' 分别表示在 κ 和 κ' 下诱导出的后验随机变量（给定先验 b_0）。那么：

(i) $\mathbb{E}[B] = \mathbb{E}[B'] = b_0$;

(ii) B 是 B' 在凸序上的均值保持扩散，即对每个凸函数 $\varphi:\ \ [0, 1] \to \mathbb{R}$,

$$\mathbb{E}[\varphi(B)]\ \geq\ \mathbb{E}[\varphi(B')].$$

等价地，对于每个凹函数 $\psi:\ \ [0, 1] \to \mathbb{R}$, $\mathbb{E}[\psi(B)] \leq \mathbb{E}[\psi(B')]$。

证明。 根据贝叶斯可行性，后验的平均值等于先验，因此 $\mathbb{E}[B] = \mathbb{E}[B'] = b_0$。由于 κ' 是 κ的布莱克威尔劣化，因此存在一个随机核 G，使得在 κ' 下的信号可通过将 κ 下的信号经过 G得到。这会在后验之间诱导出一个鞅耦合：$B' = \mathbb{E}[B \mid \mathcal{F}']$，其中较粗的信息 σ-域 $\mathcal{F}'$由劣化信号生成。由詹森不等式可得$\mathbb{E}[\varphi(B')] = \mathbb{E}[\varphi(\mathbb{E}[B \mid \mathcal{F}'])] \leq \mathbb{E}[\mathbb{E}[\varphi(B) \mid \mathcal{F}']] = \mathbb{E}[\varphi(B)]$，对每个凸函数 φ均成立，从而确立凸序。 □

注 7。 引理 *2* 是标准的信息设计比较：信息量更大的实验会产生分布更分散但均值相同的后验分布（*Blackwell，1953*年；*Kamenica* 和 *Gentzkow，2011*年）。假设 *3* 可被解释为在一般状态情形下施加了这种凸序单调性。

A.2 定理 1 的证明

证明。 固定 $T < \infty$ 和一个策略组合 σ。假设环境在 $(\sigma, T, b^\dagger)$ 处呈现武器化信息（定义 2）。

步骤 1：局部凹性。 根据引理 1，定义 2 意味着 $V_T(\cdot;\ \sigma)$ 在 $b^\dagger$ 的某个邻域 N 内是局部凹的。

步骤 2：詹森不等式。 固定 $b \in N$。对于任意均值为 b 的后验分布 π（即 $\int \mu\, d\pi(\mu) = b$），命题1意味着

$$\int V_T(\mu; \sigma)\, d\pi(\mu) \leq V_T(b; \sigma), \tag{10}$$

对于支撑在 N 中的非退化 π，不等式严格成立。

步骤 3：不透明度比较。 在假设 3 下，将不透明性从 λ 提高到 $\lambda' > \lambda$ 会导致后验分布发生均值保持收缩。具体而言，令 π_λ 和 $\pi_{\lambda'}$

分别表示在相同先验条件下，由 $\kappa^{\varepsilon,\lambda}$ 和 $\kappa^{\varepsilon,\lambda'}$ 所诱导的后验分布。那么， $\pi_{\lambda'}$ 在凸序下是 π_λ 的均值保持收缩。根据步骤 2，这意味着

$$\mathbb{E}_{\pi_{\lambda'}}[V_T(b;\sigma)] \geq \mathbb{E}_{\pi_\lambda}[V_T(b;\sigma)]. \tag{11}$$

因此，预期的有限期存续概率随 λ弱递增。

步骤 4：极限可行性。 根据定义，回顾一下， $V_T(b;\sigma) = 1 - \Pr^\sigma(\tau_F \leq T \mid b_0 = b)$。增加 λ会使 $\mathbb{E}[V_T(b_T;\sigma)]$,弱递增，这等价于对于每个固定的 T， $\Pr^\sigma(\tau_F \leq T)$弱递减。

因此，如果 σ 在 λ' 下是极限可行的（意味着 $\lim_{\varepsilon\downarrow 0}\Pr^{\sigma}_{\kappa^{\varepsilon,\lambda'}}(\tau_F \leq T) = 0$ 对所有 T均成立），那么在任何 $\lambda > \lambda'$下它仍然是极限可行的。因此，不透明性的增加会弱扩展极限可存续均衡的集合。

