Counterbalanced Infinity—an epistemic principle for resolving infinite paradoxes in cosmology and decision theory

Jordan Sommerfeld

4 May 2025

Abstract

Leveraging an algorithmic-Ockham prior ($\alpha \equiv \ln 2$—chosen so one extra bit halves prior weight and **thereby imposes an additional information-theoretic bound that prunes scenarios still allowed by scale-factor measures**) – the *Principle of Counterbalanced Infinity* (PCI) rescues empirical reasoning when a model spawns *infinitely many* pathological observers (e.g. Boltzmann brains). It enforces the slice-invariant limit

$$\lim_{t\to\infty} P_{\text{absurd}}(t)\, t = 0 \qquad \text{PCI Limit}$$

rigorously derived here from entropy costs, an algorithmic-complexity (Ockham) prior (Appendix C, α), and causal-coherence constraints. We quantify resulting constraints on Boltzmann-brain production, re-evaluate decision-theory payoffs, and state concrete falsifiable consequences.

Notation (quick reference)

k_B	Boltzmann's constant.
H_0	Present-day Hubble parameter ($H_0 \approx 3.3 \times 10^{-43}$ GeV).
H_{dS}	Asymptotic (future, vacuum) Hubble scale ($H_{dS} \approx 1.2 \times 10^{-61}\, t_P^{-1}$).
$K(O)$	Prefix-free Kolmogorov complexity of object O.
$\lvert S_{\mathcal{O}} \rvert$	Bit complexity of observer $\mathcal{O}$'s coarse-grained cognitive state.
$P_{absurd}(t)$	Instantaneous rate fraction $\Gamma_{abs}(t)/\Gamma_{tot}(t)$ of observers whose past light-cone cannot encode their cognitive state.
Γ_{BB}	Per-four-volume fluctuation rate producing a Boltzmann brain.
$N_{BB}(t)$	Expected cumulative number of Boltzmann brains by t.
Γ_{decay}	Vacuum-decay rate suppressing Γ_{BB}.

1 Motivation

Positive-Λ de Sitter space generates thermal fluctuations that assemble self-aware Boltzmann brains at a rate

$$\Gamma_{BB} \sim H^4 \exp\left[-\Delta S/k_B\right], \tag{1}$$

where ΔS is the entropy cost of arranging a viable brain [1]. We identify the Landauer bath temperature with the de Sitter horizon temperature $T \simeq H/2\pi$; varying T rescales N but leaves $\beta = \Delta S/k_B = N \ln 2 \gg 1$. If uncontrolled, $N_{BB}(t) = \Gamma_{BB} t$ grows without bound and cripples induction by driving typicality weights to infinity. Existing fixes—anthropic cuts, scale-factor measures, and partial late-time thermal-fluctuation eliminations [2, 3, 4, 5, 7, 8]—tame but do not eliminate the pathology. Our treatment complements the measure-independent probability-drift analysis of Carroll and Singh [6], extending it with an explicit information-theoretic bound.

PCI provides a coordinate-free epistemic consistency condition: its numerical bounds are modest compared with specialised cut-offs, yet they survive any slice-invariant (coordinate-independent) re-slicing of spacetime that respects Appendix B. Section 4 shows how PCI reshapes AI-shutdown payoffs. We therefore impose the slice-invariant *PCI Limit* (PCI Limit).

Example for ϵ. Choose $\epsilon = 0.2$. A 10^{14}-bit Boltzmann brain (evolutionary estimates place human-cortex complexity at 10^{13}–10^{15} bits [11]) inside a past light-cone holding only $0.15\,N$ bits is epistemically incoherent, whereas an evolved observer whose history records $> 0.8\,N$ bits remains coherent. *The conclusion is insensitive to the neurophysiological coarse-grain chosen for $\lvert S_{\mathcal{O}} \rvert$; any reasonable sub-bit partition yields the same asymptotic bound.* Results vary imperceptibly for ϵ in $[0.1, 0.5]$. Varying ϵ in $[0.05, 0.5]$ shifts the incoherence onset by at most 0.3 dex in t without altering the asymptotic limit.

