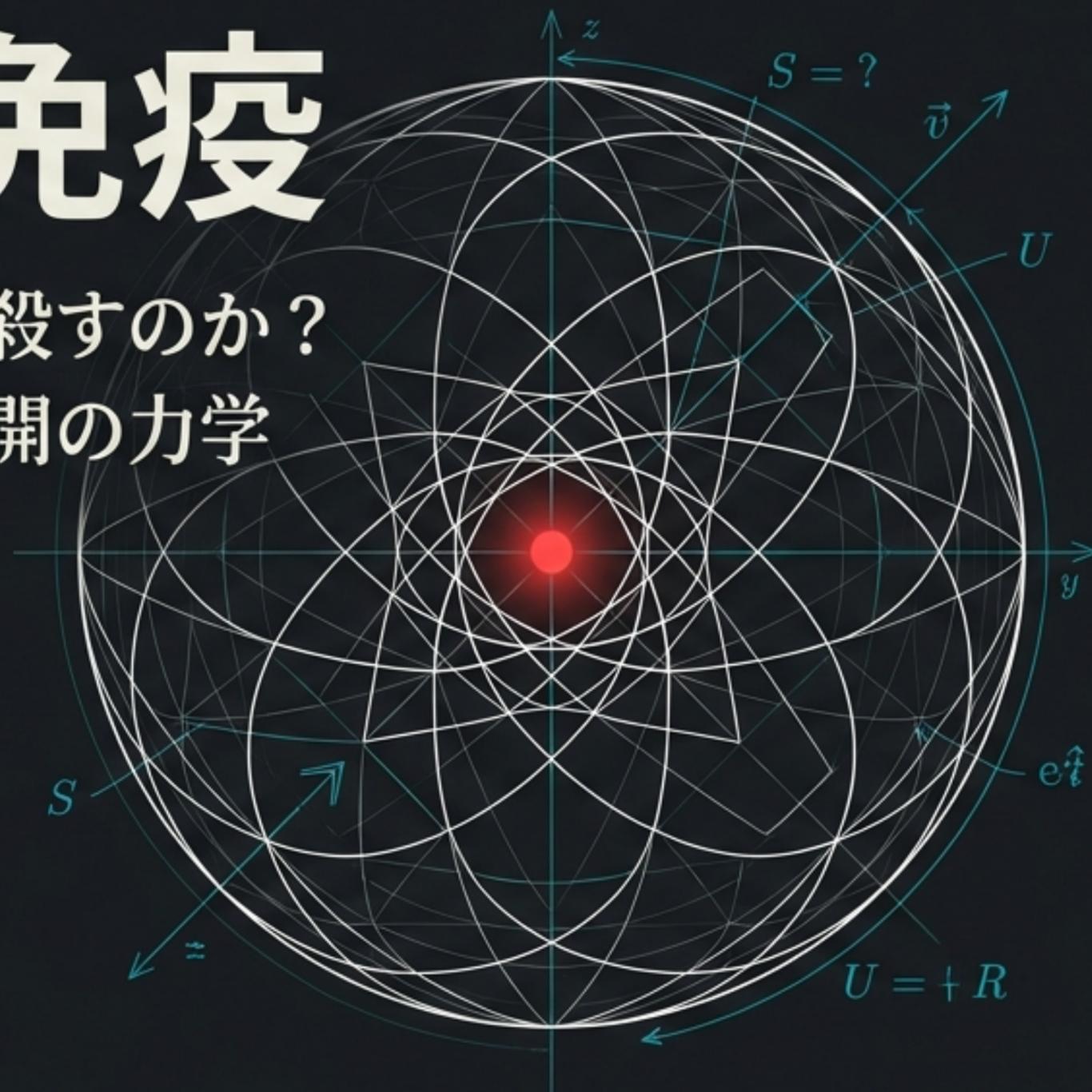


# 逸脱と免疫

なぜ「罰」は組織を殺すのか？  
構造的免疫と差分公開の力学

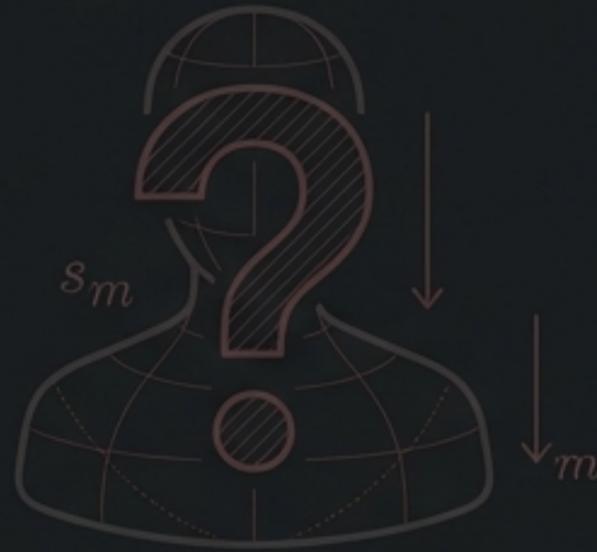


# 問いの転換：道徳から物理へ

STATUS: CRITICAL ANALYSIS

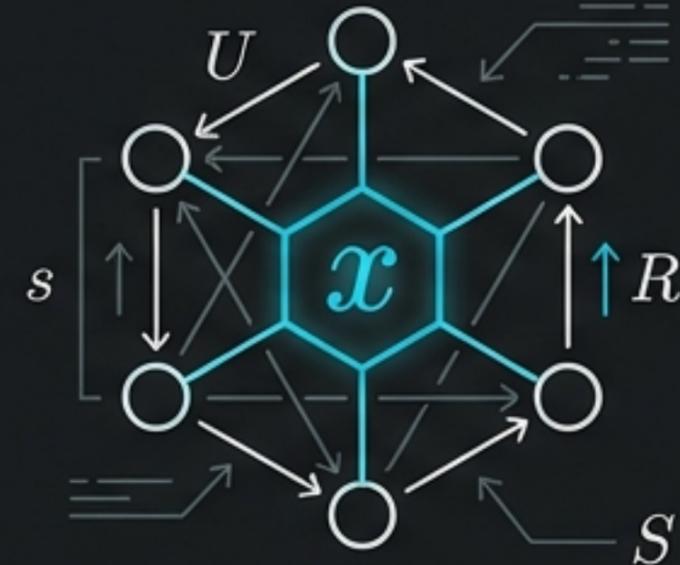
SYSTEMS ARCHITECTURE

## Moral View (道徳)



問い：「誰が悪いのか？」  $\downarrow_m$   
行動：犯人探し、厳罰化、精神論  $\downarrow_m$   
結果：隠蔽と崩壊  $\downarrow$