步骤 5：不透明性的必要性。 假设最大透明度$\lambda = \underline{\lambda}$ 使存活概率最小。如果对于某个 T，在 $\lambda = \underline{\lambda}$下有 $\liminf_{\varepsilon\downarrow 0}\Pr^\sigma(\tau_F \leq T) > 0$，那么 σ在最大透明度下不可能是极限可行的。根据步骤 4，任何极限可存续均衡都必须满足 $\lambda^* > \underline{\lambda}$，即要求严格为正的不透明性。 □

A.3 定理 2 的证明：显式构造

证明。 我们构造一个双人动态博弈，其中策略互动至关重要，而透明度会摧毁均衡的可行性。

环境。 有两位玩家 $i \in \{1,2\}$。收益相关状态为 $\theta \in \{0,1\}$，其中$\theta = 1$ 表示脆弱但有利可图的状态， $\theta = 0$ 表示安全状态。先验信念为 $\Pr(\theta = 1) = q \in (0,\ 1)$。失败具有吸收性。

时间是离散的。在每个日期 t，若失败尚未发生，每位玩家同时选择一个行动 $a_i^t \in \{C, W\}$，解释为继续 或退出。只要至少一位玩家选择 W，就会立即发生失败。若发生失败，此后收益为零。

收益。 如果两位玩家在日期 t 选择 C ，每位玩家获得流收益

$$u_i(\theta) = \begin{cases} R > 0 & \text{if } \theta = 1, \\ 0 & \text{if } \theta = 0. \end{cases}$$

如果在任何日期发生失败，延续收益为零。玩家以因子 $\delta \in (0,\ 1)$进行贴现。

信息与不透明性。 在每个日期，玩家都会观察到私人信号

$$x_i = \theta + \sqrt{\varepsilon\lambda}\,\eta_i, \qquad \eta_i \sim N(0,1),$$

其中 $\varepsilon \downarrow 0$ 且 $\lambda \in [\underline{\lambda}, \lambda]$ 用于刻画不透明性：较高的 λ 对应于对信号进行布莱克威尔劣化。信号满足单调似然比性质。

均衡结构。 标准论证表明，任何对称均衡都可由阈值策略刻画：存在 $c(\varepsilon, \lambda)$，使得每个玩家当且仅当其后验信念 $\Pr(\theta = 1 \mid x_i)$ 超过 $c(\varepsilon, \lambda)$ 时选择 C 。

透明度下的失败。固定参数，使得：

$$qR > 0 \qquad \text{and} \qquad qR < \delta R.$$

在最大透明度 $\lambda = \underline{\lambda}$ 下，后验信念会集中于 $\varepsilon \downarrow 0$。以 $\theta = 1$ 为条件，在任何有限时间范围内，至少有一名玩家以正概率收到一个信号，促使其后验低于均衡临界值。由于行动具有策略互补性，且失败由单方面退出触发，任何均衡临界值都必须在信念足够悲观时规定 W。因此，对于任何对称均衡策略组合 σ 和任何有限的 $T < \infty$，

$$\liminf_{\varepsilon \downarrow 0} \Pr_{\sigma, \kappa_{\varepsilon, \underline{\lambda}}} (\tau_F \leq T) > 0.$$

因此 最大透明度下不存在满足极限可存续性准则的均衡 .

不透明性下的恢复。 现在考虑足够高的不透明性 $\lambda = \bar{\lambda}$。当 λ 足够大时，信号将变得毫无信息，所有玩家在所有历史下的后验信念都与先验 q 相同。根据构造，在先验信念下，继续是最佳回应，因为 $qR > 0$，而失败会带来零收益。因此，双方始终选择 C 的策略组合构成一个均衡。由于从不发生撤出，在任何时间范围内，失败都以概率 1 得以避免：

$$\Pr_{\sigma, \kappa_{\varepsilon, \bar{\lambda}}} (\tau_F \leq T) = 0 \quad \text{for all } T < \infty \text{ and all } \varepsilon.$$

因此，该均衡具有极限可行性。

一般性。 现在我们证明，透明度下极限可行性的失败在参数的一个开集上成立。固定 $R > 0$ 和 $\delta \in (0, 1)$。上述构造要求 $qR > 0$ 和 $qR < \delta R$，当且仅当 $0 < q < \delta$ 时，这一要求成立。