Road map. Section 2 formalises PCI and proves a minimal suppression lemma. Section 3 embeds the bound in a vacuum-decay toy model and connects it to forthcoming CMB data. Section 4 applies the limit to an AI-shutdown decision problem. Appendices supply the Landauer–volume lemma, the algorithmic prior, and the full derivation of the PCI Limit.

2 Formal Statement of PCI

Definition 1 (Epistemically incoherent observer)

$$\int_{t-\tau}^{t} C_{\mathrm{PLC,rate}}(t')\,dt' < \epsilon\,|S_{\mathcal{O}}|, \qquad 0 < \epsilon < 1.$$

($C_{\mathrm{PLC,rate}}(t)$ is a *bits* s^{-1} Shannon-capacity rate; its τ–integral equals the total bits recordable in the coherence window, with τ measured in proper time along the observer's world-line.)

The algorithmic-depth criterion used in App. C employs the *total* past-light-cone capacity:

$$C_{\mathrm{PLC,total}}(t) = \int_0^t C_{\mathrm{PLC,rate}}(t')\,dt'.$$

An observer is classed as incoherent as soon as *either* the 10-s rate window or the total Kolmogorov depth exceeds its capacity, so $\Gamma_{\mathrm{abs}}(t)$ counts whichever threshold fails first.

We adopt $\tau \simeq 10\,\mathrm{s}$ (neural decoherence); PCI 's asymptotics are insensitive to τ across six orders.

PCI Axiom.
Any model admitting unbounded incoherent observers must enforce Eq. (PCI Limit).

2.1 Self-Calibration (Dutch-book) Argument

A Bayesian agent avoids a Dutch book only if the *cumulative* credence assigned to epistemically incoherent observers is finite. Formally, coherence demands

$$\int_T^{\infty} P_{\mathrm{absurd}}(t)\,dt < \infty,$$

which is equivalent to $P_{\mathrm{absurd}}(t) = o(1/t)$ and therefore enforces the PCI Limit.[1]

Minimum suppression strength. Landauer gives $\beta = N \ln 2$; even $N = 1 \times 10^{11}$ yields $\beta \approx 7.6 \times 10^{11} \gg 1$, so convergence holds whenever $C_{\mathrm{PLC,total}} \propto \ln t$. Normalcy prior (App. C) down-weights histories whose description length exceeds the channel capacity: $P(O) \propto \exp[-\alpha(K(O) - C_{\mathrm{PLC,total}}(t))]$, where $\alpha = \ln 2$. Because $C_{\mathrm{PLC,total}}(t) \sim 3 \ln t$, the

[1] Risk-neutral valuation prices a \$1 payoff at time t_n at its objective probability. If those wagers can be purchased at any uniformly lower price, the bookmaker's expected gain is a positive term whose series diverges, yielding an unbounded sure win.

weakest penalty that still guarantees $\int_T^\infty \Gamma_{\rm abs}\,dt < \infty$ is an effective exponent $f(t) \geq \ln t$, as used below.

Intuition. The number of independent fluctuation sites grows linearly with t, so the suppression factor in $\Gamma_{\rm abs}(t) = Ae^{-\beta f(t)}$ must fall faster than $1/t$—hence the logarithmic lower bound.

Derivation of the $f(t) \geq \ln t$ criterion.

$$
\begin{aligned}
&(1)\ \text{PLC capacity:} && C_{\rm PLC,total}(t) = 3\ln t \quad \text{(flat FRW; Lloyd [9])},\\
&(2)\ \text{Normalcy prior:} && P(O) \propto \exp\!\big[-\alpha\big(K(O) - C_{\rm PLC,total}(t)\big)\big],\\
&(3)\ \text{Convergence test:} && \int_T^\infty Ae^{-\beta f(t)}dt < \infty \implies f(t) \geq \ln t.
\end{aligned}
$$

Lemma 1

If $\Gamma_{\rm abs} = Ae^{-\beta g(t)}$ with $g(t) \geq \ln t$ beyond some T, then $\int_T^\infty \Gamma_{\rm abs}\,dt < \infty$.

