Bitcoin Exit Dominance

v2 to v3: mathematical repair with annotated redline

Reading key. Unchanged text is black. New text is red. ~~Deleted text is gray and struck through.~~ Blue boxes explain the changes and are not manuscript text. Equations are compared whole; the changed payoff table is shown as labeled rows. The clean PDF controls numbering.

Reader's note: What is substantive?

The original A4 is retained word for word. The new condition (NR) supplies an explicit coalitional payoff restriction: every returning group contains someone who would not gain at the resulting adoption level. It is additional formal content, not a consequence of the derivative signs. The proof now evaluates coordinated moves at their resulting profiles and proves the greatest high-Exit Nash equilibrium, when one exists, is strong Nash and hence BPW coalition-proof.

Reader's note: Why any changes outside the proof?

Only dependent mathematical summaries, the v3 date and revision note change. The abstract must state the condition used by the proof and the existence qualification. The introduction's final theorem summary and two conclusion sentences are aligned with those results. The title, original A1–A4, P1–P7, related-work section, acknowledgements, author declarations, and bibliographic entries are retained. The May revision note remains as historical text.

Sources. No new equilibrium concept or bibliography entry is needed. BPW (1987), printed/PDF pp. 3 and 6–7, supplies the coalition definitions and strong-Nash implication. Milgrom–Shannon (1994), printed pp. 162 and 175 (archived PDF pp. 7 and 20), supplies monotone choices and extremal-equilibrium context. The manuscript gives the finite-game proof directly. Downloaded originals, parsed passages, each change, its strongest defense and objection, and the derivation are linked in the accompanying Merkle DAG. Hashes establish provenance, not truth.

Prose comparison. v2: 4750 words; v3: 5199; net +449. Under the documented section-wise token alignment, 4390 v2 words are retained, 360 removed and 809 added. Equations and LaTeX commands are excluded; headings, revision notes and references are included.

Bitcoin Exit Dominance in Monetary Coordination Games

Sean Hash*

bitcoingametheory.com

sean@bitcoingametheory.com

Preprint: February 2026 (SSRN 6299081)

Revised: ~~May~~ November 2026 (v3)

Working Paper

JEL Codes: C72, D71, D83, E42, E51, G11, L14

Keywords: Bitcoin, game theory, coalition-proof Nash equilibrium, monetary coordination, neutral settlement, strategic complementarities, coordination failure, threshold strategy, digital scarcity, network effects

Abstract

Abstract. In a multipolar world with no trusted monetary coordinator, how do rational capital allocators preserve value when sovereign settlement systems are subject to debasement and political capture? The question is not hypothetical: the 2022 freezing of Russian foreign-exchange reserves demonstrated that sovereign settlement is capturable for any holder, central banks included. This paper models the choice as the Exit Game, a non-cooperative game in which allocators choose between capturable settlement systems, "Stay", and neutral settlement, "Exit". The model rests on four assumptions: persistent multipolarity, rational self-interest, computational hardness, and the persistence of network effects. ~~Three~~ Under the stated payoff and coalitional-persistence conditions, three results follow. First, the payoff advantage of Exit over Stay grows as adoption rises (Proposition 1). Second, no Stay coalition is internally stable once ~~any allocator~~a member's threshold is ~~reached~~ crossed at the proposed coalition outcome (Proposition 2). Third, Nash profiles in the high-Exit region ~~above a critical mass is coalition-proof against~~ resist returning Stay coalitions~~, and persistent debasement drives that~~ ; if such profiles exist, the greatest is coalition-proof. Accumulating recognition and debasement shift the critical mass monotonically downward ~~over time~~ (Theorem 1). Together these establish Bitcoin as the primary candidate for neutral settlement derived purely from ongoing incentives (Theorem 2). The formal Exit Game results show viable neutral settlement creates a new form of competition for the existing monetary order.

*1exitdP2akUuDreYSHSuzUx9zQv2n6JXu

Revision Note, ~~May 2026.~~November 2026 (v3). Readers identified that the coalition proof must evaluate payoffs after a proposed coordinated move. This revision includes each allocator's capital weight, states the coalitional-persistence condition (NR) explicitly, and proves that the greatest high-Exit Nash equilibrium, when one exists, is strong Nash and hence BPW coalition-proof. The threshold ratchet is retained; the inference that every adoption plateau must be temporary is removed. Assumptions A1–A4, Properties P1–P7, and the economic argument are retained. Theorem summaries are aligned with the repaired proof.

Previous Revision Note, May 2026. References updated and tightened. Substantive changes from the February 2026 preprint (SSRN 6299081):

- The preprint's absorption theorem is recast in ratchet form as Theorem 1. The preprint claimed full absorption at $p = 1$; the revision claims that $\{p > p_c\}$ is coalition-proof and that p_c shifts monotonically downward as the slowly-varying state (K_t, D_t) accumulates (K_t recognition; D_t debasement; §3.5). Permanent holdouts are admitted by construction.