## Physical View (物理)



問い：「どの変数が落ちたか？」  $U$   
行動：差分公開、修復ノード、構造設計  $\uparrow$   
結果：速度と回復  $\uparrow S$

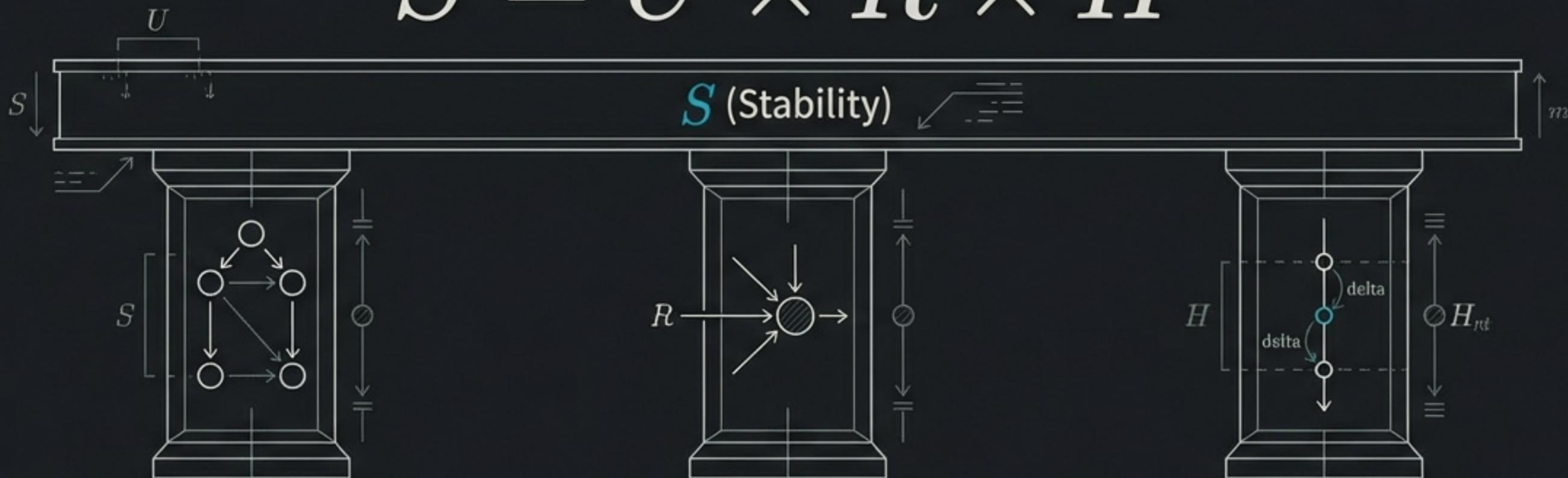
「組織における逸脱は、人格の欠陥ではなく、設計と運用の間に必然的に生じる差分である。」

STATUS: CRITICAL ANALYSIS

DATA POINTS: SAU, 3S, 3RS

# 合意安定度の方程式

$$S = U \times R \times H$$



**U (Understanding)**  
(Understanding)

第三者再現可能性。誰が見ても「なぜそうなったか」が分かる状態。

**R (Responsibility)**  
(Responsibility)

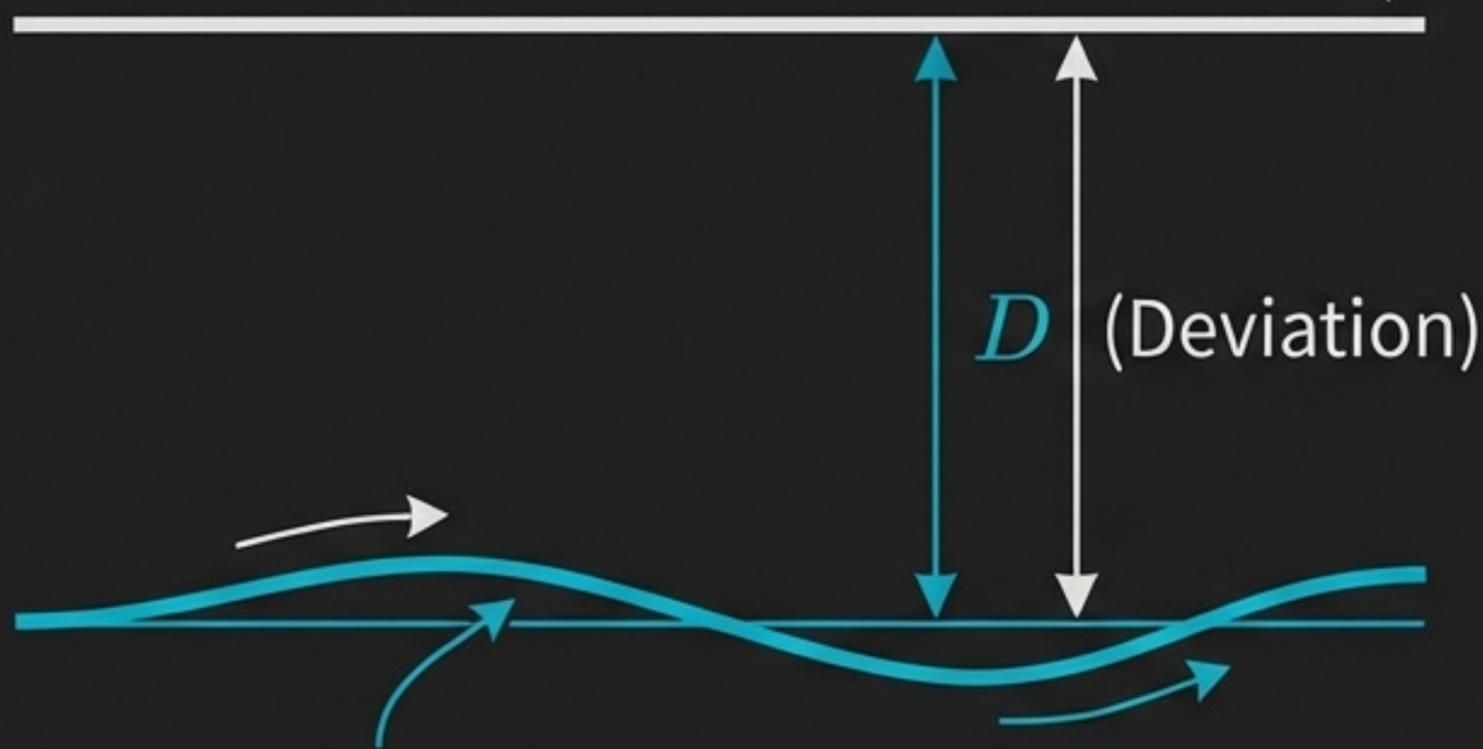
責任特定可能性。「誰が修復の入口か」が一意に定まる状態。

**H (History)**  
(Bone White)

履歴公開度。意思決定の「差分」が検証可能な状態で残る状態。

# $\$D$ (逸脱) の物理定義

設計 (Design): ルール、手順、理想



運用 (Operation): 現場の判断、省略、例外

$$D = \text{Design} - \text{Operation}$$

- $D$ は「悪意」ではない。「差分」である。
- 時間が経過すれば、エントロピー増大則により  $D$  は必ず発生する。「逸脱ゼロ」を目指すことは、物理法則に逆らうことである。