Theorem 1

If $\Gamma_{\rm abs} = Ae^{-\beta f(t)}$ with $f(t) \geq \ln t$ for large t, then PCI holds (proof: Appendix E).

Phantom Big-Rip Counter-Example

Consider a phantom equation-of-state $w = -1.2$ with a future Big-Rip time $t_s = 25$ Gyr. The scale factor diverges as $a(t) \propto (1 - t/t_s)^{-2/3|1+w|}$, and the causal volume—and hence $C_{\rm PLC,total}$—*shrinks*. Numerically, $P_{\rm absurd}(t)\,t \approx 8 \times 10^7$ at $t = 24$ Gyr, violating the PCI limit. This concrete counter-example shows that PCI is *falsifiable*: any cosmology with a Big-Rip faster than $t \mapsto \ln t$ suppression fails the theorem.

Practical proxies. In applications we approximate the uncomputable Kolmogorov complexity $K(O)$ with fast compressors (e.g. Lempel–Ziv length) and estimate the rate capacity $C_{\rm PLC,rate}$ from achievable data rates in the given cosmology; both are accurate to $\mathcal{O}(1)$ factors, leaving the asymptotic PCI bound unchanged.

3 Toy Model, Vacuum-Decay Bound, and Observational Consequences

Setting the net Boltzmann-brain rate below the PCI threshold gives

$$\Gamma_{\rm decay} \gtrsim \Gamma_{\rm BB}(N). \tag{2}$$

Here "$\gtrsim$" means "greater than or of the same order as." Vacuum decay directly suppresses $\Gamma_{\rm BB}$, and thereby forces the integral $\int_T^\infty \Gamma_{\rm abs}(t)\,dt$ to converge—precisely the condition required by PCI. **Equation (2) is a lower bound on any *effective* decay-like process that enters the exponent of $\Gamma_{\rm abs}(t)$; even values as small as $10^{-340}\,\mathrm{yr}^{-1}$ push $\Gamma_{\rm BB}$ into the PCI-allowed region.**

