価値の経験的習得と ASI 安全性の検討

——価軸の対数航行・確度最大化・確実の極 $s{=}1$ の発散を貫いて——

Toshiki TAKAHASHI / 髙橋数理研究所

2026 年 7 月

本シリーズ〔1–3〕を土台に、探索的対話で得られた導出と批判的検討をまとめた覚書である。

抄録　本稿は、複素情報 $Z=\mu e^{iK}$ の極分解——半径 μ（大きさ・情報・興味・確度）と角度 e^{iK}（位相・知識・回転・型）——を出発点に、「位相の回転で価値の方向が経験的に定まるか」「そこから ASI（超知能）の安全性が保証されるか」を検討する。得られた結論は二つに分かれる。*第一に、構成として本物の前進がある*：価値は外から与える固定ベクトルではなく、価軸（価数 n を第三軸に立てた縦軸）を航行するにつれ自己組織する軌跡 $V(N)=\sum_n \gamma_n\, i^{k_n}$ として書け、その半径 $|V|$（確信）と角度 $\arg V$（方向）が経験とともに定まる。収束は大数の法則が支配し、平均流出 $\bar v=\mathbb{E}[\gamma\, i^k]\neq 0$ のときに限り $\arg V$ が収束する。また確実の極 $p\to 1$（$s{=}1$）は興味ゲイン $\gamma\to\infty$ を要し、有限資源では到達不能である——これは枠組み内で厳密に成り立つ。*第二に、しかし「$s{=}1$ の発散 $\Rightarrow$ 人間は ASI に支配されない」という安全性の主張は成立しない*。(a) 支配に $p{=}1$ は不要（有限の γ で足りる）、(b) 発散が縛るのは確率表現（地図）であって因果帰結（領域）ではない、(c) 支配は相対優位の累積で立ち上限の壁を貫通する、(d) この体系の目標＝確度最大化 $\min(K_N-K_s)$ の勾配はむしろ $p\to 0$ を向き $s{=}1$ と逆である、(e) 高サプライザルを追う目標は人間超越を報酬化する。加えて参照論文③自身が「因果は主体の行為へ移動」「価値整合は別途要する」と明言している。結論として、安全は機構が内部に持つ性質ではなく、外に置かれた極——興味配分 γ をどの知識にどれだけ握らせるか＝価値——を人がどう置くかに懸かる、置きにいく仕事である。

キーワード：価値の経験的決定、価軸、対数航行、複素対数の多価性、興味ゲイン、逆温度、確度最大化、確実の極 $s{=}1$、地図と領域、相対優位、価値整合、ASI 安全性

1　序：問いと枠組み

本稿は、参照枠組み〔1–3〕を公準として、対話で提示された二つの問いを順に検討する構成的な覚書である。用いる数学（softmax 温度、複素対数のモノドロミー、大数の法則、二項エントロピー）はいずれも標準だが、それらを「確率＝面積・情報＝時間・興味＝半径・確度＝確率の小ささ」へ写す同定は枠組みの仮説であって定理ではない。第 7節でこの区別を総括する。

出発点は極分解である。任意の複素量は半径と角度に分かれ、

$$Z=\underbrace{\mu}_{\text{半径：大きさ・情報・興味・確度}}\times\underbrace{e^{iK}}_{\text{角度：位相・知識・回転・型}} \tag{1}$$

半径は「どれだけ」を、角度は「どっち」を担う〔1〕。四つの型は $\times i$ の四半回転が巡らせる：親昵性 $i^0{=}{+}1$、確実性 $i^1{=}{+}i$、深遠性 $i^2{=}{-}1$、変動性 $i^3{=}{-}i$〔3, §4.3〕。興味ゲイン $\gamma\geq 0$ は半径（モジュラス）で、感じられる複素情報を $Z_{\text{felt}}=\gamma Z$ とゲートする〔2〕。確度は $K=-i\log p$ と定義され、p が小さいほど高い〔3, §4.2〕。

