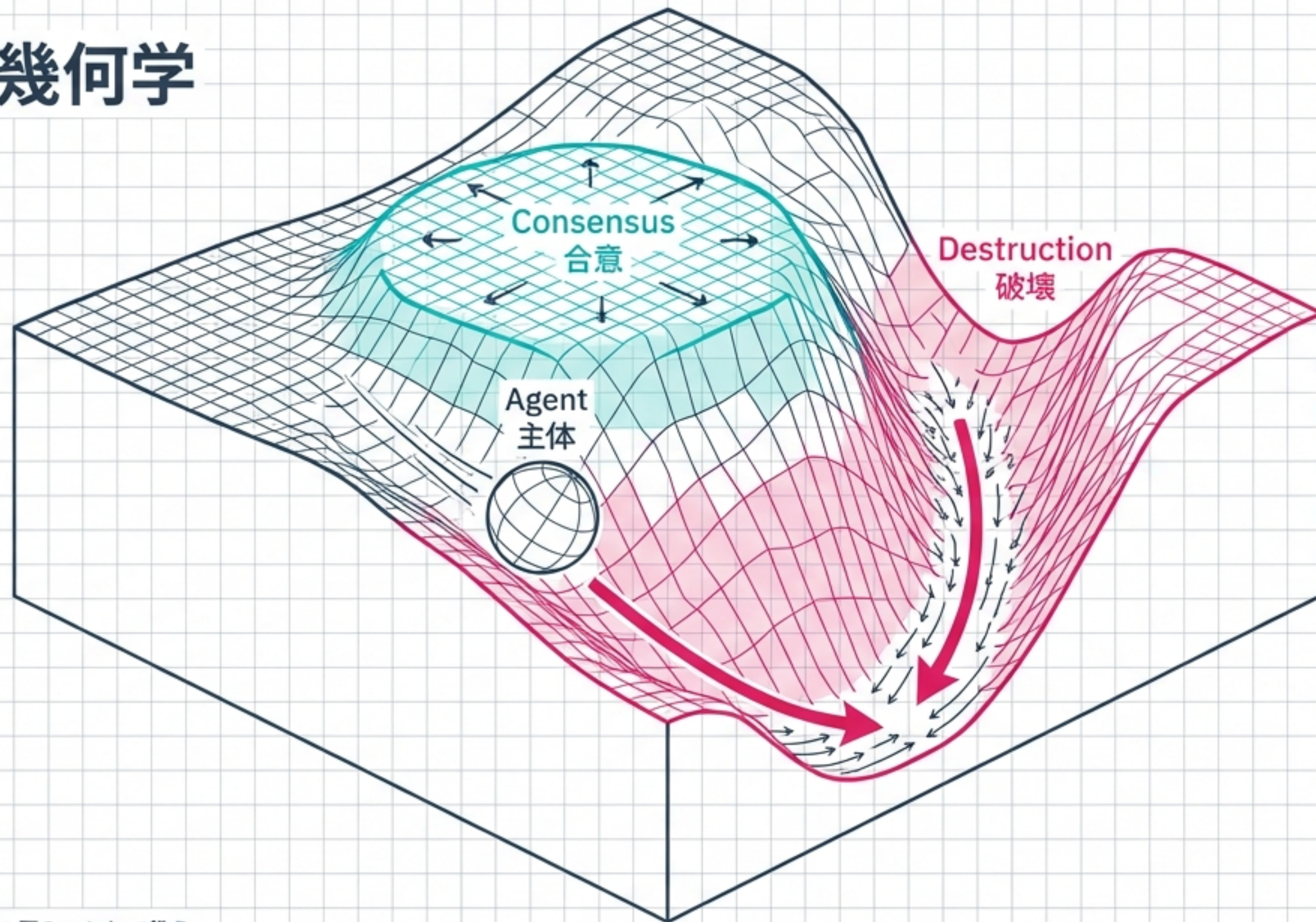


合意形成の物理

第11論：誘因の幾何学

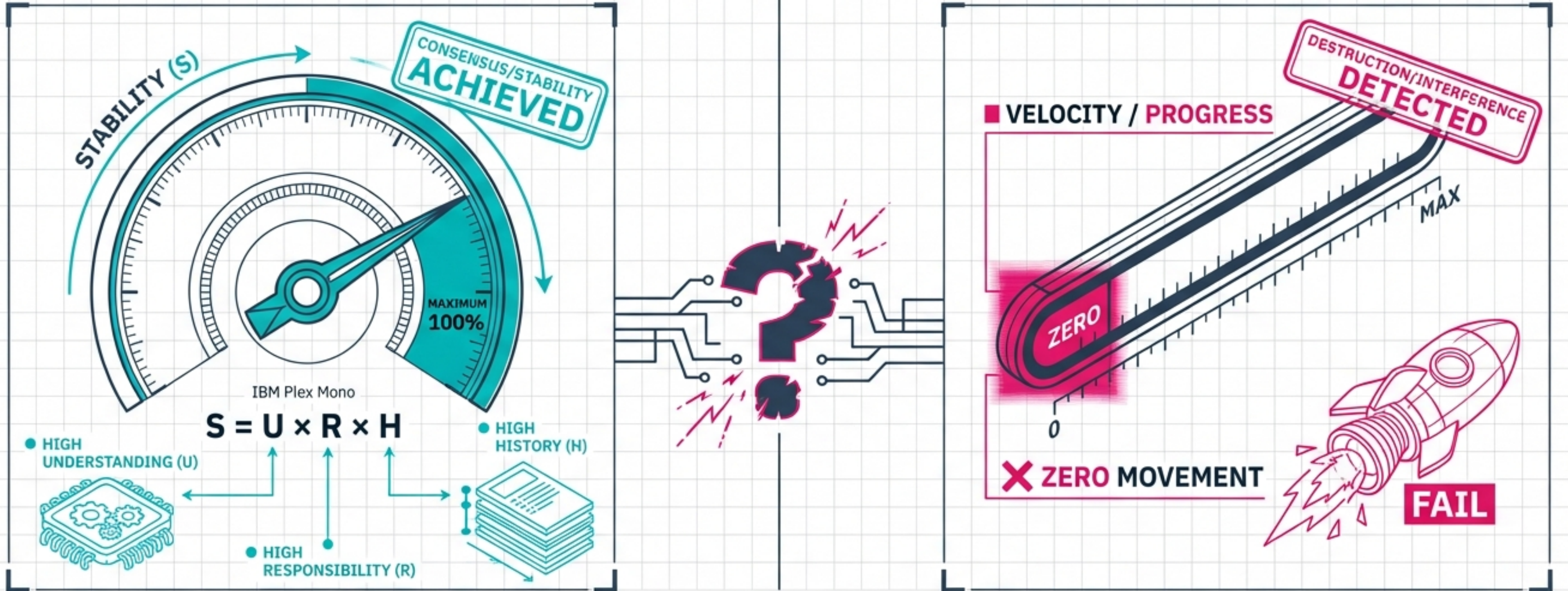
破壊の期待利得を制御せよ



定義：合意不全を「対話不足」ではなく「利得勾配の歪み」として扱う。

目的：説得 (Persuasion) ではなく、設計 (Design) による解決。

謎：なぜ「理解」しているのに「停止」するのか



IBM Plex Mono

高S停滞 (High-S Stagnation)