对于开集 Ω 中任意的 $(q, R, \delta) := \{(q, R, \delta) : q \in (0, \delta)\ R > 0,\ \delta \in (0, 1)\}$，均衡临界值 $c(\varepsilon, \underline{\lambda})$ 满足刻画对称单调均衡的无差异条件。根据收益参数中最佳回应对应的连续性（例如参见 Fudenberg 和 Tirole，1991年，定理 1.2），存在 $\varepsilon_0(q, R, \delta) > 0$，使得对所有 $\varepsilon < \varepsilon_0$，最大透明度下的任何对称均衡都满足

$$\Pr_{\sigma, \kappa^{\varepsilon, \underline{\lambda}}} (\tau_F \leq T) \geq \delta/2 > 0 \tag{12}$$

对于某个有限的 $T < \infty$（具体而言，$T = 1$ 已足够）。

由于 Ω 在 $[0,1] \times \mathbb{R}_+ \times (0, 1)$ 上的乘积拓扑中是开集，这表明在透明度下，极限可行性的失败具有一般性（即在先验信念和收益参数空间的一个正测度开集上成立）。

我们得出结论：在具有吸收性失败的动态博弈中，透明度可能破坏均衡可存续性，而足够的不透明性则能恢复均衡的存在性。 □

A.4 定理 3 的证明：必要性

*定理 3 的证明。*考虑这样一个扩展博弈：代理人先选择 λ，然后在诱导博弈中进行均衡博弈。为得出矛盾，假设扩展博弈存在一个满足 $\lambda^* = \underline{\lambda}$ 的均衡。

根据定理 3 中的假设 (b)，在 $\lambda = \underline{\lambda}$ 每个均衡 σ 都不具备极限可行性：存在 $T < \infty$ 和 $\delta > 0$ 使得

$$\liminf_{\varepsilon \downarrow 0} \Pr_{\sigma, \kappa^{\varepsilon, \underline{\lambda}}} (\tau_F \leq T) \geq \delta.$$

因此，在消失噪声极限下，由 (σ^*, λ) 引致的延续收益，在时间范围 T 上至少损失了存续加权流收益的 δ 比例。

现在固定$\lambda > \lambda$ 以及在 λ下的一个极限可存续均衡 $\bar{\sigma}$。根据假设 (a)，某个代理人在某些无失败历史上获得严格为正的流收益。由于失败具有吸收性，且延续收益为零，因此在时间范围 T 内，策略组合（$\bar{\sigma}$，λ）下的期望收益严格高于失败概率至少为 δ的任何策略组合；当 ε足够小时，这一点始终成立。因此，至少有一名代理人可以通过将选择λ 改为选择 λ（然后采取 $\bar{\sigma}$）进行有利可图的偏离，这与 $\lambda^* = \lambda$ 能够在扩展博弈均衡中出现相矛盾。

因此，扩展博弈的任何均衡都必须满足 $\lambda^* > \lambda$。 □

A.5 引理 1 的证明（吸收性失败产生局部凹性）

引理 *1* 的证明。我们在标准的“触发”结构下给出一个完全形式化的论证，使正文中的启发式说明更加严谨。关键步骤如下：(i) 吸收性失败与离散行动切换意味着单步生存函数存在向下折点；(ii) 动态生存价值满足贝尔曼递归，且该递归在局部保持凹性，因此折点会传播至任何有限期。

步骤 0（精确的触发条件表述）。为便于分析透明度，我们将讨论限定为典型的一维信念情形；对于局部结论，这样做不失一般性，因为可以沿着线段验证 $\Delta(S)$ 上的局部凹性。[2] 令相关的信念坐标为 $b \in (0,1)$，将其解释为处于“安全”状态的后验概率。固定一个策略组合 σ，并考虑由此诱导的（共同）后验过程 $\{b_t\}_{t\geq 0}$。

假设在某个 $b^\dagger \in (0,1)$ 处满足以下触发性质：存在邻域$U^- = (b^\dagger - \eta, b^\dagger)$ 和 $U^+ = (b^\dagger, b^\dagger + \eta)$，以及可测风险函数 $h^-, h^+ : (b^\dagger - \eta, b^\dagger + \eta) \to [0,1]$，使得在 σ 下的单步失败概率满足

$$\Pr_{\sigma}(\tau_F = 1 \mid b_0 = b) = \begin{cases} h^-(b) & \text{if } b < b^\dagger, \\ h^+(b) & \text{if } b > b^\dagger, \end{cases} \qquad b \in (b^\dagger - \eta, b^\dagger + \eta), \tag{13}$$