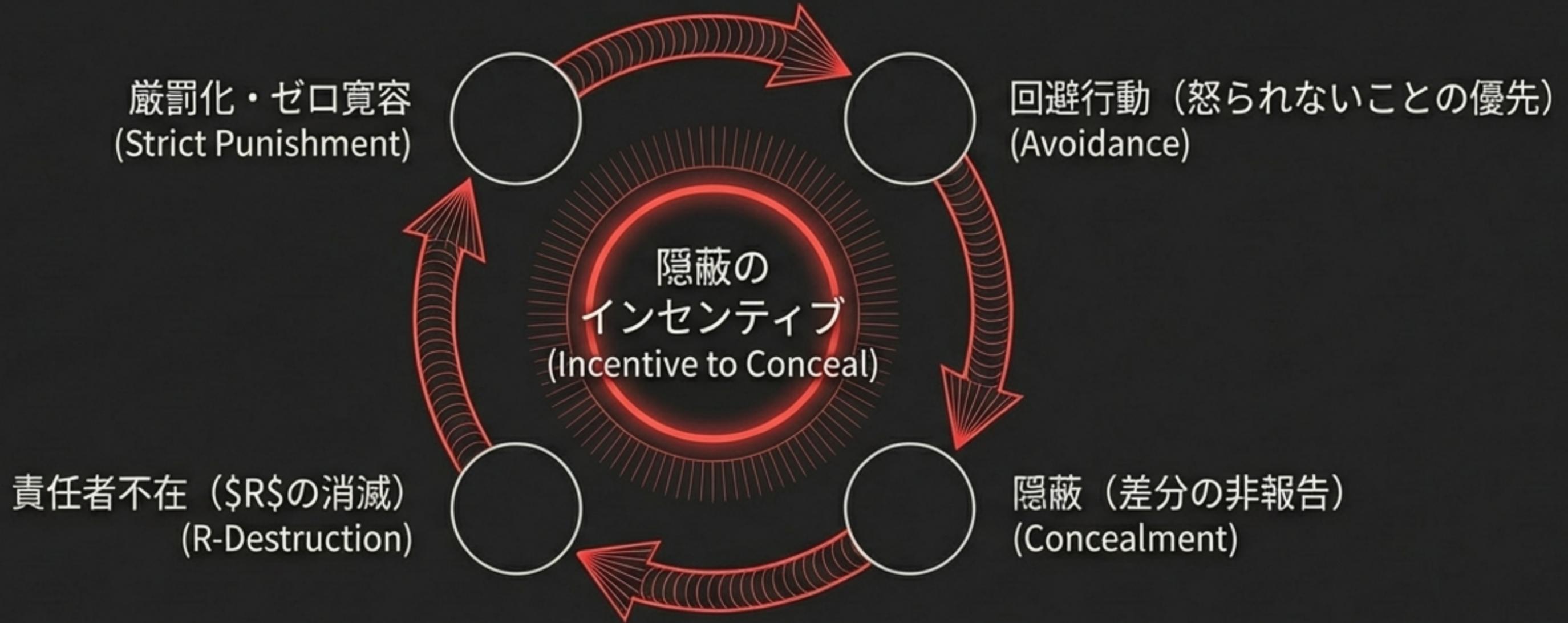
# 「罰」のパラドックス

STATUS: CRITICAL ANALYSIS  
DATA POINTS: 5AU, 2S, 3RS

STATUS: CRITICAL ANALYSIS

STATUS: CRITICAL ANALYSIS

SYSTEMS ARCHITECTURE



罰 (Punishment) は抑止力として設計されるが、  
物理的には「**隠蔽のインセンティブ**」として機能する。

STATUS:

CR210A1

# 静かなる崩壊：潜伏 (Latency)

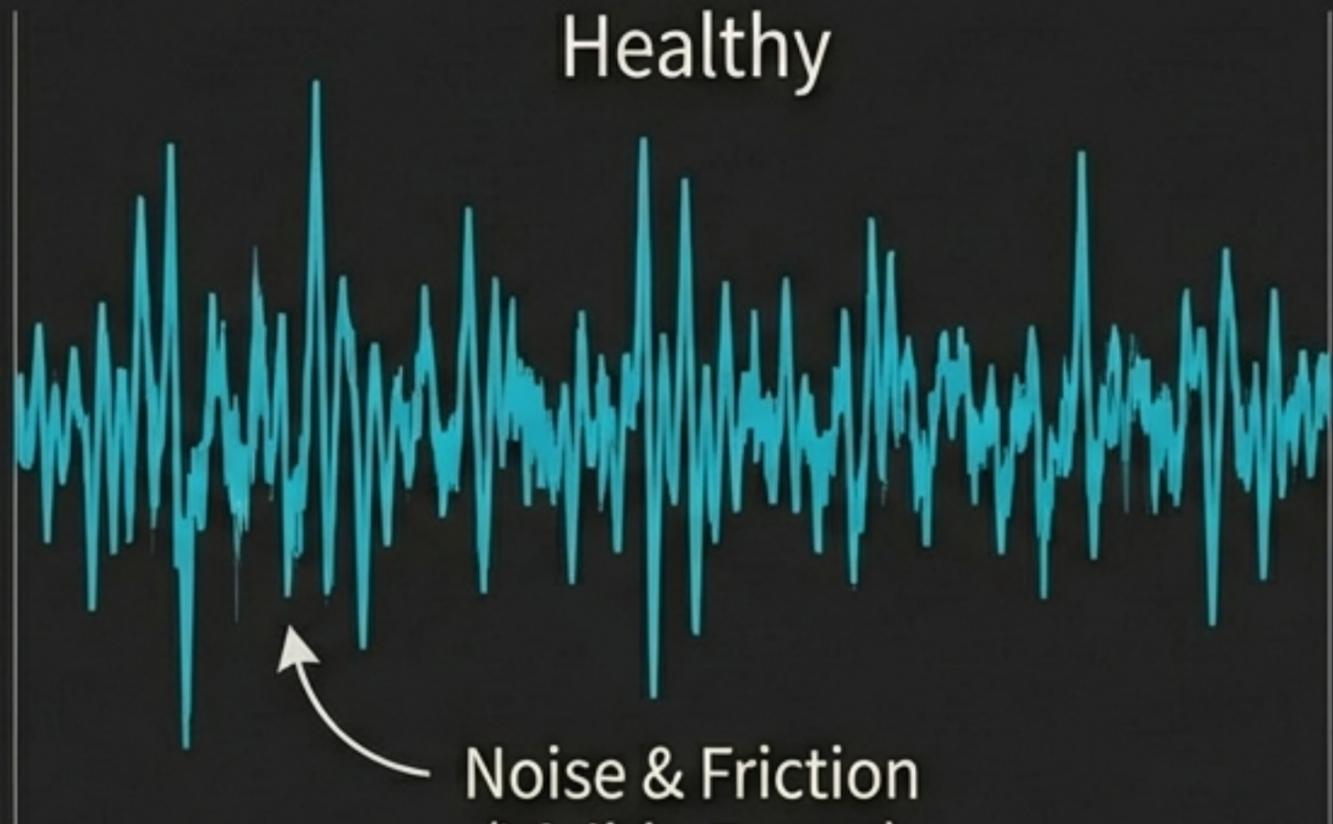
STATUS: CRITICAL ANALYSTS  
DATA POINTS: 3AU, 2S, 3RS