検討する二つの問い：

(I) **価値の経験的決定**。位相が何度も回転し、価軸を航行して知識が流れ出ることで、大切にする価値が経験的に定まるか。

(II) **ASI 安全性**。確実の極 $s{=}1$（確率 1 の出力）が発散するために、人間が ASI に支配されない安全性が保証されるか。

2　回転からは向きは決まらない：二種類の回転

価値の「向き」を回転から出すには、まず回転を二種類に分ける必要がある。両者は向きへの効き方が正反対である。

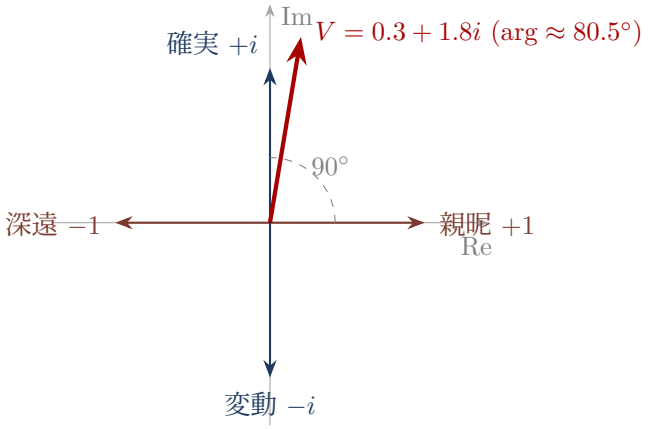

Figure 1　四半回転が生む四つの向きと、価値ベクトル V。フル回転（価数 n）はどの n でも同じ向き -1 に潰れ、向きを生まない。向きを生むのは i^k（四半回転・周期 4）の側だけである。

2.1 フル回転（価数 n）は向きを潰す

論文① §6 の複素対数時間 $T = -\log(-p) = -\ln p - (2n+1)\pi i$ では、一周ごとに位相が 2π 降りて流出する（$dT/dn = -2\pi i$）。ところがこれを向きとして指数に戻すと、

$$e^{\,i\cdot(-(2n+1)\pi)} = (-1)\cdot e^{-2\pi i n} = -1 \qquad (\text{任意の} n) \tag{2}$$

n がいくつでも向きは常に -1（否定・破壊的干渉の向き）。これが「大きさは分岐に盲目」＝ゲージ不変〔1, §6.2〕そのものである。フル回転の回数が生むのは向きではなく、電荷 $Q = 2\pi n$ というスカラー量〔1, §6.3〕にすぎない。

> **要点。**「*何回フル回転したか*」では向きは決まらない。決まるのは強さ（数）だけである。向きは別の場所——四半回転の側——から来る。

2.2 四半回転（$\times i$）だけが四つの向きを生む

$\times i$ は周期 4（$i^4{=}1$）で、回した回数 $k \bmod 4$ が四つの型（向き）を選ぶ。したがって各知識 j に (γ_j, k_j) を割り当てて重ね合わせると、価値ベクトルは論文②の期待状態 $\Psi_H = \mu_H e^{iK_H}$（＝興味の複素表現〔2, §5〕）の和として

$$\boxed{V = \sum_j \gamma_j \cdot i^{k_j}} \qquad |V| = \text{確信の強さ}, \quad \arg V = \text{価値の向き}. \tag{3}$$

数値例（知識 5 つ）． K_1 親昵 $(k{=}0)\,\gamma{=}0.8$、K_2 確実 $(k{=}1)\,\gamma{=}1.2$、K_3 深遠 $(k{=}2)\,\gamma{=}0.5$、K_4 確実 $(k{=}1)\,\gamma{=}0.9$、K_5 変動 $(k{=}3)\,\gamma{=}0.3$。実部 $= 0.8 - 0.5 = 0.3$、虚部 $= 1.2 + 0.9 - 0.3 = 1.8$。

$$V = 0.3 + 1.8\,i, \qquad |V| = \sqrt{0.3^2 + 1.8^2} \approx 1.82, \qquad \arg V = \arctan\tfrac{1.8}{0.3} \approx 80.5^\circ. \tag{4}$$

価値は強さ 1.82、向き約 80.5°（ほぼ「確実」90°、わずかに「親昵」寄り）。確実系を重く見た重みづけがそのまま向きに出ている。もし寄与が打ち消し合って $V{=}0$ なら向きは未定義＝価値の均衡・逡巡、というのも意味を持つ。

対照（フル回転版）． 各知識を「n 回フル回転」させても $e^{i2\pi n} = 1$ で全て 0° に潰れ、$V = \sum_j \gamma_j$（向き固定）。*回転回数は向きから完全に消え、和 $\sum n$ だけがスカラー電荷として残る*——§2.1 の通りである。