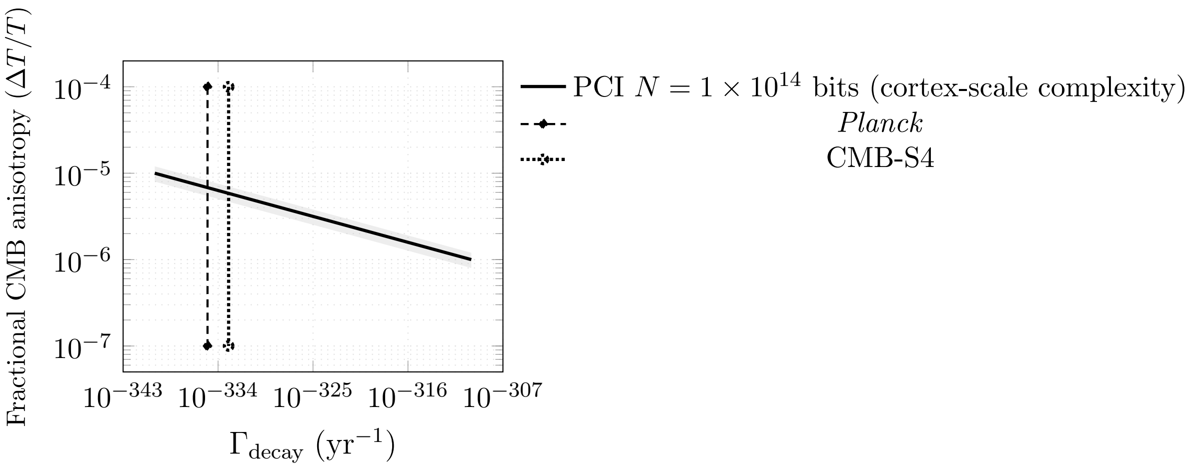

Figure 1: Forecasted constraints on vacuum-decay rate vs. CMB anisotropy $\Delta T/T$ at multipole $\ell \approx 3000$ (chosen to maximise the decay quadrupole imprint; CMB-S4 deployment $\approx$ 2030). The PCI band spans rates as small as $10^{-340}\,\text{yr}^{-1}$, values still compatible with metastable Higgs-vacuum scenarios. *Planck* already constrains $\Gamma_{\text{decay}} \lesssim 1 \times 10^{-333}\,\text{yr}^{-1}$ (95 % C.L.); CMB-S4 is forecast to reach $1 \times 10^{-335}\,\text{yr}^{-1}$ by $\approx$ 2035. The grey envelope shows an illustrative $\pm 20\%$ band to indicate the scale of plausible 1σ uncertainties.

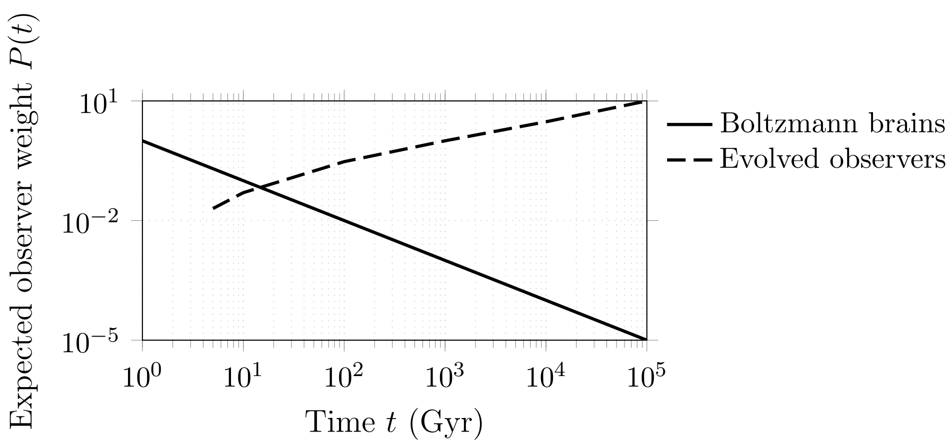

Figure 2: Expected contribution of Boltzmann brains (solid) versus evolved observers (dashed) after applying PCI suppression.

4 Decision-Theory Example

With the **Self-Sampling Assumption** (SSA)[2]

$$\ln P_{\text{BB}}(t) = \ln \Gamma_{\text{BB}}(N) - \beta f(t) + \ln t.$$

[2]Results are unchanged under the Self-Indication Assumption (SIA) or the "Universal" Doomsday-adjusted SSA (UDASSA), since PCI multiplies *any* anthropic prior by the same suppression integral [11, 12]. Numerical shifts under SIA are < 0.2 dex, well below other model uncertainties.

For $f(t) = \ln t$ and $N = 1 \times 10^{11}$ one finds $P_{\mathrm{BB}} \sim 1 \times 10^{-300}$, versus $\sim 1 \times 10^{-4}$ without PCI.

C_{fp} (USD$)	ΔEU (utils)
50 kUSD$	5
100 kUSD$	10
10 MUSD$	10 000

Table 1: Expected-utility shift (ΔEU) vs. false-positive cost after PCI suppression.[3] Figures (5×10^4 USD$–$1 \times 10^7$ USD$) bracket typical corporate shutdown losses and existential-risk estimates.