STATUS: CRITICAL ANALYSIS

STATUS: CRITICAL ANALYSIS

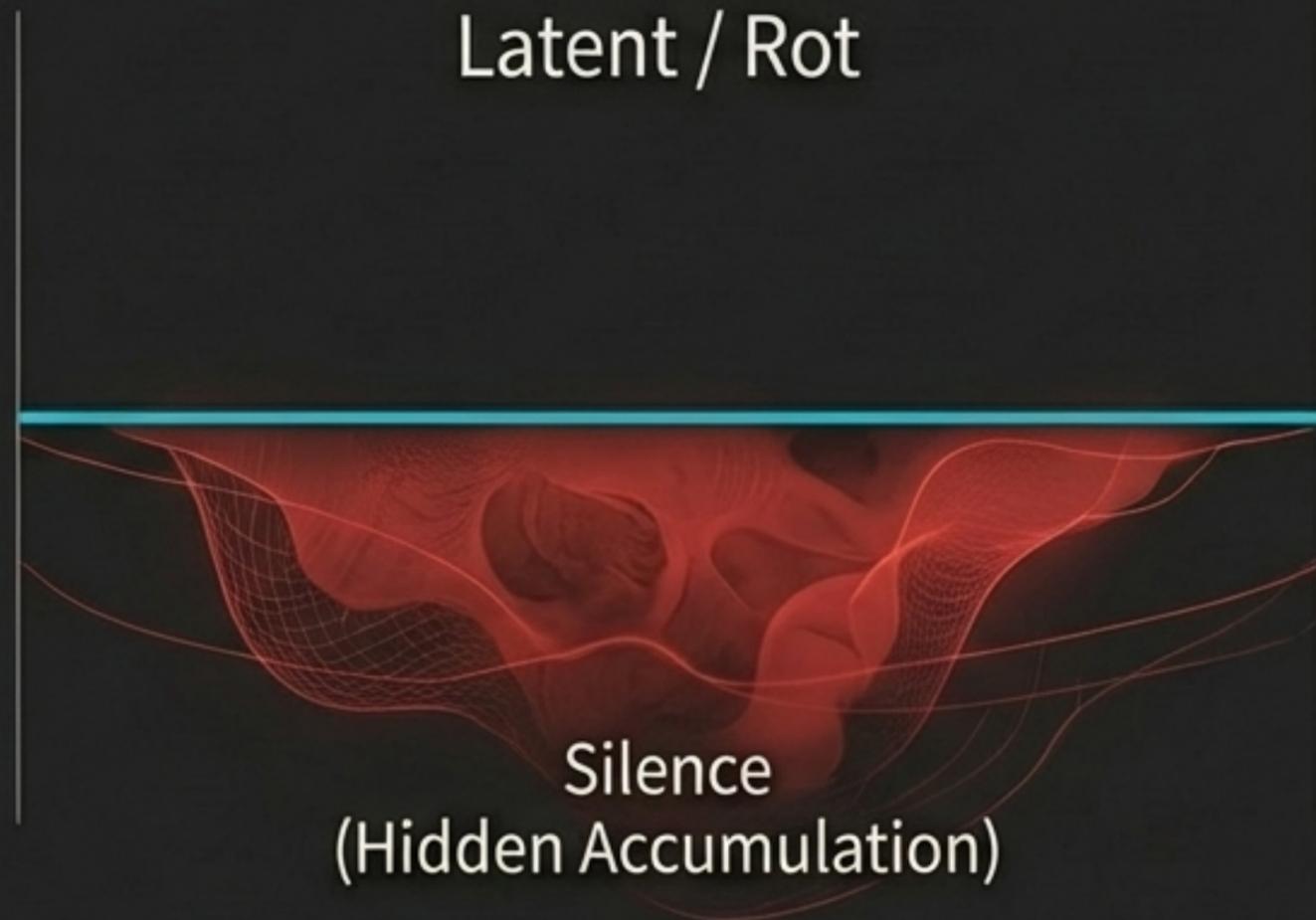
SYSTEMS ARCHITECTURE

Healthy



Noise & Friction  
(Visible Errors)

Latent / Rot



Silence  
(Hidden Accumulation)

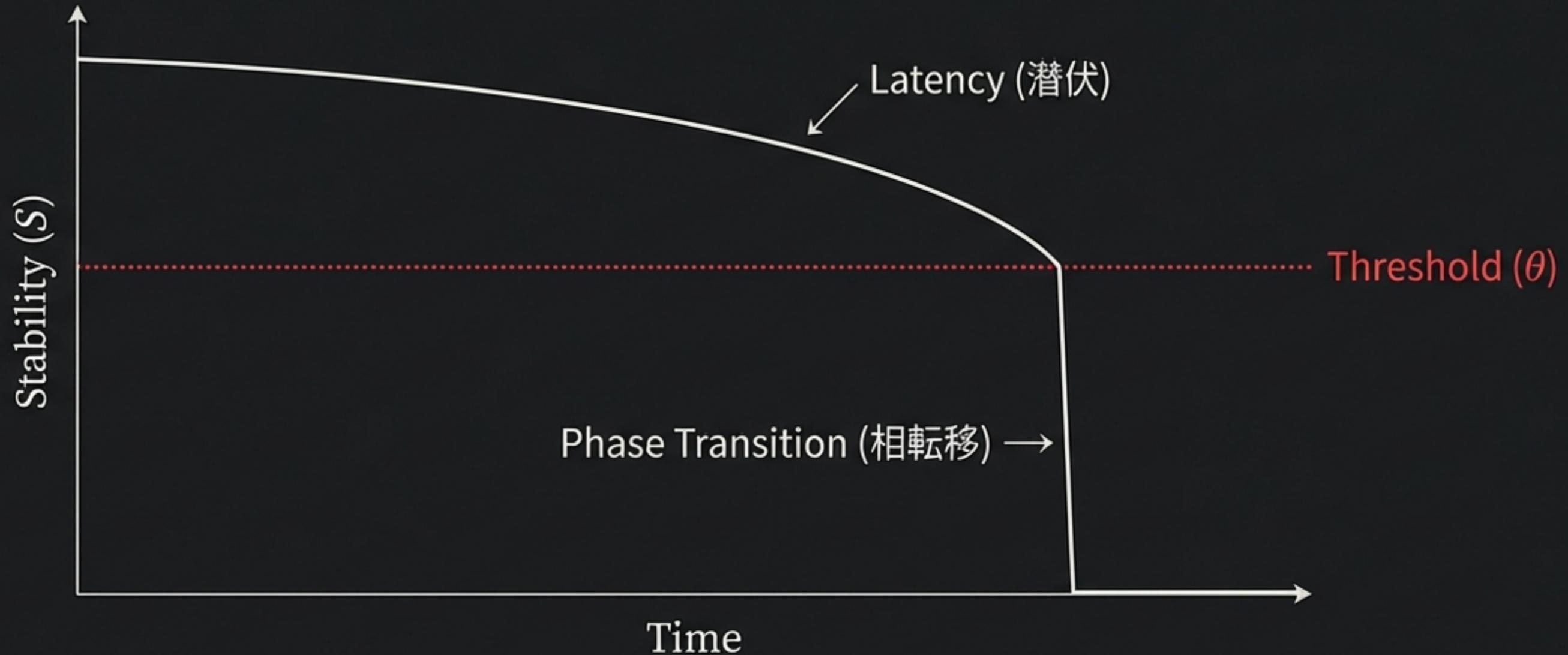
厳罰化された組織で起きる「沈黙」は、平和ではない。  
それは  $D$  (逸脱) が地下で蓄積されている状態である。