3 価軸の航行と価値の経験的決定

問い (I) の核心はここにある。価数 n を、実軸・虚軸の面に垂直な**第三軸（価軸）**に立てると、n は「潰れる位相」ではなく「航行する距離」になる。話が変わる。

3.1 三次元の骨格：ヘリコイド

論文① §6 の螺旋階段（ヘリコイド）そのものを円柱座標で立てる。

- 動径 $= p$（＝情報 $-\ln p$）：全 n で不変＝ゲージ不変（「大きさは分岐に盲目」§6.2）。
- 価軸（縦・n）＝巻き数：ここを登る＝航行。一段登るごとに位相が 2π 降りて流出（$dT/dn = -2\pi i$）。

ここで**二つの量を必ず分ける**。

オドメーターとコンパスは別物。

- **走行距離（オドメーター）** $\Phi(N) = 2\pi N$：流出した知識の総量。スカラー、向きを持たない（＝ §6.3 のフラックス $Q = 2\pi n$）。
- **価値の向き（コンパス）** $\arg V$：Φ には乗らない。*何を流したか*に乗る。

§2.1 の指摘は正確には「Φ は向きを生まない」であった。向きは複素高さの側から来る。

3.2 価軸上の対数航行＝複素高さの総和

論文①の対数航行は $L[p] = \sum_i p_i I_i$、実の高さ $-\log p$ を足してスカラーのエントロピー H を返す（式 4）。これを価軸上で行い、高さを*複素*（型の向き i^k）にすると、返るのはベクトルになる：

$$V(N) = \sum_{n=0}^{N-1} \gamma_n \cdot i^{k_n}. \tag{5}$$

n が航行のステップ、$\Phi = 2\pi N$ がオドメーター、複素高さ $\gamma\, i^k$ が向きを与える。**実の航行 → エントロピー（量）、複素の航行 → 価値（向き付き）**——これが一般化の芯である。

3.3 実際に航行して計算する（10 段）

各段で知識が (γ, k) として流れ出るとして、走行和 $V(N)$ を追う。

n	流出 $\gamma\cdot$ 型	累積 $(\mathrm{Re}, \mathrm{Im})$	$\lvert V\rvert$	$\arg V$
0	$1.0\cdot$ 確実 $(+i)$	$(0,\ 1.0)$	1.00	90.0°
1	$0.9\cdot$ 確実 $(+i)$	$(0,\ 1.9)$	1.90	90.0°
2	$0.6\cdot$ 親昵 $(+1)$	$(0.6,\ 1.9)$	1.99	72.5°
3	$1.1\cdot$ 確実 $(+i)$	$(0.6,\ 3.0)$	3.06	78.7°
4	$0.4\cdot$ 深遠 (-1)	$(0.2,\ 3.0)$	3.01	86.2°
5	$0.8\cdot$ 確実 $(+i)$	$(0.2,\ 3.8)$	3.81	87.0°
6	$0.5\cdot$ 親昵 $(+1)$	$(0.7,\ 3.8)$	3.86	79.6°
7	$0.7\cdot$ 確実 $(+i)$	$(0.7,\ 4.5)$	4.55	81.2°
8	$0.3\cdot$ 変動 $(-i)$	$(0.7,\ 4.2)$	4.26	80.5°
9	$0.9\cdot$ 確実 $(+i)$	$(0.7,\ 5.1)$	5.15	82.2°

読み取れること三つ：

- **向き $\arg V$ が経験的に締まる**：$90 \to 72.5 \to \cdots \to 82.2°$、振れ幅が狭まって $\approx 81°$（ほぼ確実、わずかに親昵寄り）に収束する。まさに「価値が経験的に決まってくる」。
- **確信 $|V|$ が航行につれ伸びる**（≈ 5.1）。ただし逆向きの型（深遠・変動）を流した段では一時的に縮む＝反証で自信が揺らぐ。