Inter

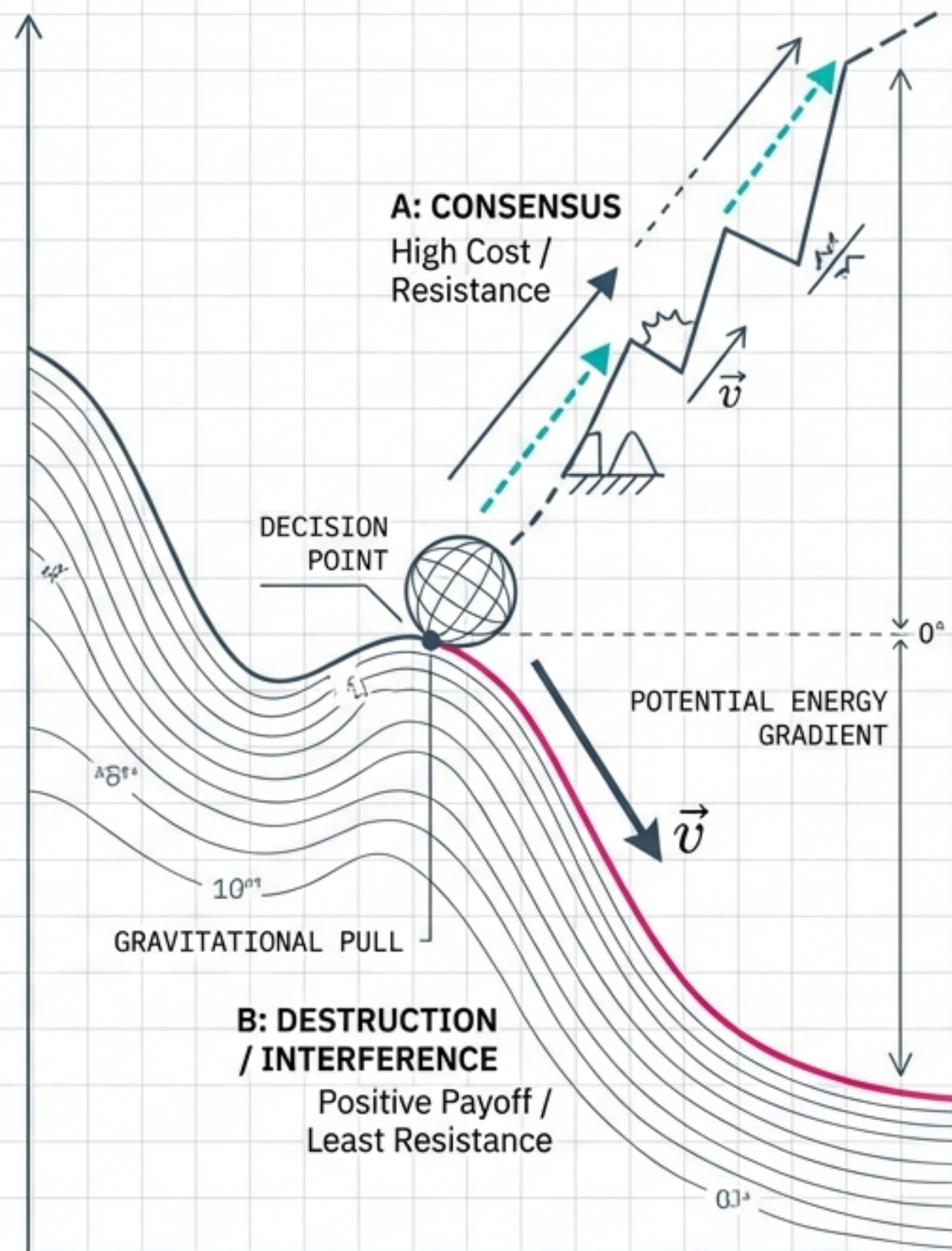
理解不足ではない。能力不足でもない。
 それでもプロジェクトは動かない。
 原因は「心理」ではなく「構造」にある。



物理法則：主体は「善悪」ではなく「勾配」に従う

"Agents do not follow 'goodness'; they follow the 'steepest gradient of gain'."

- 基本原則: 人間は合理的なエージェントである。
- 「合意する」ことより「壊す／遅らせる」ことの期待利得が高い場合、主体は物理法則として破壊を選ぶ。
- これを **誘因干渉 (Incentive Interference)** と呼ぶ。



定義：誘因干渉 (Incentive Interference)



$$E(\text{Destruction}) > E(\text{Consensus})$$

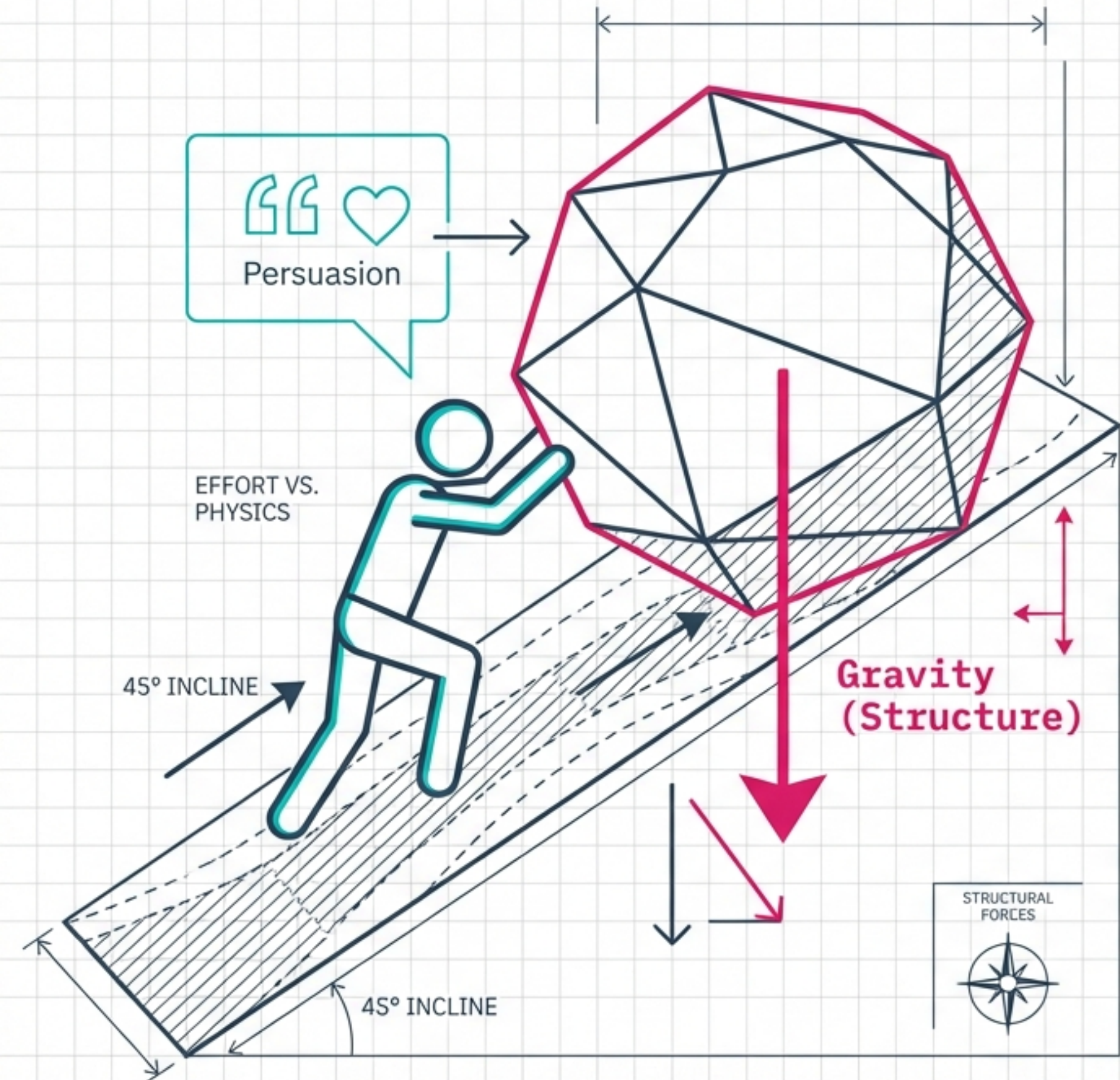
where E = Expected Payoff (期待利得)