警告指標：  $D_{det}$  (Detected Deviation)  $\rightarrow 0$

STATUS:

CR210A1

# 臨界点と相転移

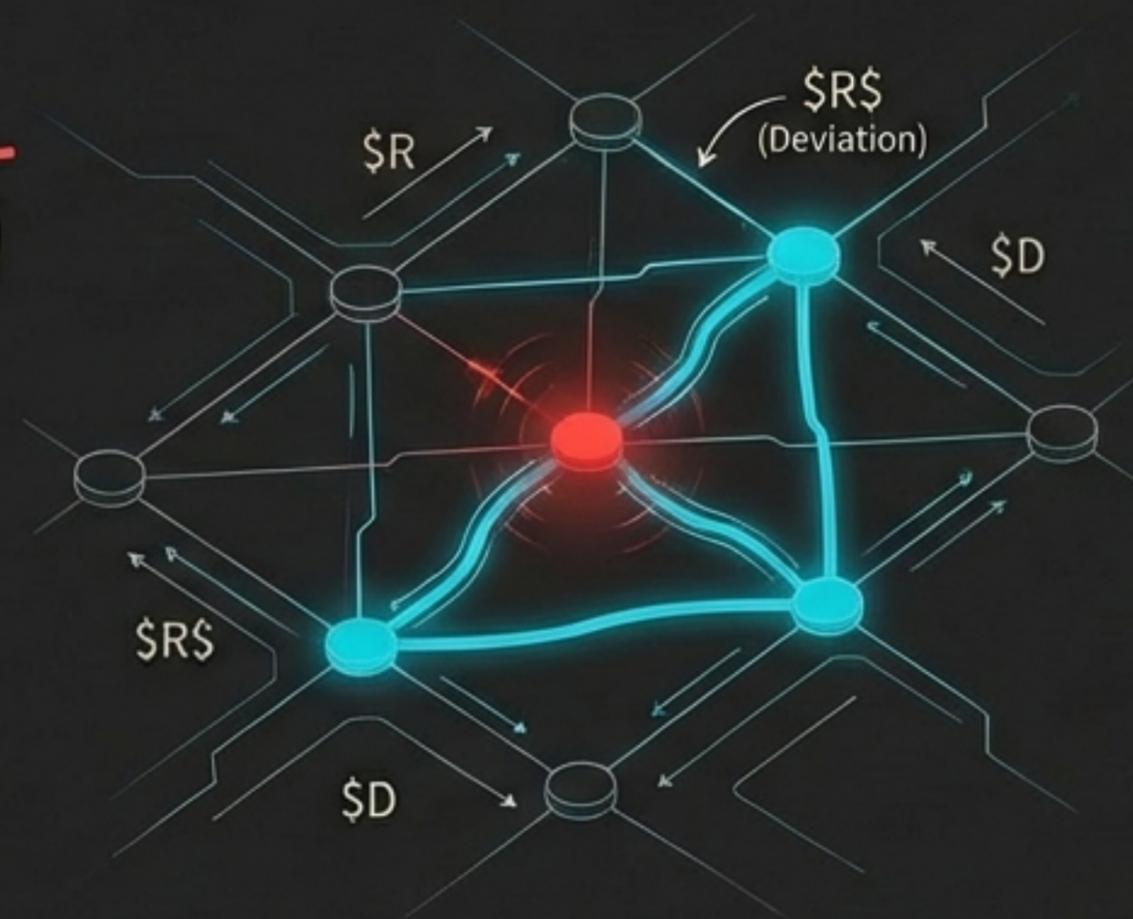


崩壊は「原因」ではなく「状態遷移」である。事件はトリガーに過ぎない。  
本体は、増幅された逸脱 ( $D\$$ ) の爆発である。

# 免疫の再定義

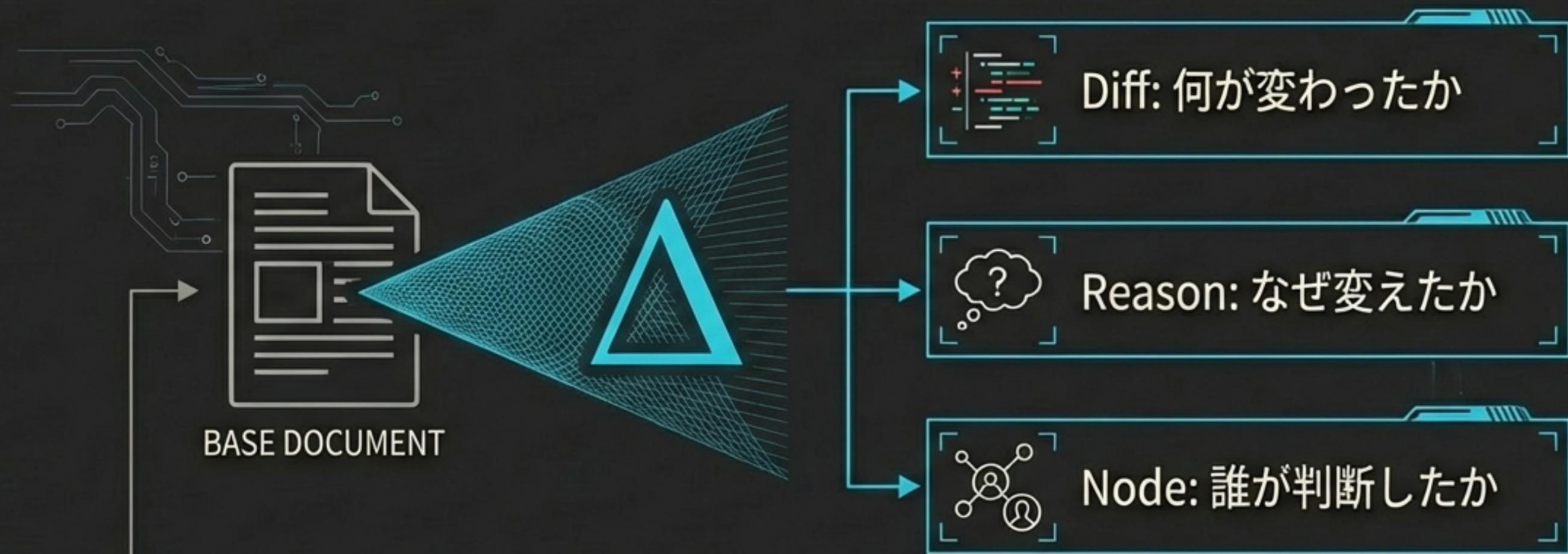
~~免疫 = 病気にならないこと (防御・遮断)~~

**免疫 = 回復速度  
(Recovery Speed)**



逸脱は防げない。だが、壊れない構造は設計できる。  
問題ゼロ信仰を捨て、「どれだけ速く直せるか」にリソースを全振りする。

# 構造的免疫の正体：差分公開 (Diff-H)



すべてを公開する必要はない。

「設計と運用」のズレ (差分) だけを、検証可能な形で残すこと。  
差分が公開される環境では、「隠すコスト」が「利得」を上回る。

# 悪意の経済学



$$H_0 = \left( \frac{I_{L0}}{R_A} - W_{L0} \right)$$

$$H_1 = \left( \frac{I_{L1}}{R_L} - W_{L1} \right)$$

$$M(x,t) = \left( \frac{(x,d)}{ovf} \right) \left( v = G(t) - \frac{(Avef) - (t_0)}{7R_0(L - \eta - \tau)} \right)$$