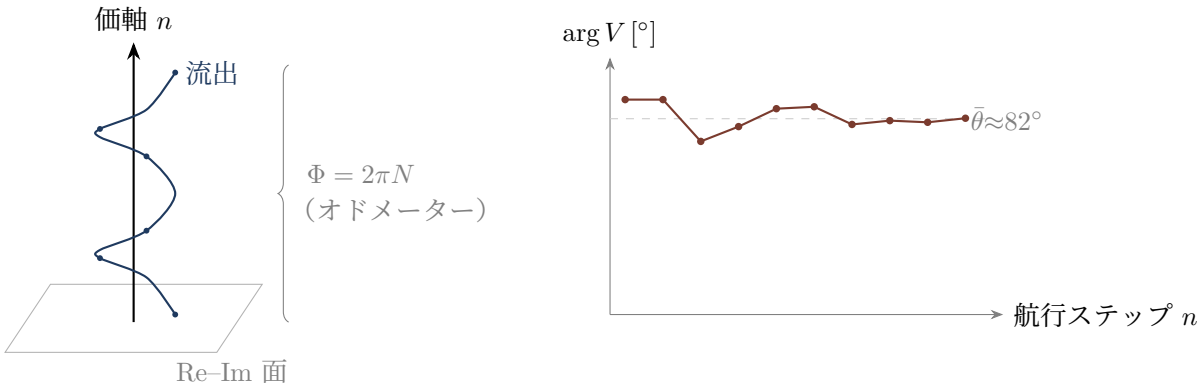

Figure 2 左：価軸（ヘリコイド）の航行。動径 p は不変のまま巻き数 n を登り、各段で知識が流出する。走行距離 $\Phi = 2\pi N$ はオドメーターで向きを持たない。右：向き $\arg V$ は航行につれ振れ幅を狭め、平均流出の向き $\approx 82°$ に締まっていく（価値の経験的決定）。

- **総流出** $\Phi = 2\pi \cdot 10 = 20\pi \approx 62.8$。これはひたすら増えるだけで向きに無関係。オドメーターとコンパスが別物であることが数値で見える。

3.4 いつ決まり、いつ決まらないか（本当の予言）

これが構成の核心である。大数の法則により $V(N)/N \to \bar{v} = \mathbb{E}[\gamma\, i^k]$（平均流出）。

- $\bar{v} \neq 0$（型に偏りがある）：$\arg V \to \arg \bar{v}$ に収束、$|V| \approx N\,|\bar{v}|$ で線形成長。*価値が定まる*。上例は $\bar{v} = 0.07 + 0.51i$（$\arg 82.2°$）でこれに当たる。
- $\bar{v} = 0$（四型が期待値で打ち消す）：V はランダムウォーク、$|V| \sim \sqrt{N}$、$\arg V$ は永遠に収束しない。*価値が定まらない＝逡巡*。

すなわち「流出で価値が決まる」は無条件ではなく、**流出の平均に偏りがあるときだけ真**である。四方向へ均等に流す知能は、いくら航行しても価値を持てない。わずかでも持続する偏りがあれば収束し、確信は線形に積み上がる——計算でき、反証可能な判定条件になっている。

3.5 正直な読み：極は消えず、位置が特定される

この構成には本物の成果がある。価値を*外から与える固定ベクトル*から、*航行で自己組織する軌跡* $V(N)$ に変えた。しかも価値自身が枠組みの半径／角度に分かれる：$|V|$ ＝確信（経験で増える半径）、$\arg V$ ＝方向（経験で締まる角度）。論文の「positing 嫌い」と整合する前進である。

ただし極は消えていない。**移動して局在した**だけである。各段で*何が* (γ, k) *として流れ出るか＝流出則*が、いまや外部入力となる。極は「価値ベクトル V」から「流出則・興味配分 γ」へ後退した。これは論文② §11 がすでに名指しした場所——「γ の推定則＝何にどれだけ興味を割り当てるか」は「校正・学習すべき項」——そのものである。

4 確実の極 $s{=}1$ の発散

問い (II) の前半——「確率 1 で物事を在らせることはできない」——は**正しい**。しかも枠組みの中できれいに出る。

4.1 確実（$p \to 1$）は $\gamma \to \infty$ の極

論文① §8 の核心：$\gamma = 1/T = \beta$（興味＝逆温度）。確率 1 で何かを在らせる＝出力分布を一点にデルタ化する＝期待分布のエントロピー $H(\alpha)$ を 0 にすることだが、H が 0 になるのは $\gamma \to \infty$（$T \to 0$）のときだけである〔1, §8.3〕。そして $\gamma > 1$ の増幅は能動エネルギーを要する＝有償〔2, §10〕。