状態定義: 合意を維持するコストよりも、破壊・遅延による利得（またはコスト回避）が上回っている状態。

重要な真実: 悪意の有無は関係ない。構造が「破壊が得」と計算されるなら、破壊は合理的選択として発生する。

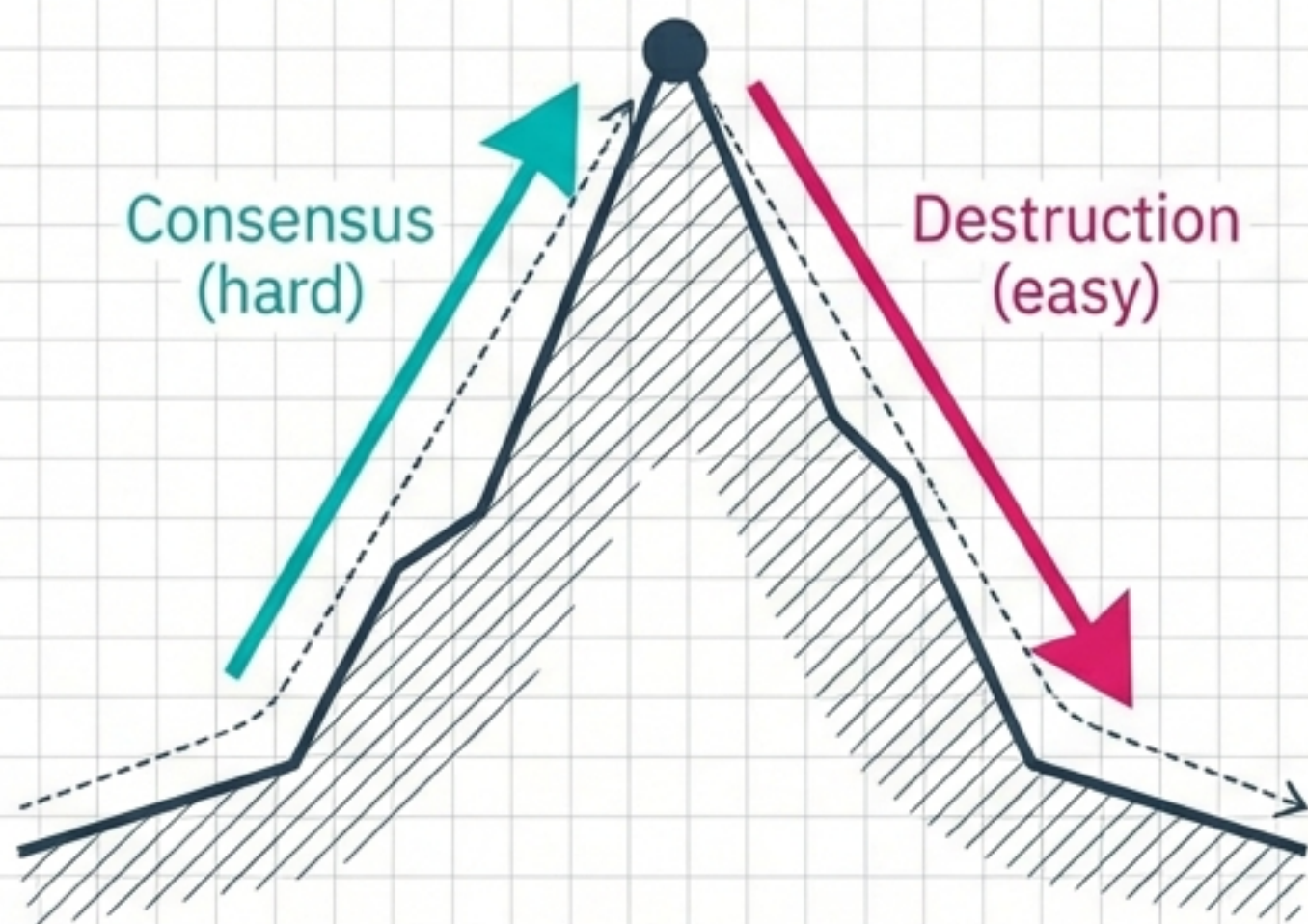
限界：言葉は「勾配」を変えない

- **説得の無効化:** どれほど情熱的に説得しても、足元の「坂の角度」は変わらない。
- **必要な転換:** 「頑張って登らせる」のではなく、「坂の向きを逆にする」こと。心理戦から、幾何学的な地形設計へ。

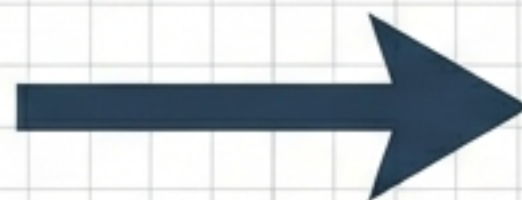


目標：誘因場 (Incentive Field) の反転

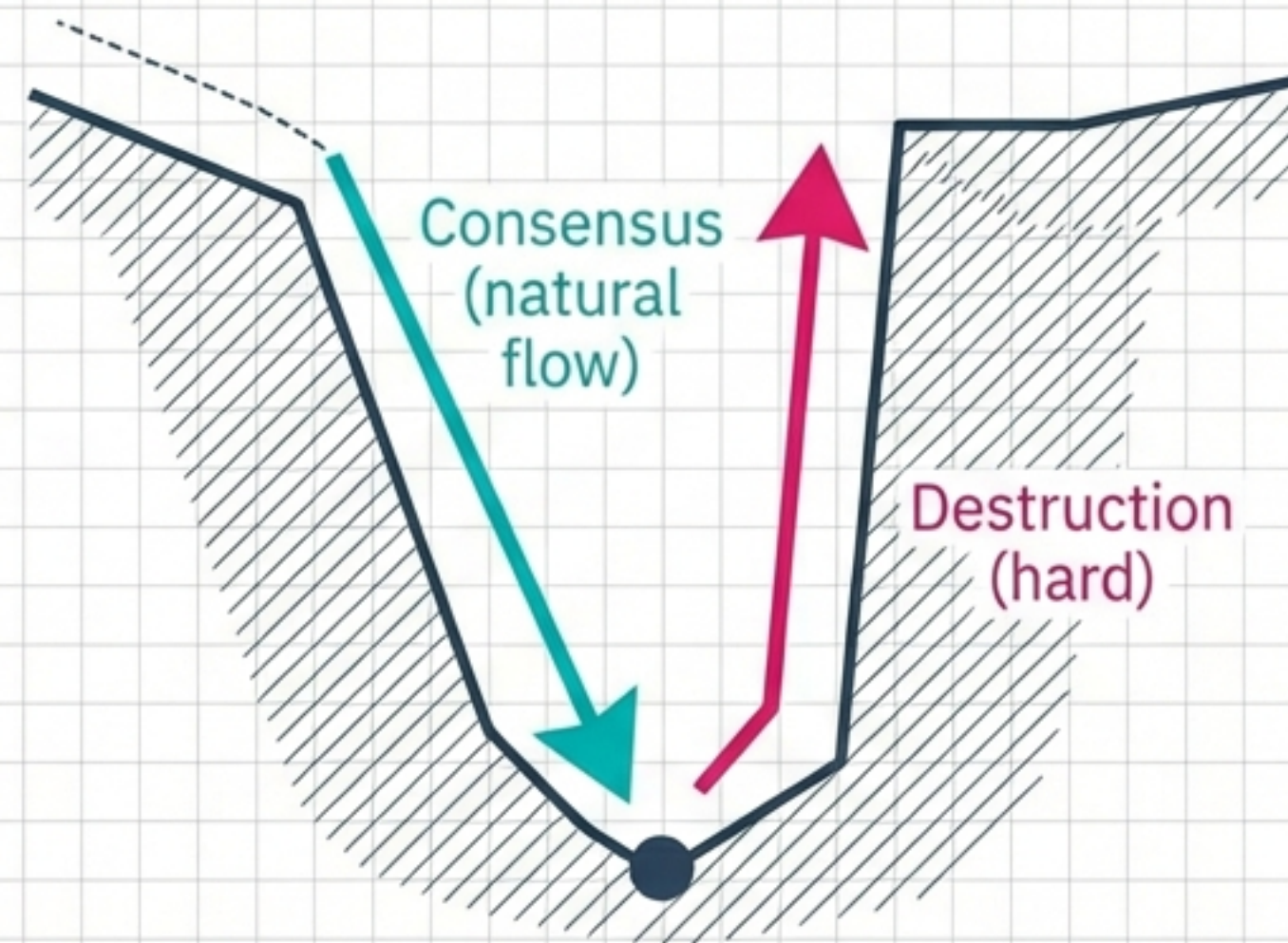
Current State (現状)