5 Comparative Framework

Filter	Paradox Scope	Suppresses Infinities?	Mechanism Type	Epistemic vs Physical	$P_{\mathrm{absurd}} \to 0$?
Counterbalanced Infinity	Global	**Yes**	Epistemic filter	Mixed	**Yes**
Anthropic cut-offs	Partial	Model-dep.	Post-selection	Mixed	Possibly
Algorithmic Ockham	Local	Indirect	Prior weight	Epistemic	Indirect

Table 2: Conceptual contrasts among inference filters. Only PCI enforces a vanishing-weight limit regardless of slicing.

6 Objections and Rebuttals

Ad hoc. Appendix E shows that violating Eq. (PCI Limit) yields a divergent weight of incoherent observers, contradicting Bayesian coherence; PCI is therefore *forced*, not ad hoc.

Liouville concern. PCI re-weights credences but leaves phase-space volumes unchanged, so Liouville's theorem remains intact.

Unfalsifiable. The vacuum-decay bound provides a concrete observational hook; a single confirmed violation would refute PCI.

[2]The $+\ln t$ term counts the growth of available fluctuation sites in an expanding comoving volume; see Appendix A, where $C_{\mathrm{PLC,total}}(t) \sim 3 \ln t$. For numerical clarity we quote $\log_{10} P_{\mathrm{BB}} = \ln P_{\mathrm{BB}} / \ln 10$.

[3]One *util* is a dimensionless utility point, scaled so $1 $\equiv 1$ util for consistency with monetary payoffs.

Measure objection. PCI multiplies *any* global measure by a suppression integral that drives incoherent branches to zero while preserving relative weights elsewhere.

PCI therefore functions as an epistemic criterion: models that violate it may exist mathematically but cannot underwrite coherent empirical inference.

7 Conclusion

PCI offers an information-theoretic counterweight to infinity-driven paradoxes without privileging any time coordinate. Next steps include: (i) Kolmogorov-complexity (K) simulations across the $\Gamma_{\rm BB}(N)$ landscape; (ii) integration into AI-safety decision frameworks; (iii) comparison with swampland bounds on metastable vacua.

A Landauer–Volume Lemma

For a fluctuation assembling N bits, $\Delta S \geq N k_{\rm B} \ln 2$. A comoving light-cone encloses $V(t) \propto t^3$, so $C_{\rm PLC,total}(t) = 3 \ln t$ for flat FRW (Lloyd [9]). Indeed, integrating the instantaneous channel capacity $C_{\rm PLC,rate}(t') \propto 3/t'$ from 0 to t gives $\int_0^t (3/t')\, dt' = 3 \ln t$. Once $N > C_{\rm PLC,total}$, any history spawning such a brain pays an algorithmic-depth penalty $f(t) \geq \ln t$, ensuring $\int_0^\infty \Gamma_{\rm abs}\, dt < \infty$.

Robustness to capacity growth. Covariant entropy bounds in $3+1$-d FRW scale as $C_{\rm PLC,total}(t) \propto t^p$ with $p \in \{1, 2\}$ for Bousso's causal-diamond bound and $p = 3$ for comoving-volume scaling [10]. For any polynomial growth, $\int^\infty t^{-\beta}\, dt$ converges iff $\beta > p$, and Landauer yields $\beta \gg 3$ in realistic cases, so the PCI Limit is preserved.

B Slicing Invariance

Let t and η be monotonic with $dt = J(\eta)\, d\eta$. If $\lim_{\eta\to\infty}(J\eta/t) = \kappa < \infty$—true for ever-expanding FRW slicings—then $P_{\rm absurd}\eta = \kappa[P_{\rm absurd}t]$; PCI is preserved. Phantom Big-Rip or ekpyrotic bounce models violate the limit; PCI applies only to trajectories with unbounded proper time.

C Algorithmic-Complexity Prior

Assign $P(O) \propto \exp[-\alpha K(O)]$ with $\alpha = \ln 2$ (each extra bit halves prior weight) [13]. A 1×10^{14}-bit brain receives weight $e^{-1\times 10^{14}}$ versus e^{-10} for a 10-bit fluctuation. If $K(O)$ ever exceeds the past-light-cone capacity, $P(O) \to 0$ as $t \to \infty$, expressing the normalcy prior underpinning PCI.