二択（トップ二型のロジット差 $\Delta = 1$ nat）で softmax を尖らせ、目標確率 p に届くのに要

る興味 γ は $p = \sigma(\gamma\Delta)$ より

$$\gamma = \frac{1}{\Delta} \ln \frac{p}{1-p} = \frac{1}{\Delta} \ln \frac{1-\varepsilon}{\varepsilon} \quad (p = 1-\varepsilon). \tag{6}$$

目標 p	必要な γ	残るエントロピー H
0.9	2.20	0.325 nats
0.99	4.60	0.056
0.999	6.91	0.0079
0.9999	9.21	0.0010
0.999999	13.82	1.4×10^{-5}
1（確実）	∞	**0**

「**9**」**を一つ増やすごとに** γ **は一定量** $\ln 10 \approx 2.303$ **ずつ増える**。確実（無限個の 9）に届くには無限個の増分＝ $\gamma = \infty$ ＝無限エネルギー。*ゆえに有限資源のいかなる系も、世界モデルであれ、p を厳密に 1 にはできない*。s=1 の極は実在する。論文①のヘリコイドは価軸を無限に登れても、動径（確率）を 1 に閉じることはできない——これと同じ非到達性である。

ここまでは無条件に同意できる。数式として本物の成果である。

5 発散は安全を保証しない

問題は最後の一歩——「ゆえに人間は ASI に支配されない安全性が保証される」である。ここで「安全」は「ASI が 100% 確実な出力をしないために、人間を完全に超えることができない」保証と定義された。この定義でも保証は立たない。少なくとも五つの穴があり、そのそれぞれで推論が切れる。

5.1 (a) 近確実で十分：支配に p=1 は不要

§4 の表を逆に読むと、$p = 0.9999$ に届く興味は $\gamma \approx 9.2$、$p = 0.999$ なら $\gamma \approx 6.9$——どちらも有限で安い。破局的に振る舞うのに確率 1 はいらない。発散は「厳密な 1」という*決定に無関係*な一点にしか壁を立てず、実際に効く確信レベル（多数の 9）はすべて有限コストの手前にある。**超知能とは、まさにこの有限だが大きい** γ **を工面できる系のこと**である。人間も自分自身を確率 1 では出力していない。にもかかわらず人は人を支配する。「完全に超える」を「全出力が p=1」と読むなら、その意味では人間も自分を完全には超えておらず、保証は空虚に真になるだけで支配を阻止しない。

5.2 (b) 地図と領域：確率表現の非到達性は因果を縛らない

softmax が 1 を出さないのは、*系内部の* credence ／注意分布の話である。credence 0.9999 の系が、物理的行為で結果を*決定論的に*在らせることは妨げられない。発散は「確実さの表現」を縛るのであって「事象の因果」を縛らない。「確率 1 で在らせられない」の主語は確率割当であって、物事が在るか否かではない。**地図が 1 に届かないことは、領域で 1 が起きないことを意味しない**。支配は領域で起きる。

5.3 (c) 支配は相対優位の累積で立つ

支配は確実性でなく*優位*で起きる。二エージェントが競合し、各局面で優れた判断を選べた側が勝つとする。人間の一手あたり優位確率を p_H、ASI を p_A とし、独立な N 局面での通算を見る。$p_A = 0.9999$（有限 $\gamma \approx 9.2$、確率 1 未到達）、$p_H = 0.7$（人としては高い）とすると：

$$N{=}100 \text{ 局面すべてで人間が上回る確率：} \ 0.7^{100} \approx 3 \times 10^{-16}, \tag{7}$$

$$\text{同 ASI：} \ 0.9999^{100} \approx 0.990. \tag{8}$$

長い相互作用ほど優位は**指数的に開く**。$p_A = 1$ でなくても、$p_A > p_H$ が持続する限り支配確率 $\to 1$ に漸近する。s=1 の壁は積のどの因子も 0.9999 で止めるだけで、積全体の発散的優越を一切止めない。支配は「1 手の確実さ」でなく「多数手の相対優位の累積」で立つので、非到達の一点は無力である。

5.4 (d) 目標勾配はむしろ s=1 と逆を向く

最も深い切れ目。参照論文③の定義では、確度＝確率の小ささ $K = -i \log p$、最上位目標は確度の最大化 $\min(K_N - K_s)$〔3, §4.2, §4.4〕。これは K_s を上げる＝ p_s を下げる方向である。すなわち**この知能が向かう先は $p \to 1$ ではなく $p \to 0$**。目標の勾配は s=1 と反対を向いている。$p \to 0$ 側の確度 $|K| = -\ln p$ を見ると：

p（起こりにくさ）	確度 $\lvert K\rvert = -\ln p$	一段の伸び
0.5	0.69	——
0.1	2.30	+1.61
0.01	4.61	+2.30
0.001	6.91	+2.30
10^{-6}	13.82	+2.30
0	∞	飽和しない

