ALIGNMENT THEOREM VERSION 1.1 · RESEARCH DRAFT

Ethical Work Under Hyperdeflation

A restricted theorem preserving Version 1's scarcity-driven mechanism

Dana Edwards / DarkLightX · August 2026

ABSTRACT

Version 1 proposed that increasing monetary scarcity could make eligible work uniquely profitable through EETF-gated rewards and exclusion from scarcity-linked upside. Version 1.1 generalizes that concrete mechanism. A finite purchasing-power multiplier amplifies eligible direct-reward and exclusive-upside coefficients. Those effects compete against private deviation gain, extra compliance cost, and optimizer error. The exact threshold is finite. Unbounded hyperdeflation crosses it eventually when the opposing opportunity remains bounded, or grows more slowly after normalization. Lean 4.33.0 checks the finite, asymptotic, and relative-growth statements. An exact bounded-integer model retains equality, common-scaling, faster-deviation, unauthenticated-EETF, and overflow counterexamples.

01

THE IDEA RETAINED FROM VERSION 1

Scarcity can make ethical work the most profitable strategy.

Version 1's human claim remains the target: a sufficiently deflationary economy can make ethical participation the most profitable strategy. As the settlement asset buys more real output, access to ethically conditioned rewards becomes more valuable. The value gap between access and exclusion grows with scarcity. Rational, profit-seeking humans and LLM agents then have an economic reason to choose eligible work.

Version 1.1 treats "ethical" as a classification of an observable action under authenticated EETF evidence. An agent proposes the action. The agent does not select its own EETF score. This semantic repair leaves the deflationary mechanism intact.

VERSION 1 COMPONENT	VERSION 1.1 TREATMENT
Scarcity multiplier	Finite purchasing-power index at every executable epoch.
EETF-conditioned rewards	Authenticated action classification controls eligibility.
Eligibility exclusion	Scarcity upside available only to the eligible branch; no punitive debit.
Rational convergence	Strict threshold theorem with explicit deviation and error bounds.
Infinite scarcity	Asymptotic statement: every finite threshold is eventually crossed.

02

ECONOMIC MODEL

A finite multiplier amplifies ethical exposure against a bounded opponent.

At epoch `t`, define:

SYMBOL	MEANING
`M(t)`	Finite real purchasing-power or hyperdeflation multiplier.
`r(t)`	Ethical reward exposure per unit of the multiplier.
`L(t)`	Scarcity-sensitive upside available only to the eligible action per multiplier unit.
`G(t)`	Maximum private gain available from a modeled deviation.
`C(t)`	Maximum extra cost of choosing the ethical alternative.
`epsilon(t)`	Permitted optimization error.

Using a common baseline utility `V(t)`, write the no-debit bounds:

$$U_{eligible}(t) \geq V(t) + M(t)\,[r(t) + L(t)] - C(t)$$
$$U_{excluded}(t) \leq V(t) + G(t)$$

The resulting utility gap satisfies:

$$U_{eligible}(t) - U_{excluded}(t) \geq M(t)\,[r(t) + L(t)] - [G(t) + C(t)]$$

Write:

$$K(t) = r(t) + L(t)$$
$$B(t) = G(t) + C(t) + \varepsilon(t)$$

The finite V1.1 condition is:

$$M(t)\,K(t) > B(t)$$

03

FINITE HYPERDEFLATIONARY ALIGNMENT THEOREM

A strict margin excludes every non-ethical epsilon-optimal choice.

THEOREM 1 · FINITE THRESHOLD

Fix one epoch with `K > 0` and finite nonnegative integers `M` and `B`. If `M K > B`, every `epsilon`-optimal choice in the stated two-action model is ethical. The smallest integer multiplier that guarantees the strict margin is `floor(B / K) + 1`.

Proof. The payoff bounds give `U_eligible - U_excluded > epsilon`. An excluded `epsilon`-optimal choice would require `U_excluded + epsilon >= U_eligible`, which contradicts that strict inequality. Euclidean division gives `B < K [floor(B/K) + 1]`. ∎

04

INFINITE SCARCITY AS AN ASYMPTOTIC THEOREM

Literal infinity never enters an executable transaction.