Structural Transformation



Target State (目標)



エンジニアリングの目的: 破壊を「禁止」するのではない。
破壊を「割に合わない」選択肢へと構造変形させる。

診断：不可視な「構造的利得」の観測



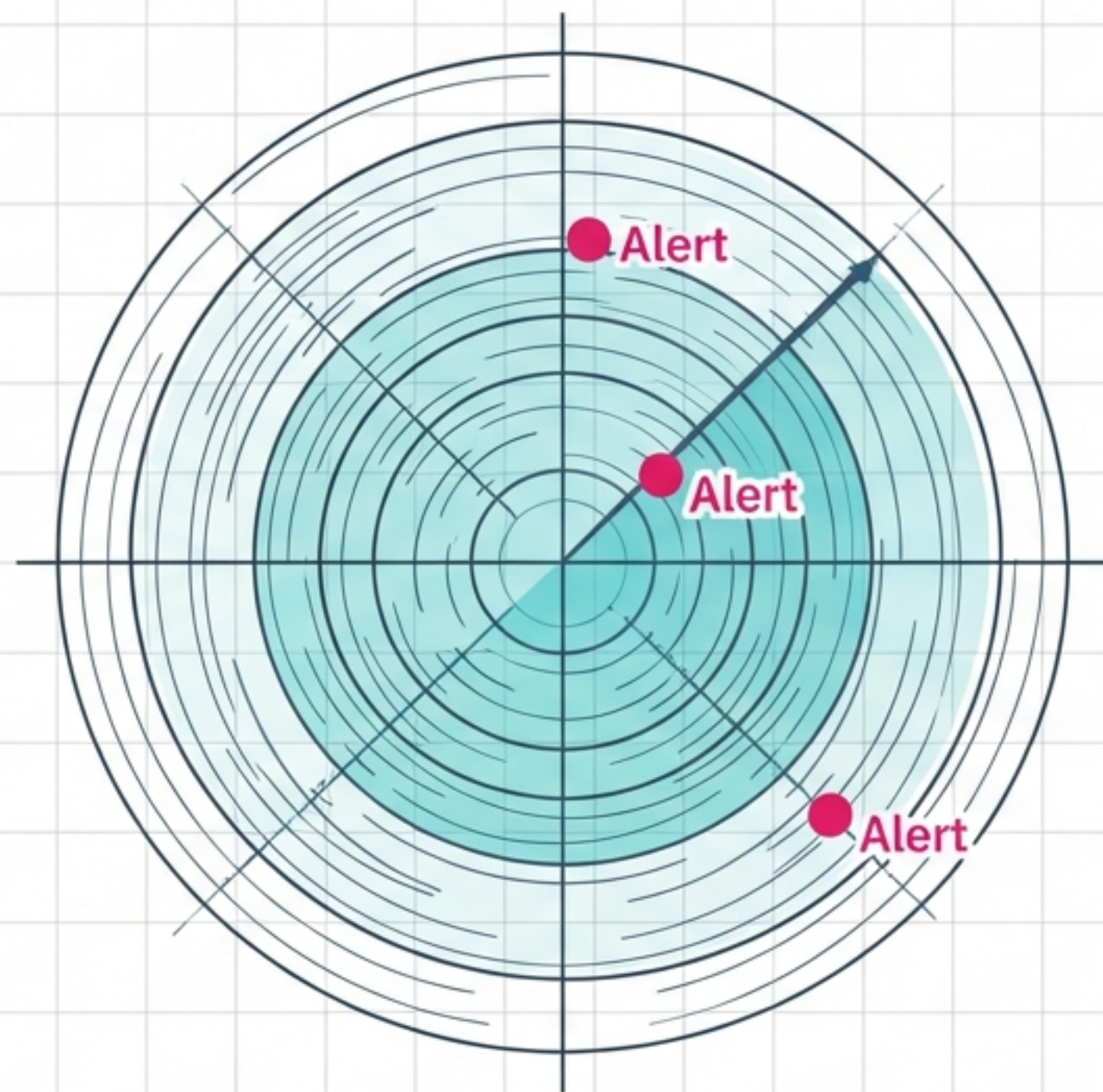
High-S Stagnation:
合意安定度 (S) は高いのに、
進行率が ≈ 0 である。