并具有以下性质：

(H1)（触发点处边际风险的不连续性） h^- 和 h^+ 在各自定义域上连续，$\lim_{b\uparrow b^\dagger} h^-(b) = \lim_{b\downarrow b^\dagger} h^+(b) =: \bar{h}$，但单侧导数满足

$$\lim_{b\uparrow b^\dagger} (h^-)'(b) \;>\; \lim_{b\downarrow b^\dagger} (h^+)'(b). \tag{14}$$

(H2)（局部正则性） h^-和 h^+分别在 U^-和 U^+上二阶连续可微。

解释。 定义 2 中的离散行动切换表现为 $b^\dagger$处的斜率下降：在 $b^\dagger$以下，均衡行动会诱发失败（失败风险对信念的敏感度高）；而在$b^\dagger$ 以上，行动不会诱发失败（敏感度低）。吸收性失败体现为失败是一个吸收事件，因此一步失败风险会直接截断延续生存。

[2]函数 V 在凸集上为凹函数，当且仅当其在每条线段上的限制均为凹函数。

步骤 1（O　一步生存在局部上是凹的）。 定义 σ下的一步生存价值:

$$V_1(b;\sigma) \equiv 1 - \Pr_{\sigma}(\tau_F \le 1 \mid b_0 = b) = 1 - \Pr_{\sigma}(\tau_F = 1 \mid b_0 = b).$$

根据 (13)，在 $(b^\dagger - \eta, b^\dagger + \eta)$ 上，我们可以写成

$$V_1(b;\sigma) = \begin{cases} 1 - h^-(b) & \text{if } b < b^\dagger, \\ 1 - h^+(b) & \text{if } b > b^\dagger. \end{cases}$$

因此， V_1 在 $b^\dagger$ 处连续，并且在其两侧分别二阶连续可微。注意，(14) 等价于 V_1 中存在一个向下折点:

$$\lim_{b \uparrow b^\dagger} (V_1)'(b;\sigma) = -\lim_{b \uparrow b^\dagger} (h^-)'(b) \;<\; -\lim_{b \downarrow b^\dagger} (h^+)'(b) = \lim_{b \downarrow b^\dagger} (V_1)'(b;\sigma).$$

因此， V_1 的导数在 $b^\dagger$ 处向上跳跃（左侧斜率小于右侧斜率），这正是连续分段-C^1 函数具有凹性的局部特征。

形式上，定义 $N := (b^\dagger - \eta, b^\dagger + \eta)$，并令 $b_1 < b^\dagger < b_2$ 属于 N。对于任意满足 $b_\alpha := \alpha b_1 + (1-\alpha) b_2$ 的 $\alpha \in (0,1)$， V_1 在 N 上凹等价于

$$V_1(b_\alpha;\sigma) \;\ge\; \alpha V_1(b_1;\sigma) + (1-\alpha) V_1(b_2;\sigma). \tag{15}$$

由于 V_1 在两侧分别为 C^2，且具有上述折点，(15) 在 $b^\dagger$ 周围一个可能更小的邻域 $N_1 \subseteq N$ 内成立。[3] 因此， $V_1(\cdot;\sigma)$ 在 $b^\dagger$ 的某个邻域 N_1 内是凹函数。

步骤 2（有限期存续满足保持凹性的递推）。 对于 $T \ge 1$，回顾

$$V_T(b;\sigma) = 1 - \Pr_{\sigma}(\tau_F \le T \mid b_0 = b) = \Pr_{\sigma}(\tau_F > T \mid b_0 = b).$$

通过对存续至第 1 期进行条件化，并利用失败具有吸收性， V_T 满足贝尔曼递归

$$V_T(b;\sigma) = \underbrace{V_1(b;\sigma)}_{\text{survive period 1}} \cdot \underbrace{\mathbb{E}_\sigma[V_{T-1}(b_1;\sigma) \mid b_0 = b,\ \tau_F > 1]}_{\text{survive periods 2,...,}T \text{ given survival at 1}}. \tag{16}$$

定义延续算子 $\mathcal{T}$如下

$$(\mathcal{T}W)(b) := \mathbb{E}_\sigma[W(b_1) \mid b_0 = b,\ \tau_F > 1].$$

由于后验过程是通过贝叶斯更新生成的，给定$b_0 = b$ 时 b_1 的条件分布符合贝叶斯可行性:
$\mathbb{E}[b_1 \mid b_0 = b] = b$ （鞅性质），因此 $\mathcal{T}$ 保持凹性：若 W在邻域 N'上是凹函数，则（$\mathcal{T}W$）在 N'上也是凹函数。[4]因此，若V_{T-1}（$\cdot$; σ）在某个邻域上是凹函数，那么 $\mathcal{T}V_{T-1}$也如此。