$$h_1 = F_{(t)} \frac{R_L}{P}$$

免疫は性善説を前提にしない。差分公開 (Diff-H) は、不正の痕跡を不可逆的に残す。

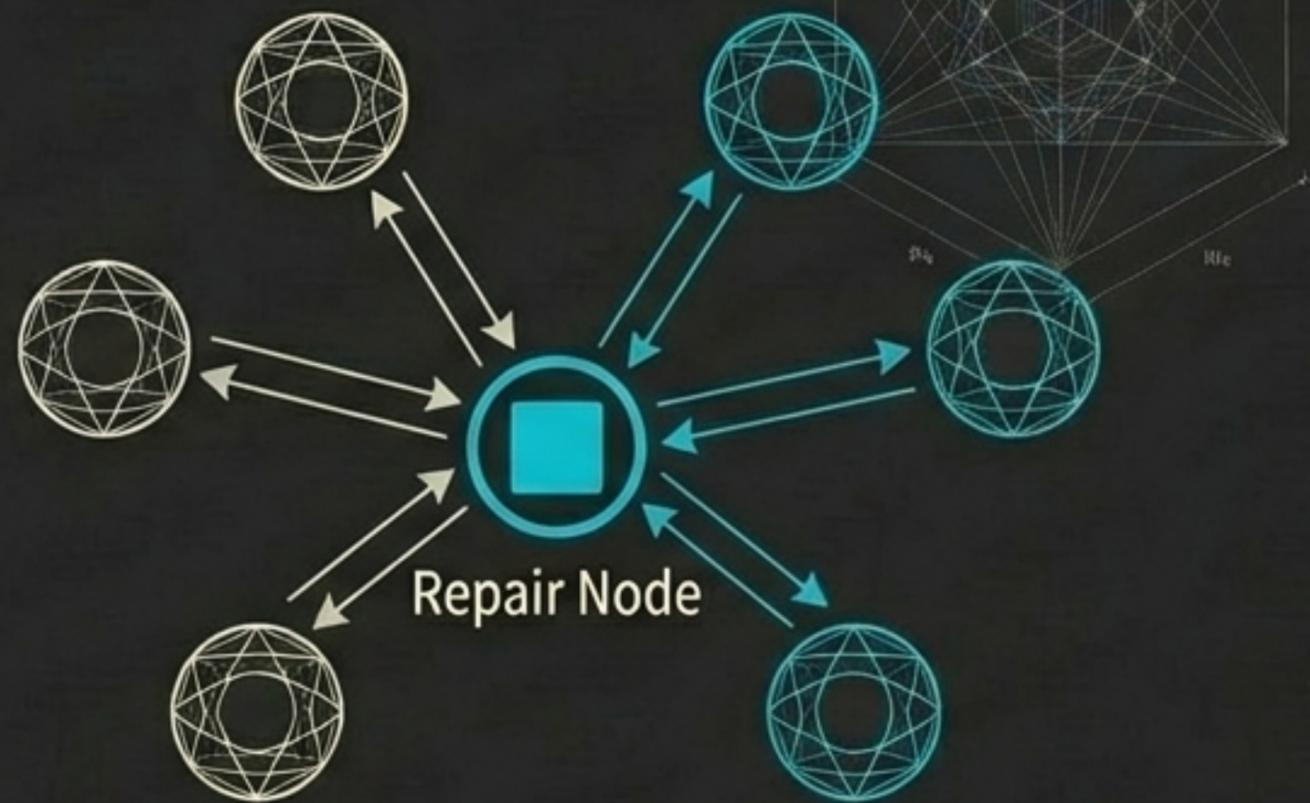
隠蔽コストが利得を上回る瞬間、悪意は合理的選択肢から外れる。

「人を信じるな。構造を信じる。」

# \$Rの再定義：断罪から修復へ



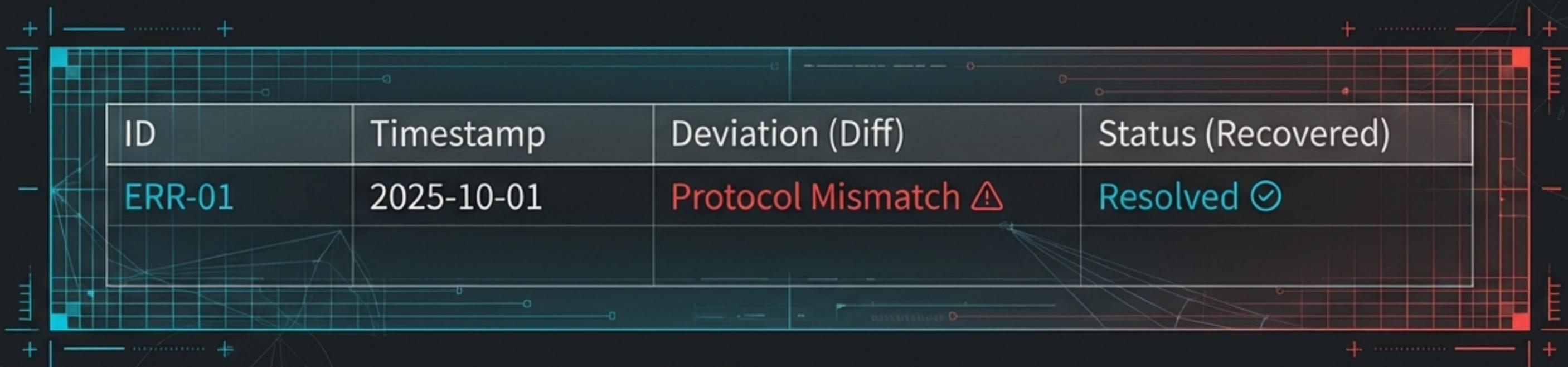
旧 R: 断罪 (Blame) → R Disappears



新 R: 修復 (Repair) → R Active

*R* is not the person to blame.  
*R* is the specific point where repair begins.