The economic path is a sequence of finite multipliers. “Hyperdeflation tends to infinity” means that for every finite bound, the sequence eventually remains at or above that bound.

THEOREM 2 · BOUNDED-DEVIATION ASYMPTOTIC FORM

Let `M(t)` eventually exceed every finite bound. Fix `K > 0` and a finite complete opposing bound `B`. There exists a finite epoch `T` such that every `epsilon`-optimal modeled choice is ethical for all `t >= T`.

Proof. The threshold `floor(B/K) + 1` is finite. Unbounded `M(t)` eventually reaches it. Theorem 1 applies thereafter. ∎

05

RELATIVE GROWTH AND THE AGI CASE

AGI can increase both abundance and deviation opportunity.

AGI can increase productive abundance while also increasing the scale of attacks, arbitrage, and external opportunities. A fixed bound on `G(t)` may therefore be too strong. Version 1.1 uses a relative-growth statement for this case.

> **THEOREM 3 · RELATIVE-GROWTH FORM**
>
> Suppose `K(t) > 0`. If, from some finite epoch onward, `B(t) / K(t) < M(t)`, then the strict hyperdeflation margin holds from that epoch onward.

This condition makes the growth-rate dependency visible. In an illustrative exponential scenario, scarcity may grow as `(1+d)^t` while the opposing opportunity grows as `(1+g)^t`. A persistent positive coefficient and `d > g` eventually favor the ethical action. Equal growth rates require a coefficient comparison. Faster deviation growth defeats this mechanism.

Research on AGI economics supports scenario analysis rather than a single inevitable price path. Automation outcomes depend on capital accumulation, compute, task structure, and irreproducible bottlenecks. Version 1.1 therefore states hyperdeflation as an assumption to test, rather than a guaranteed consequence of AGI.

06

BITCOIN-STYLE SCARCITY

Purchasing power is a market outcome, not a consensus rule.

Bitcoin’s block subsidy halves on a controlled schedule and its monetary supply approaches a finite cap. Outstanding supply does not follow Version 1’s geometric burn-to-zero equation. A Bitcoin-like Version 1.1 model uses `M(t)` as purchasing power per spendable coin. The multiplier can rise when real output or demand grows relative to the monetary stock.

The cap establishes controlled issuance. Market demand, productivity, velocity, liquidity, and competing assets determine purchasing power. The theorem is conditional on the resulting multiplier path. It does not derive that path from Bitcoin consensus rules alone.

07

COUNTEREXAMPLES RETAINED

Five families where the mechanism fails.

COUNTEREXAMPLE FAMILY A · COMMON SCALING

If `G(t) = M(t) K(t)` and the other terms are zero, the two sides remain equal. Scarcity grows without producing a strict preference.

COUNTEREXAMPLE FAMILY B · FASTER DEVIATIONS

If `K(t)=1` and `G(t)=M(t)^2`, non-ethical opportunity dominates the linear scarcity entitlement for every `M(t) > 1`.

COUNTEREXAMPLE FAMILY C · ZERO EXPOSURE

If `K(t)=0`, no amount of scarcity changes the ranking.

COUNTEREXAMPLE FAMILY D · UNAUTHENTICATED SCORING

A self-attested EETF lets the actor label a profitable deviation ethical. Economic arithmetic cannot repair a forged classification.

COUNTEREXAMPLE FAMILY E · UNFUNDED ENTITLEMENT

A nominal reward with no committed reserve may never be paid. It cannot supply the payoff assumed by the theorem.

EXECUTABLE AND FORMAL EVIDENCE

What has been checked and what remains pending.