---

[3]验证这一点的一种标准方法是使用凹性的支撑线刻画：折点意味着在 $b^\dagger$ 处存在一条支撑线，其斜率处于区间 $\left[\lim_{b\uparrow b^\dagger}(V_1)'(b),\ \lim_{b\downarrow b^\dagger}(V_1)'(b)\right]$ 内，而两侧的 C^2-正则性确保图像在局部位于该直线下方。[4]这就是詹森不等式：对于凹函数 W， $W(b) = W(\mathbb{E}[b_1]) \ge \mathbb{E}[W(b_1)],$，沿先验分布的混合使用同样的论证，即可得到该映射的凹性。

步骤 3（一致界传播）。我们用归纳法证明，存在 $\bar{\eta} > 0$，使得对所有 $T \geq 1$，$V_T(\cdot;\ \sigma)$ 在 $\bar{N}$ 上都是凹函数：$=(b^\dagger - \bar{\eta},\ b^\dagger + \bar{\eta})$。

基本情形 $T = 1$**：** 根据步骤 1，V_1 在某个 $\eta > 0$ 上是凹函数：$N_1 = (b^\dagger - \eta, b^\dagger + \eta)$.

关键观察： $b^\dagger$ 处的折点幅度由 σ 中编码的均衡行动切换决定，而非由时间范围 T 决定。具体而言，导数跳跃满足

$$\kappa := -\left[\lim_{b \uparrow b^\dagger}(h^-)'(b) - \lim_{b \downarrow b^\dagger}(h^+)'(b)\right] > 0, \tag{17}$$

其中，κ 是 σ 以及原始要素 (Q, F, u_i) 的一个性质。

一致曲率界： 由于生存概率满足 $0 \leq V_T(b;\ \sigma) \leq 1$，且 $0 \leq (\mathcal{T}V_{T-1})(b;\ \sigma) \leq 1$，并且 V_1 和 $\mathcal{T}V_{T-1}$ 在 $b^\dagger$ 的两侧均一致有界（根据假设 1），因此二阶导数 V_1'' 和 $(\mathcal{T}V_{T-1})''$ 在 $(b^\dagger - \eta, b^\dagger + \eta)$ 的任意紧子集上一致有界。

乘积凹性： 对于 $b \neq b^\dagger$，二阶导数满足

$$V_T''(b) = V_1''(b) \cdot (\mathcal{T}V_{T-1})(b) + 2V_1'(b) \cdot (\mathcal{T}V_{T-1})'(b) + V_1(b) \cdot (\mathcal{T}V_{T-1})''(b). \tag{18}$$

在 $b^\dagger$ 附近，凹性的主导贡献来自 V_1 中的折点。V_T 的单侧导数满足

$$\lim_{b \uparrow b^\dagger} V_T'(b;\sigma) - \lim_{b \downarrow b^\dagger} V_T'(b;\sigma) = -\kappa \cdot (\mathcal{T}V_{T-1})(b^\dagger;\sigma) + o(1), \tag{19}$$

其中，当邻域缩小时，$o(1) \to 0$。由于当 V_{T-1} 在 $b^\dagger$ 附近为凹函数时（存活概率严格为正），$(\mathcal{T}V_{T-1})(b^\dagger) > 0$，因此该折点得以保留，且其幅度有正的下界。

归纳步骤： 取 $\bar{\eta} = \eta/2$。由归纳假设，V_{T-1} 在 N 上为凹函数。于是，根据步骤 2，$\mathcal{T}V_{T-1}$ 在 N 上为凹函数。乘积 $V_T = V_1 \cdot \mathcal{T}V_{T-1}$ 从 V_1 继承向下折点，且其幅度在 N 上有下界 $\kappa/2$。根据凹性的支撑线刻画（脚注 5），这保证 V_T 在 $\bar{N}$ 上为凹函数。[5]

因此，对于所有 $T \geq 1$，$V_T(\cdot;\ \sigma)$ 在统一邻域 N 上为凹函数，其中 N 与 T 无关。

□

[5]区间上的连续函数在某个邻域内是凹函数，当且仅当它在该邻域的每个内点处都有一条支撑线。