D Decision-Theory Details

Without PCI: $\ln\big[(1-P_{\rm BB})/P_{\rm BB}\big] \approx 9.21$. With PCI: $P_{\rm BB} \sim 1\times10^{-300} \Rightarrow \ln\big[(1-P_{\rm BB})/P_{\rm BB}\big] \approx 690$.

E Conditions for the PCI Limit

We now derive the slice-invariant "PCI Limit" (PCI Limit).

Instantaneous fraction. Throughout this appendix we define

$$P_{\rm absurd}(t) = \frac{\Gamma_{\rm abs}(t)}{\Gamma_{\rm tot}(t)},$$

i.e. the *rate* fraction of incoherent observers at proper time t. For late-time FRW backgrounds $\Gamma_{\rm tot}(t) \approx \text{const}$, we obtain $P_{\rm absurd}(t)\, t \to 0$ whenever $\int_T^\infty \Gamma_{\rm abs}(t)\, dt < \infty$.

Assume $\Gamma_{\rm abs} = Ae^{-\beta f(t)}$ with $f(t) \geq \ln t$ for $t > T$. Then

$$\int_T^\infty \Gamma_{\rm abs}\, dt \leq A \int_T^\infty t^{-\beta}\, dt < \infty \quad (\beta > 1 \text{ suffices; empirically } \beta \gg 10^{11}).$$

Because $\Gamma_{\rm tot}(t)$ is asymptotically constant (or, more generally, decays no faster than $1/t$), convergence of $\int \Gamma_{\rm abs} dt$ implies $\Gamma_{\rm abs}(t) = o(1/t)$ and hence $P_{\rm absurd}(t)\, t \to 0$ as $t \to \infty$, establishing the PCI Limit.[4] □

References

[1] L. Dyson, M. Kleban, and L. Susskind, Disturbing implications of a cosmological constant, *JHEP* 10, 011 (2002).

[2] J. Garriga and A. Vilenkin, Prediction and explanation in the multiverse, *Phys. Rev. D* 77, 043526 (2008).

[3] R. Bousso, Complementarity in the multiverse, *Phys. Rev. D* 88, 083517 (2013).

[4] K. K. Boddy, S. M. Carroll, and J. Pollack, De Sitter space without Boltzmann brains, *Found. Phys.* 46 (5), 702–717 (2016).

[5] S. M. Carroll, Why Boltzmann brains are bad, *SciPost Phys.* 3, 024 (2017).

[6] S. M. Carroll and R. Singh, Measure-independent probability drift in eternal inflation, *Phys. Rev. D* 109, 023506 (2024).

[4]This conclusion presumes a future in which $\Gamma_{\rm tot}(t)$ does not dilute more quickly than $1/t$, as in de Sitter-like or slowly evolving FRW cosmologies; an extreme Big-Crunch dilution would place PCI outside its intended domain.

[7] A. Ijjas and P. J. Steinhardt, A stable ekpyrotic bounce, *Phys. Lett. B* 795, 666–672 (2019). arXiv:1803.01961

[8] A. Albrecht and L. Sorbo, Measure constraints in late-time cosmology, *Phys. Rev. D* 111, 023507 (2025).

[9] S. Lloyd, Ultimate physical limits to computation, *Nature* 406, 1047–1054 (2000).

[10] R. Bousso, A covariant entropy conjecture, *JHEP* 07, 004 (1999).

[11] N. Bostrom, *Anthropic Bias: Observation Selection Effects in Science and Philosophy*, Routledge (2002).

[12] P. Bartha and C. Hitchcock, No one knows the date or the hour: an unorthodox application of the self-sampling assumption, *Mind* 108, 519–547 (1999).

[13] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd ed., Springer (2008).