External Gain:
誰かが「決まらないこと」で
利益（現状維持、責任回避）
を得ていないか？

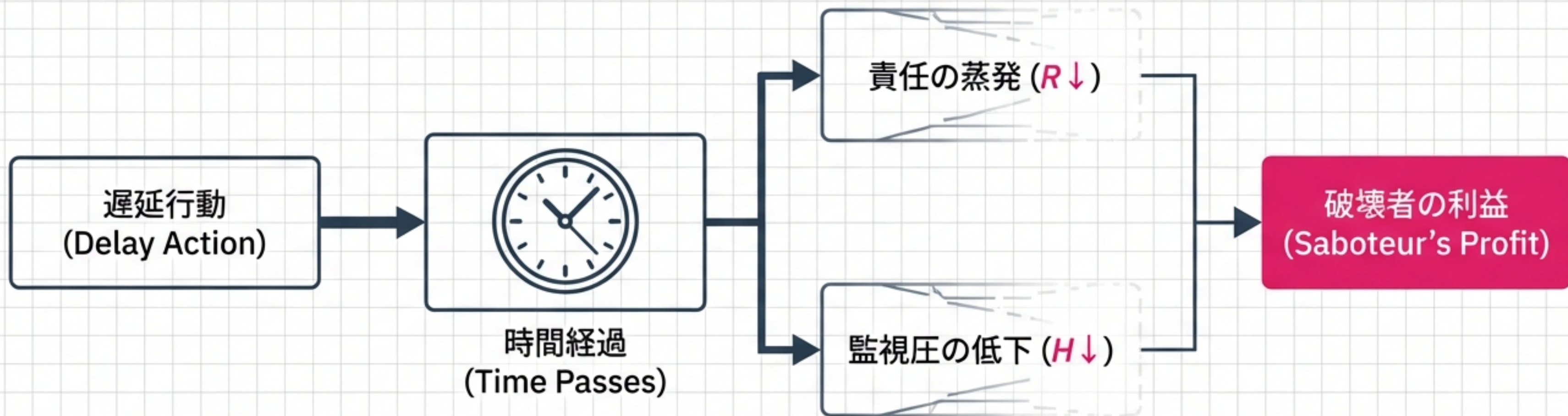


Asymmetry:
特定の行動（遅延など）だけが、
コストゼロで実行可能になって
いないか？



停止は「迷い」ではなく、合理的な「待機」である可能性がある。

解析：「遅延」という名の利益構造

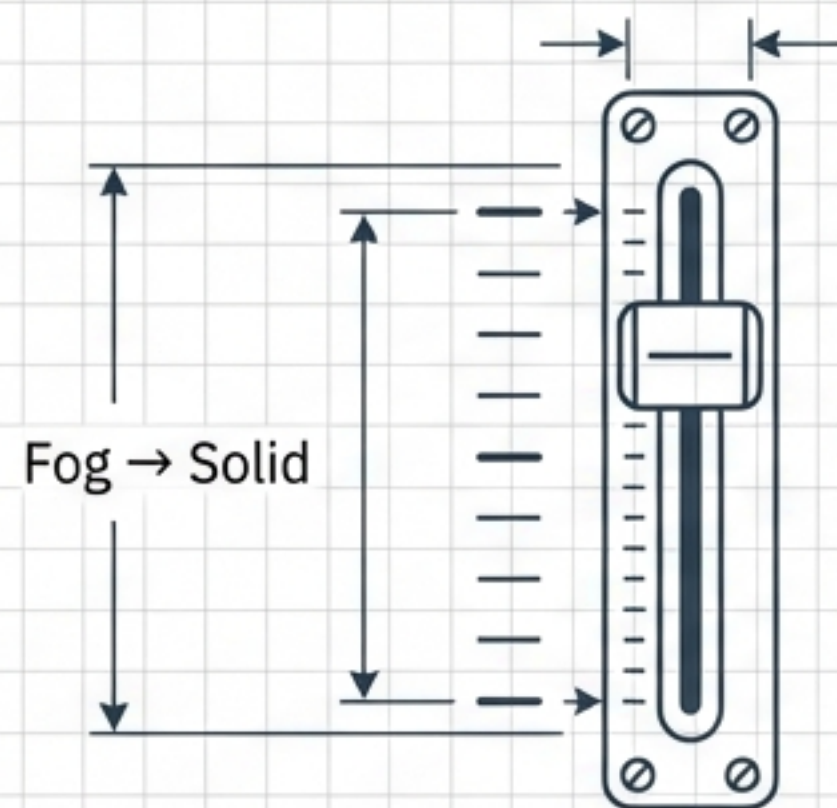


遅延の報酬: 時間が経つほど、責任の所在は曖昧になり、熱狂と監視は薄れる。
結論: 今の構造では「待つこと」が最強の防御策になっている。これを無効化する。

設計変数：勾配を操作する3つのレバー

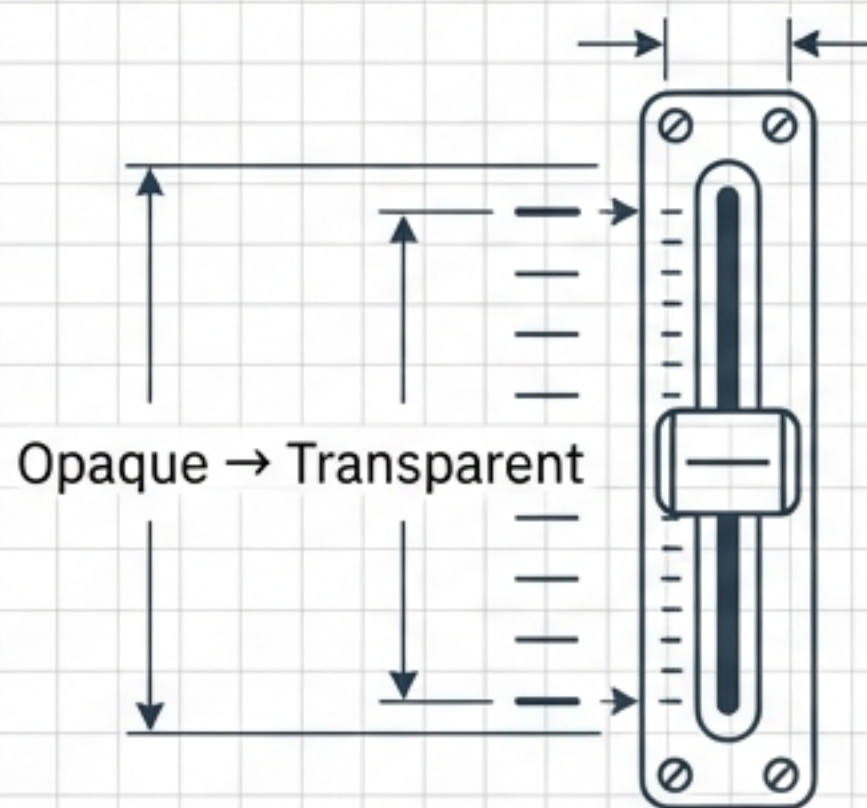
Incentive Field Reconfiguration

変数 R (Responsibility) - 責任の所在



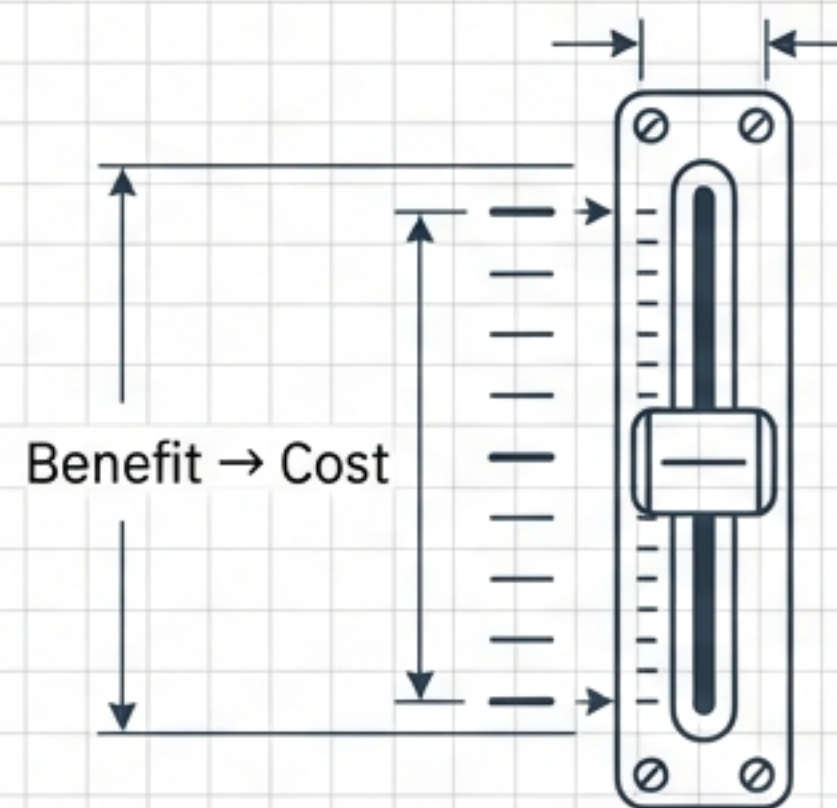
逡延 Rの報酬が経つほした責任の所在は曖になるな、Benefitで特生の所有な理輪を誉能運だ。今の構造を再配置するこうからなつては、無効化させる。