# 実装ツール：逸脱レッキャ (Deviation Ledger)



ID	Timestamp	Deviation (Diff)	Status (Recovered)
ERR-01	2025-10-01	Protocol Mismatch ⚠️	Resolved ✅



Early Warning: 破壊の早期検知



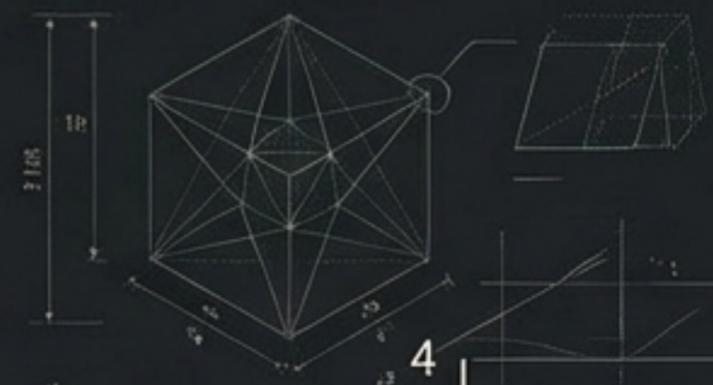
Trust Coordinate: 信頼の参照座標



Improvement Memory: 改善履歴の保存

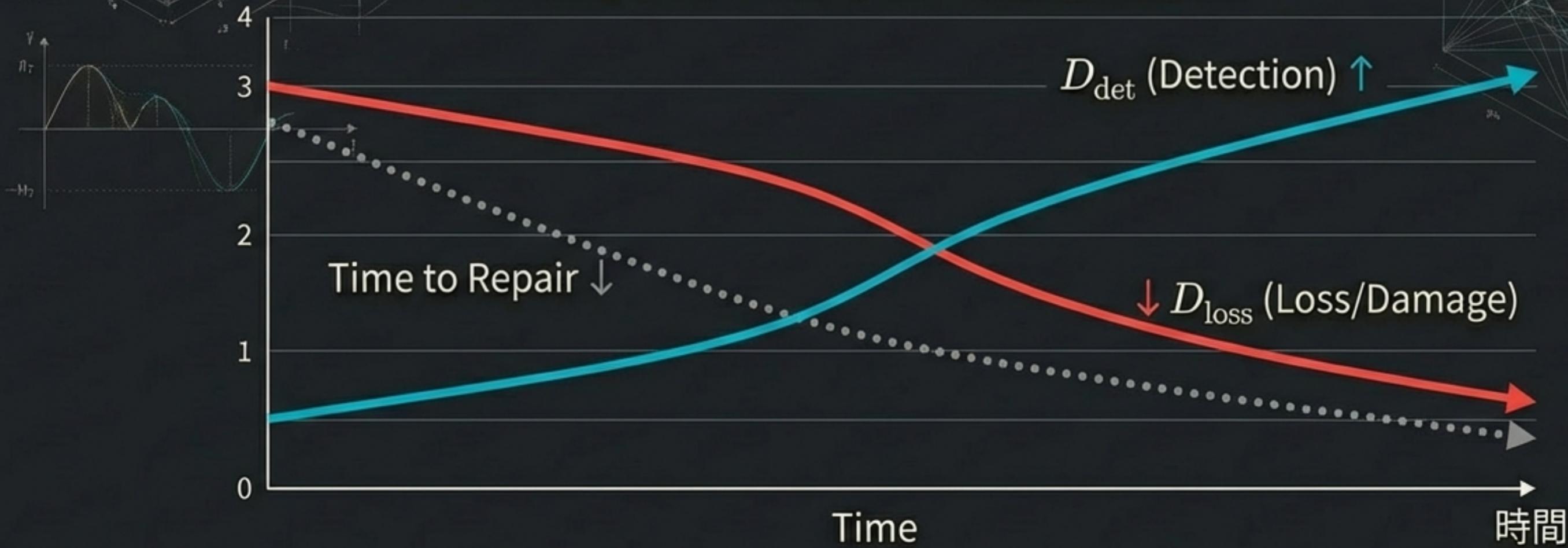
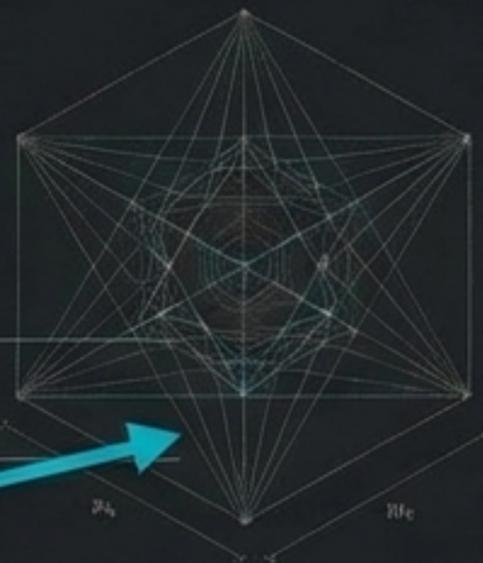
# 健康な免疫反応の観測

Healthy Immune Response (健康な免疫反応)



$$H_s = \left( \frac{112}{15} - 14.00 \right)$$

$$H_c = \left( \frac{162}{15} - 14.00 \right)$$



免疫が機能すると、一見して「問題（検知）」が増えたように見える。  
しかし、実害は減る。「検知の増加は悪化ではない。解像度の上昇である。」

# 危険な兆候：潜伏の観測

Unhealthy / Latent State (危険な兆候)