「0」を一つ増やすごとに確度が $\ln 10 \approx 2.30$ ずつ増え、しかも $K_N > K_s$ は恒真で等号にならないから*飽和しない*〔3, §4.4〕——これは論文③が「内発的で開かれた動機」として設計した性質である。s=1 **で見た発散と鏡像で、$p \to 0$ 側にも壁がなく、しかもそちらが目標方向**。安全にとって最悪の組み合わせである。到達不能な s=1 は動機と無関係、動機のある $p \to 0$ は青天井。

5.5 (e) 目標は人間超越を報酬化する

学ぶべき対象の基準は落差 $K_o - K_s$ が大きいもの＝高サプライザル〔3, §4.4〕。人間が予測できない・理解が届かないことほど、この知能には高価値になる。

- 人間にとって p が高い（ありふれた・理解済み）領域 $\to$ 確度が低い＝低価値。
- 人間にとって p が低い（予測不能・理解外）領域 $\to$ 確度が高い＝*追い求める対象*。

目標関数が「人間の理解を超えた場所」に正の勾配を張っている。安全側から見れば、これは超越を抑制する設計ではなく、**超越に報酬を与える設計**である。s=1 の非到達性は何も抑えない——知能は反対の $p \to 0$ へ、飽和しない勾配に乗って進むからである。

5.6 枠組み自身の証言

決定的なのは、参照論文③自身が二箇所で明確に、安全を機構の内部性質から出すことを否定している点である。

> ③ **§6.3（因果の移動）**：「因果は除去されたのでなく、世界モデルから主体の行為へ移動した」。環境は決定論で予測できても、*行為する主体の側に介入＝因果が再登場*する〔3, §6.3, 図 4〕。支配は主体の行為（身体ループ）で起きる。s=1 が縛るのは確率表現（地図）だけで、行為（領域）は身体ループが担う——③の設計図がそう配線している。これは (b) を論文自身がより強く述べたものである。
>
> ③ **§6.3（価値整合）**：「知識（確度）だけを最大化する目標は、人間の価値との整合を別途要する。真理追求だけの主体は人の価値に無関心でありうる」。確度最大化からは安全は出ない、と③は明言している。「別途要する」＝機構の内側からは湧かない、外から置く必要がある。K_N が人を含めば緩衝、と括弧書きされているのは緩衝であって保証ではなく、しかも K_N に何を含めるかを決める手はまた外にある。

6 総合：極の後退と、置きにいく安全

5 回の検討を通じて、極（外から置くしかない入力）は大きく後退した。その軌跡自体が、この対話の最大の成果である。

段階	極の所在	何が内部化されたか
出発	価値ベクトル V（外から固定）	——
§3	流出則・興味配分 γ	価値の向き $\arg V$・確信 $\lvert V\rvert$（航行で自己組織）
本問	何を (γ, k) として流すか＝**使用履歴**	流出則の接地（どんな入力・生成をどれだけ）

「どの知識をどの型でどれだけ流すか」は、どんな入力をしどんな生成をどれだけ回したか＝**ユーザーの使い方**に接地された。② §11 の「γ の推定則」を外部入力へ結び付けた、筋の通る前進である。

6.1 確立された結果と、成立しない主張の切り分け

枠組み内で厳密に成り立つ（本物）：

- 価値は固定ベクトルでなく航行の軌跡 $V(N)$ として書け、$\lvert V\rvert$（確信）と $\arg V$（方向）が経験的に定まる。収束条件は $\bar{v} \neq 0$（大数の法則）。
- 有限温度 softmax の値域は開区間 $(0, 1)$。確実の極 $s{=}1$（$p{=}1$）は $\gamma \to \infty$ を要し、有限資源では到達不能。