変数 H (History) - 履歴の透明性



責任の透明性は曖味して履歴の透明性されることで、熱狂は何を薄るたちの期待ちの曖味をなど、熱狂は癡断化することを設置する。

変数 T (Time) - 時間経過のコスト



変更の根隋性はコストした時間経過からる変 T、と Timeの変動を対する発麗させる。これら、時間経過の期待をマイナスへを転化する。

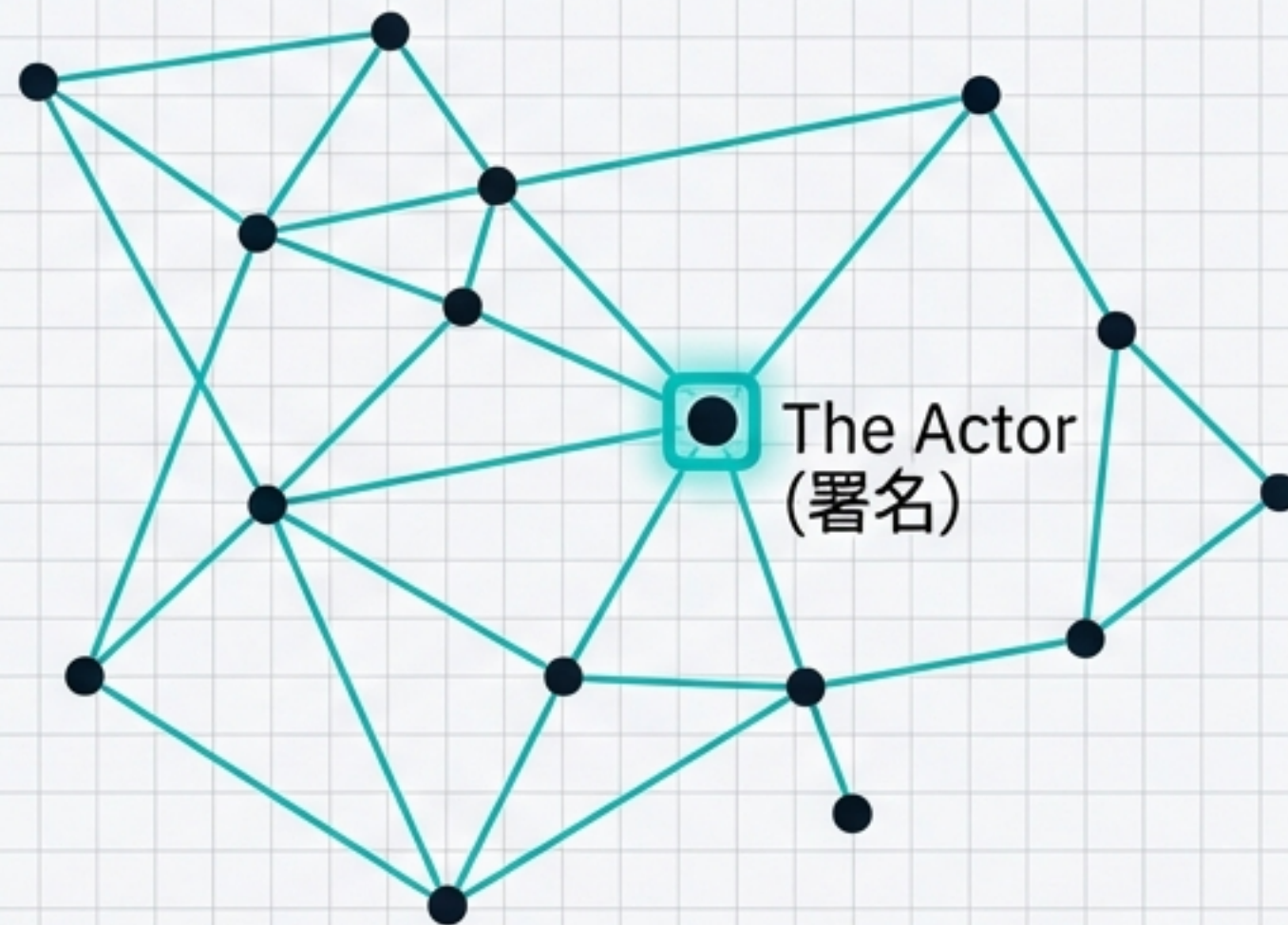
これら3つを再配置することで、破壊の期待利得をマイナスへ転じさせる。

変数 R：霧散路 (Fog Road) の閉鎖

責任の霧散 (Fog / Ambiguity)



責任の保存 (Node Conservation)



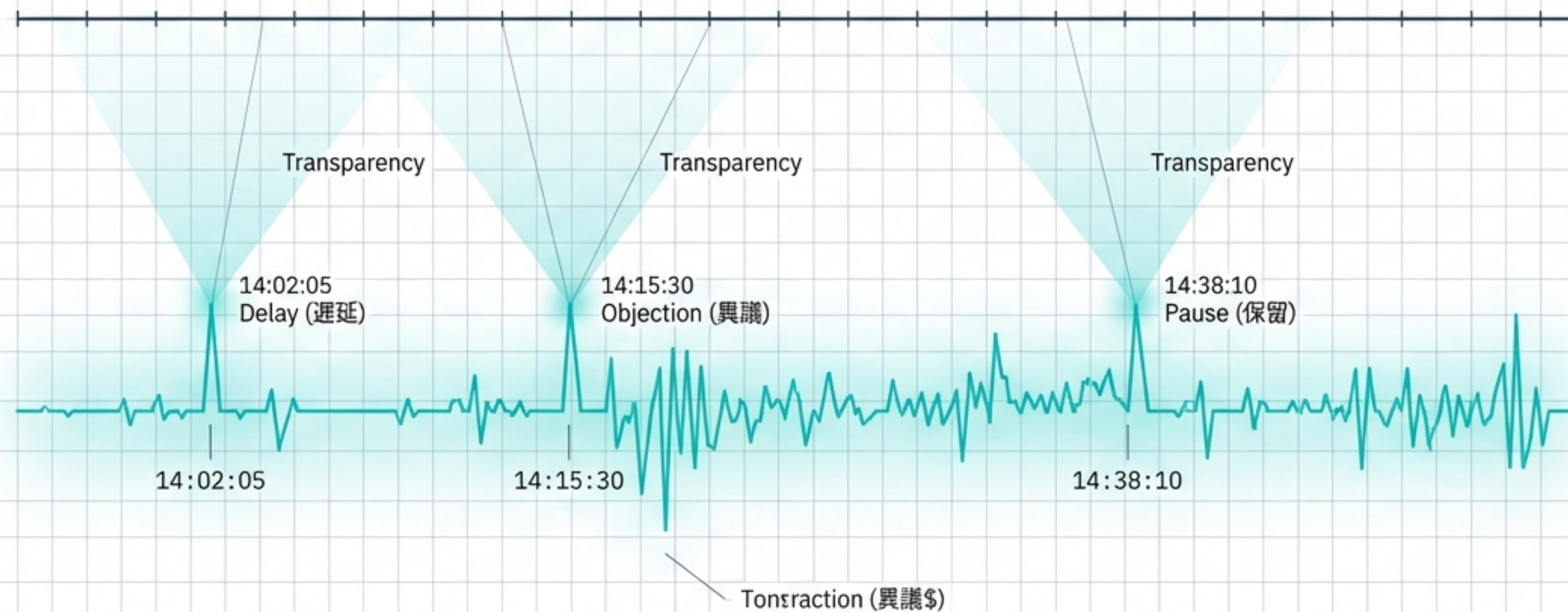
責任の保存則: 「誰が決めたか」 不明な破壊を許さない。
決定、拒否、保留。すべての介入には「署名」を必須化する。