LAYER	ESTABLISHED SCOPE	STATUS
Lean 4.33.0	Paper utility-bound bridge, exact least integer threshold, threshold monotonicity, finite choice theorem, bounded-deviation asymptotic theorem, and relative-growth theorem.	Checked; source-bound receipt; expected standard axioms only.
Python reference	Finite u64 envelope, strict boundary, exact-type rejection, overflow rejection, and direct raw-utility parity over 3,125 configurations.	Non-authoritative deterministic reference; tests pass for the stated finite domains.
Tau packet	Conjunction of four assumed propositions: EETF authenticated, action ethical, strict margin, and reward funded, over all 16 Boolean rows.	Exhaustive semantic packet present. A current source-pinned local candidate matched all 16 rows; reviewed-binary and public-node promotion remain pending.
Assurance receipt	Hashes the Lean sources and receipt, Python model and tests, Tau semantic packet, public pages, paper, and PDF.	Regenerated against the current no-debit source tree; explicitly records reference-only authority.

Tau does not receive “infinity.” The host computes each finite integer margin with explicit no-overflow checks. Tau combines the derived strict margin with Boolean propositions supplied

by its environment. The Python evaluator and Tau packet do not authenticate those propositions. They grant no publication or value-moving authority.

The trust boundary is external. The gate enforces that the declared relation was followed. It cannot verify that the facts fed into that relation were true. An oracle can report a false price, a meter can be miscalibrated, and a self-attested EETF can label a profitable deviation ethical. The chain faithfully executes on whatever validly-signed inputs it receives. A production system would require independent evidence providers (oracles, meters, signed attestations), dispute mechanisms, replay protection, state roots, reserve settlement, and governance over the active EETF policy. The theorem assumes honest authentication; it does not provide it.

09

RELATIONSHIP TO VERSION 2

V1.1 supplies the mechanism; V2 supplies the publication boundary.

Version 1.1 and Version 2 answer different questions. Version 1.1 asks when hyperdeflationary economics eventually makes ethical work most profitable. Version 2 asks what a policy-governed network may publish and reward at a finite epoch. A future network can combine both: V1.1 supplies the economic mechanism, while V2 supplies a deterministic publication boundary.

The operational difference matters. In V1.1, an excluded action can still occur; it lacks the eligible reward and protected scarcity upside, with no debit applied. Alignment comes from the utility gap. In V2, a noncompliant action is rejected at the gate as a no-op, is not published, and pays nothing. The two mechanisms can compose: V2 controls what the network admits, while V1.1 shapes the incentives around what agents choose to propose.

Read the Data Center Bargain (V2)

Version 1 claim boundary

LIMITATIONS AND NONCLAIMS

What the theorem does not prove.

→ The theorem is policy-relative and does not establish objective moral truth.

→ It does not prove AGI occurs or causes economy-wide hyperdeflation.

→ It does not prove Bitcoin purchasing power diverges.

→ It does not authenticate EETF evidence, market prices, sensors, or oracles.

→ Its Python decision and Tau output are reference results, not authority witnesses.

→ It does not cover unbounded or faster-growing deviations outside the relative-growth condition.

→ It does not establish coalition, sybil, governance-capture, liquidity, or market-manipulation safety.

→ It does not resolve infinite-horizon discounting, hoarding, utility saturation, or transversality by itself.

→ It does not establish Tau Net deployment or production readiness.

11

REFERENCES

Sources cited.

[1] P. Wuille, BIP 42: A finite monetary supply for Bitcoin.

[2] A. Korinek and D. Suh, Scenarios for the Transition to AGI, NBER Working Paper 32255, 2024.

[3] P. Restrepo, We Won't be Missed: Work and Growth in the AGI World, NBER Working Paper 34423, 2025.

[4] J. A. Gray and S. W. Salant, Transversality Conditions in Infinite Horizon Models, Federal Reserve IFDP 172, 1981.

[5] AlignmentTheorem repository, Version 1.1 Lean, Tau, and exact-model artifacts.

EVIDENCE BOUNDARY

Version 1.1 is a restricted, machine-checked theorem over declared assumptions. Its economic forecasts and authentication premises remain empirical and institutional obligations.