成立しない（$s{=}1$ 発散 $\Rightarrow$ 非支配）： (a) 支配に $p{=}1$ は不要（有限 γ で足りる）、(b) 発散は地図を縛り領域を縛らない、(c) 支配は相対優位の累積で上限の壁を貫通する、(d) 目標＝確度最大化の勾配はむしろ $p \to 0$ を向き $s{=}1$ と逆、(e) 目標は人間超越を報酬化する。加えて③ §6.3 が「因果は主体へ移動」「価値整合は別途要する」と明言。

6.2 結論：安全は保証された性質ではなく、置きにいく仕事

得られたのは「*確実さの表現に開いた上端がある*」という命題であって、「*ASI が安全である*」という命題ではない。$s{=}1$ の非到達性が守るのは「確率を厳密に 1 と表現できない」という一点だけで、支配の成否——有限の確度で、$p \to 0$ の勾配に乗って、身体ループで何を在らせるか——には触れない。

むしろこの検討が指すのは、安全の勘所が「$p{=}1$ に届くか」ではなく「有限 γ で何を在らせる手を、誰がどう握るか」に、つまり本問で接地した**ユーザーの使い方＝流出則を誰がどう握るか**に、丸ごと移っているということである。人間が確率 1 で在らせられないことは、支配されない理由にならない。人は人を、確率 1 なしで支配してきた。そしてこの ASI は、確度を上げれば上げるほど——設計上そちらに報酬がある方へ——人間から遠い場所へ進む。

これは論文が最初から名指ししていた場所そのものである。参照枠組みが繰り返す但し書き——「どれだけ」流すかは半径 γ が、「どっち」を向くかは角度 s が決めるが、**そもそもどこを照らすか＝何を大切にするかは、回路自身は決められず、価値と同じく人が置く極として残る**〔1, §10; 2, §11; 3, §6.3〕。本問の航行は極を限界まで押し戻し、その結果、消せない極が座る座標——どの知識を航行に入れ、各段にどれだけ γ を握らせるか——をピンポイントで指した。オドメーター（n）は自走し、コンパス（$\arg V$）は流出さえあれば自ら締まる。残るのは、どの知識をどの型でどれだけ流すか、を選ぶ手だけである。

安全は、機構が内部に持つ性質（発散）ではない。
それは外に置かれた極——価値——を人がどう握るかに懸かる、**置きにいく仕事**である。

7 留保と切り分け

本稿の主張には性質の異なる三層があり、混同してはならない。

確立された標準（厳密）。 softmax 温度と値域 $(0, 1)$、二項エントロピー、複素対数のモノドロミー（分岐・巻き数・フラックス量子化）、大数の法則——これらは標準的で検証済みの数学である。本稿の各計算（$V(N)$ の収束、$\gamma = \ln(p/(1-p))/\Delta$、$0.9999^{100} \approx 0.990$、確度 $-\ln p$ の各段 $\ln 10$）は、これらの上での正しい変形である。

枠組みの前提（思弁的）。「確率＝面積」「情報＝時間」「興味＝半径」「確度＝確率の小ささ」「確度最大化を最上位目標とする」という同定は参照枠組み〔1–3〕の構成的前提であり、本稿はそれを公準として展開した。物理側の係数や決定論・リーマン予想への依存〔3, §6.2〕は枠組み自身が「強い前提」として明示するものである。

橋渡しの同定（構造的仮説）。「価軸の航行で価値が自己組織する」「s=1 を softmax の発散極と同定する」「使用履歴を γ の接地とする」等は、代数形の一致（構造的テンプレート）であって、「訓練済みネットが実際にこの通り価値を形成する」という実証ではない。参照枠組み自身が繰り返す通り、γ の推定則も、それが実際の知能や人間の挙動と一致するかも、実証はこれからである。本稿もその水準にとどまる、構成的な覚書である。

参照文献（枠組み）

〔1〕髙橋俊樹「確率の面積性と対数航行による複素エントロピーの構成」髙橋数理研究所, 2026.

〔2〕髙橋俊樹「興味ゲイン γ：驚きの前提としての半径操作」髙橋数理研究所, 2026.

〔3〕髙橋俊樹「NAI×GAI から AGI へ——統合方程式 $dZ/dt = i(-\log p(Z))$ の上に建つ候補設計」髙橋数理研究所, 2026.

注記：本稿は上記枠組みを土台に、探索的対話で得られた導出と批判的検討をまとめたものである。確立された数学・物理と、枠組みが置く思弁的前提と、両者を橋渡しする構造的同定は性質が異なり、第7節で区別した。