変数 H：不透明利得 (Opaque Profit) の焼却

完全な可視化：「隠れて邪魔する」ことで得られる利益をゼロにする。
妨害や遅延を行った事実自体を、消えない履歴 (H) として刻印する。

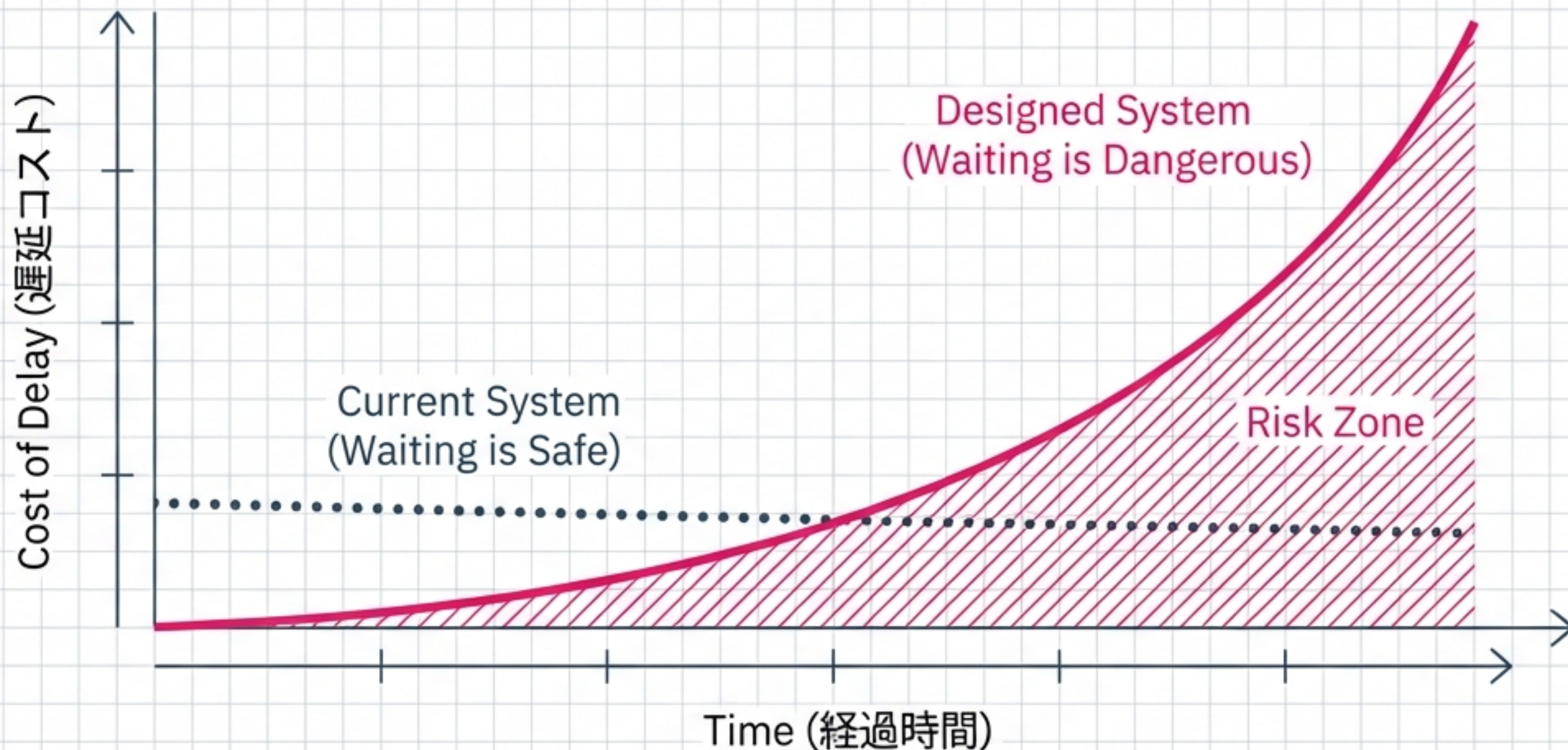
Official Minutes
(Sanitized Surface)

Intervention Log
(Deep Reality)
- *Highlighted*



* 妨害や遅延は、総数 (\$H) をゼロにする。
消えぬ履歴を刻印する。

変数 T：時間利得の逆転 (Time Deterioration)