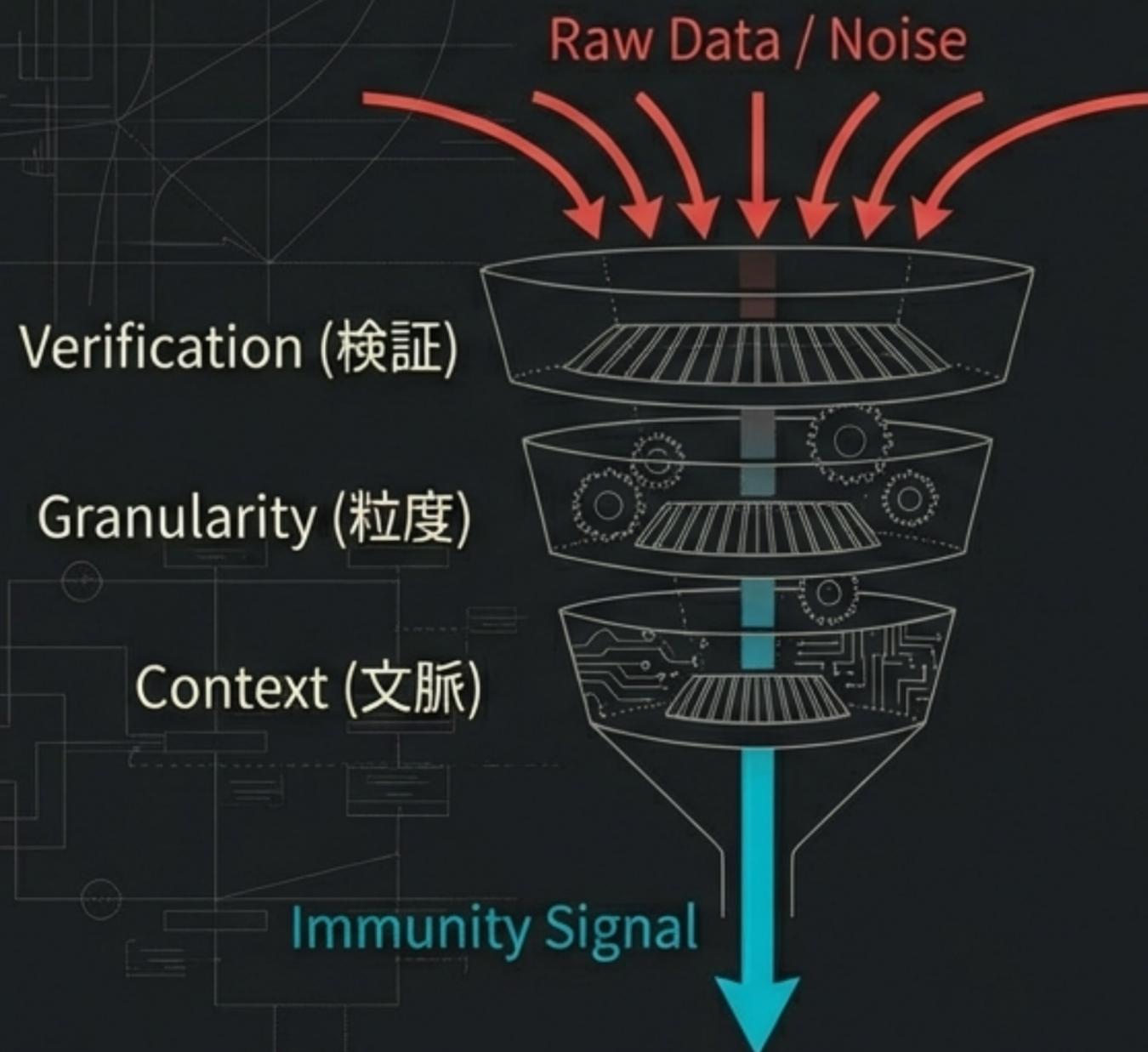


$D_{\text{det}}$  (Detection)

$D_{\text{loss}}$  (Potential Loss)

「何も起きていない」は「順調」ではない。  
報告がない =  $R$  (修復ノード) が死んでいる。  
状態：臨界点への接近。

# 公開の「帯域」設計



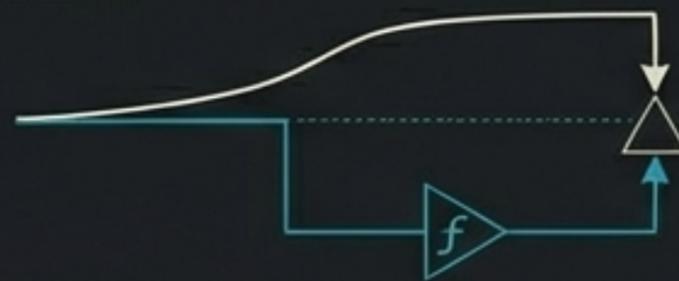
無秩序な公開は免疫にならない。  
ノイズ ( $K$ 超過) になる。

1. 検証可能性：第三者が「なぜ」を追えるか。
2. 粒度設計：認知帯域を超えない階層化。
3. 攻撃化の防止：差分を「改善の種」として扱う。

# AI・社会構造への応用

AI Context

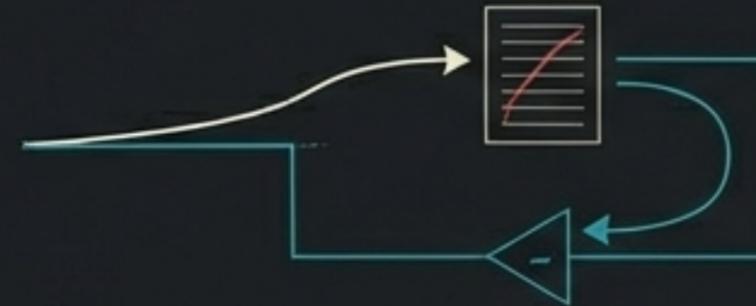
Hallucination  $\approx$  Deviation ( $D$ ).



Solution: Origin Traceability  
& Process Diff.

Social Context

Scandal/Flaming  $\approx$  Deviation ( $D$ ).



Solution: Deviation Ledger &  
Self-Correction.

Common Physics:  $S = U \times R \times H$

Based on AI構造監査レポート & NCL- $\alpha$

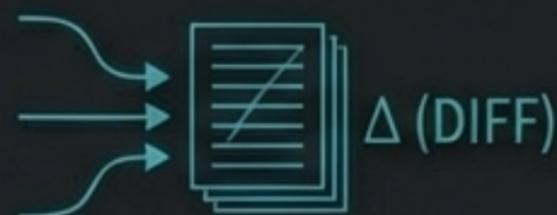


# 構造的移行へのステップ

$$A = U \times R \times (p, \ell)^2$$

$$D = \begin{bmatrix} \text{utd} & 1 & a_{\text{det}} & 0 & \text{fst} & 2 \times 1 \\ 0 & 1 & -k_1 & 1 & \text{err} & k_{\text{det}} \end{bmatrix}$$

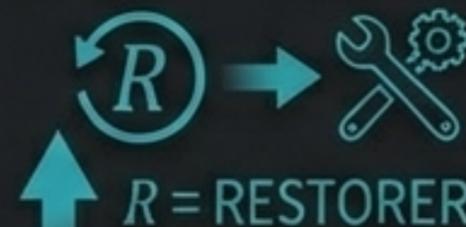
Build the Ledger  
(事実・差分の記録開始)



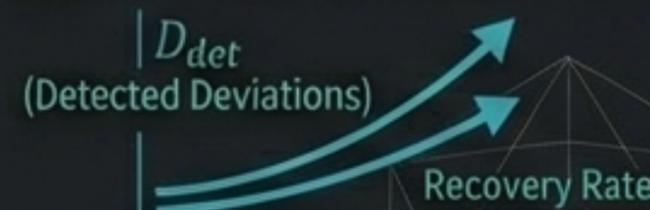
Stop the Bleeding  
(厳罰の停止・免責)



Redefine R  
(責任者を「直す人」へ)



Monitor Flow  
( $D_{det}$  増加と回復速度の観測)



Step 4

Step 3

$$\text{Restorability} = \tau_{det} - \pi_{det}$$
$$\tau_{det} \times \tau_{rec} f_a(D_{det} - \pi_{det})$$

$$\hat{f}(a) = \left( \left( \frac{(1 - \tau_{det})}{\delta^{-k_1}} \right) \right)_{l \in \mathbb{Q}}$$

Step 2

Step 1

# 新しい社会契約



私たちは「人の善意」で社会を守ることを諦める必要がある。  
 人を信じるな。構造を信じる。

# 結論：回復の物理

免疫とは、安心の演出ではない。  
傷を直す「速度」の物理である。

罰を捨て、差分を抱け。

それが、壊れない構造への唯一の道である。