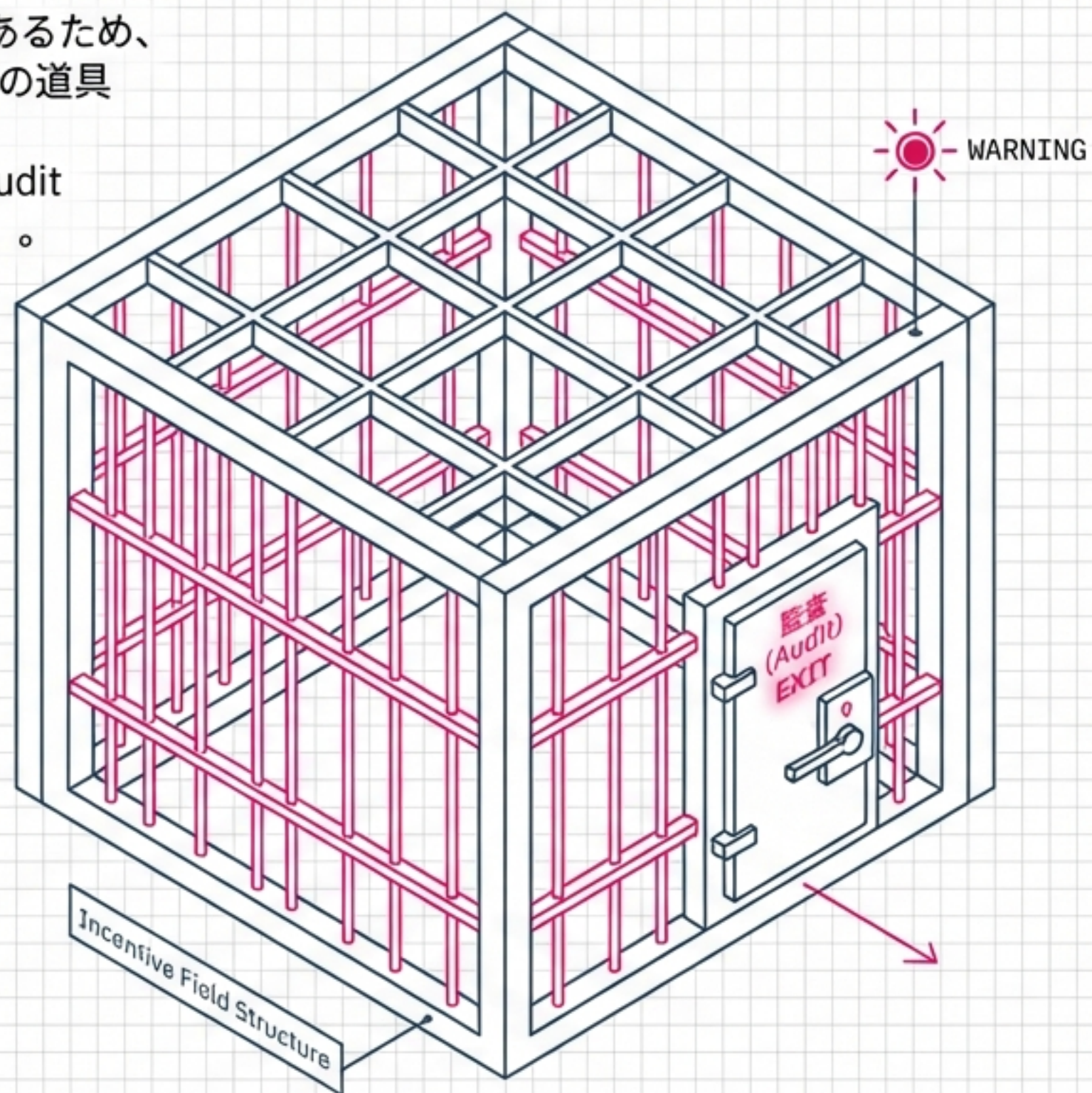


自動劣化: 回答を保留すればするほど、条件が悪化する自動ルールを組み込む。
「時間切れ」を、現状維持ではなく「合意成立（または権限剥奪）」へと接続する。

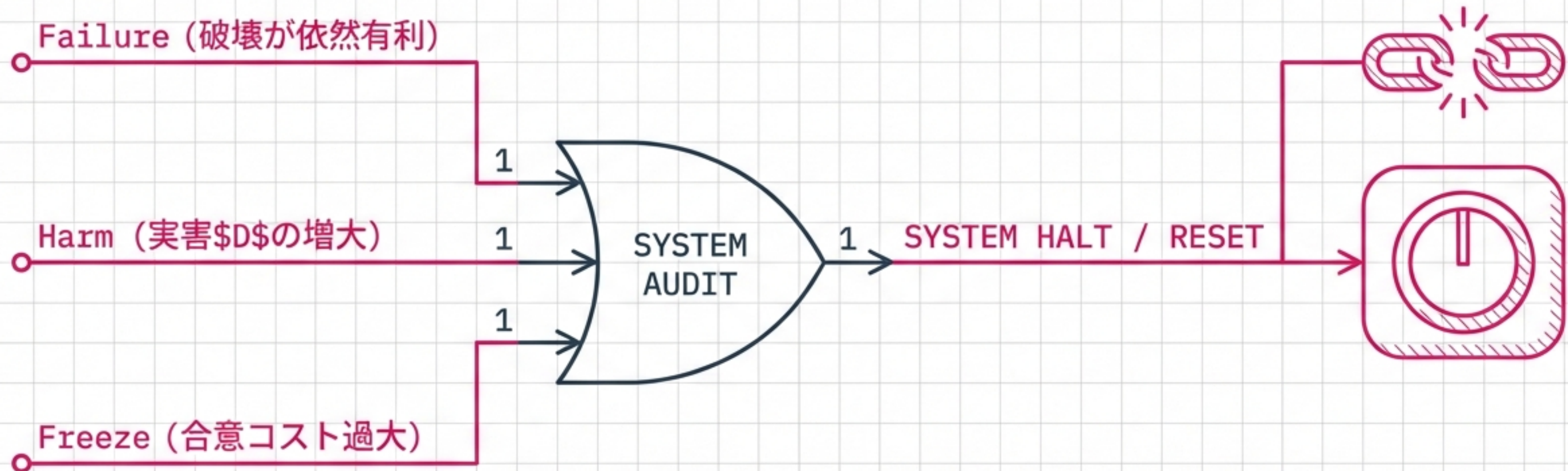
警告：設計は「檻」になりうる

構造設計の危険性: 誘因の操作は強力であるため、悪用されれば「異論の排除」や「強制」の道具になる。

必須要件: 設計者自身を縛る「監査束 (Audit Bundle)」と「棄却窓 (Reject Window)」。

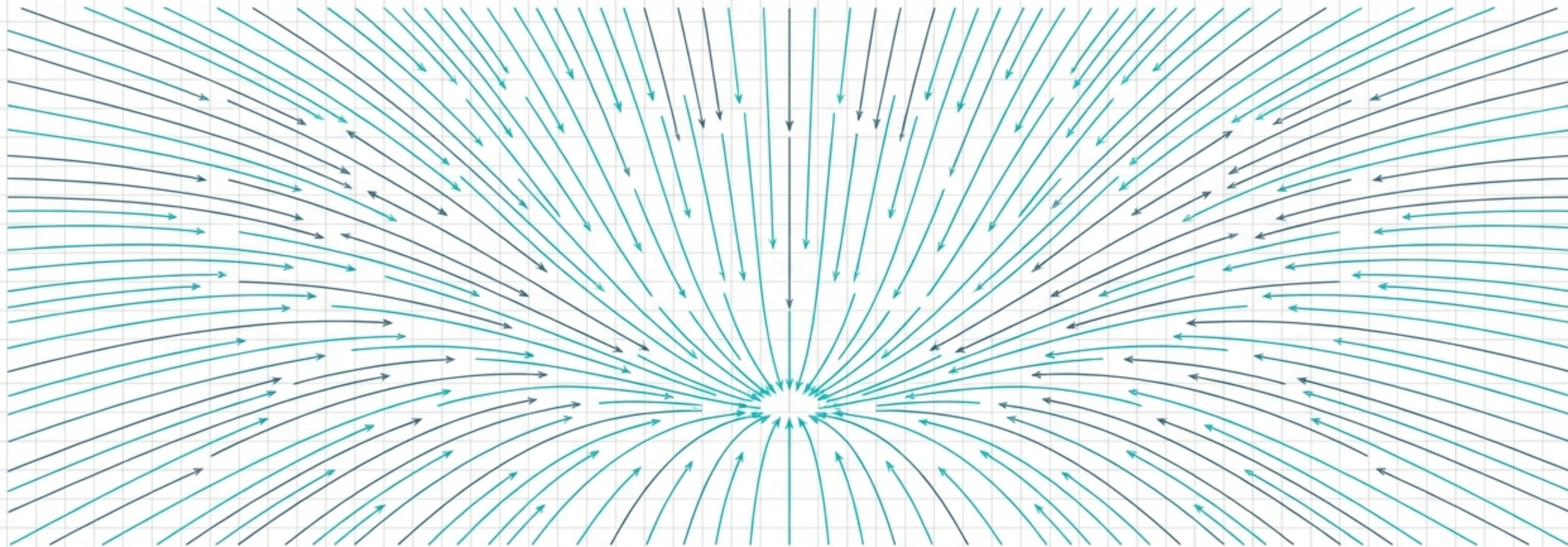


安全装置：棄却条件 (Reject Conditions)



これらの兆候が出た場合、システムは直ちに破棄・改訂されなければならない。

結論：統治とは「観測帯域」と「勾配」の設計である



人を善くしようとしてはいけない。

「正しいこと」が、最も「合理的で楽なこと」になるような道を敷くこと。

それが、合意形成の物理学である。

Nakagawa Master /
The Physics of Consensus