- Adoption is capital-weighted: $p_t = \sum_{i \in \text{Exit}_t} a_i$. Preference falsification (Kuran[24]) provides the psychological reasoning for why actors say Stay but allocate Exit. Aggregate p_t counts the allocation, not the stated posture.

- Volatility apparatus is dropped. σ_B, σ_F, and the risk-aversion weight λ_i are removed; monotonicity conditions reduce from (M1)–(M5) to (M1)–(M4). The allocator-side friction is now money illusion (Fisher[21]): nominal stability is read as safety and lumpy real outperformance as risk.

- Definition 2 (Capturable System) replaces the preprint's Capture Surface. D_t in §3.5 now indexes three post-1945 financial-repression channels: debt overhang (IMF[22]), the inflation–capital-gains tax wedge (Feldstein[20]), and the rate-cap and reserve-requirement toolkit (Reinhart and Sbrancia[23]).

- The preprint's Limitations section is folded into the Conclusion; per-assumption failure conditions are retained at conclusion register, with A2's clause reformulated to a primitive of agent rationality.

- Coalition Instability (Proposition 2) restated as Nash-stability, with P3 (permissionless access) as the load-bearing institutional clause; Bernheim-Peleg-Whinston coalition-proofness is now reserved for Theorem 1 (Ratchet), where it does formal work.

The four assumptions and the central claim of Bitcoin Exit Dominance remain unchanged; the result is restructured into Theorem 2 for clarity.

1. Introduction

A capital allocator with purchasing power to protect faces a choice between two settlement systems. The first is the prevailing one: sovereign currency, custodied by regulated intermediaries, settled through banking and securities infrastructure. Call this *Stay*. The second is a settlement asset that no government issues, no intermediary custodies, and no political body can debase or freeze. Call this *Exit*. This paper frames the choice as the *Exit Game* and models its outcome with formal game theory.

The paper proceeds as follows. §2 summarizes related work. §3 builds the model. §4 proves Bitcoin Exit Dominance. §5 concludes with falsifications and future work.

While partial exit via gold has existed for thousands of years, true exit only became possible when Nakamoto[1] mined his first block on 3 January 2009. In the genesis block of Bitcoin he encoded that morning's front-page headline from *The Times* of London: "Chancellor on brink of second bailout for banks." Since 2009, hashrate on the Bitcoin network has only ever trended higher, despite sovereign mining bans and several hostile hard-fork events. Adoption by ever larger cohorts of capital (ETFs, pension funds, sovereigns) suggests rational allocators continue to defect to Exit.

The 2022 reserve freeze was the starting gun. The 2026 US-Iran conflict produced the first credible plan to use Bitcoin as neutral settlement. The United States ordered $344 million of USDT frozen (OFAC[2]), and Tether complied, because a stablecoin, unlike Bitcoin, has an issuer who can be compelled. With the stablecoin route closed, Iran's state proposal was to settle Strait of Hormuz cargo insurance in Bitcoin itself.

High government debt is not new. What is new is the scale of forward spending now committed with no tax base behind it, which leaves debasement as the path of least resistance. The United States is stepping back from its role as world military referee, and its former allies and competitors are rearming. The logic in every arms race is the same: spend more or get left behind (Richardson[3]; Powell[4]). AI is the same race, since better AI means better weapons: AI capex now runs at the same order of magnitude as military capex, and neither side can stop without ceding the lead (Armstrong, Bostrom, and Shulman[5]).

The deeper point in the genesis block is the nod to the *Cantillon effect*: in every fiat system, those nearest to issuance capture the wealth that debasement transfers, debasing the many to enrich the few (Cantillon[6]). This paper asks whether capital will flee to a settlement asset with no issuer, and so no Cantillon effect. The answer is an equilibrium result: once debasement pushes ~~adoption~~ a Nash allocation past a critical mass, ~~the move to Exit is self-reinforcing, and no coalition can bring capital back~~coalitional persistence (NR) prevents a profitable return, and the greatest Nash allocation in that region is coalition-proof.

2. Related Work

The closest prior policy treatment is Brunnermeier, James, and Landau[7], who describe the coordination problem verbally without formal derivation and prescribe central bank digital currency. A CBDC remains sovereign-issued, freezable, and debaseable, fails Definition 1, and is an instance of Stay that Theorem 2 covers.

The existing Bitcoin game-theoretic literature concentrates on the mining subgame, which sustains itself through miner skin in the game: hardware built for one purpose, rewards paid in the coin a successful attack would destroy. Two dissents press against this view. Budish[8] extends his high-frequency-trading rent-seeking critique to Bitcoin mining. Eyal and Sirer[9] identify a selfish-mining limit under which a concentrated pool earns above its honest share by withholding blocks. Neither attack has been profitably executed on chain: shorting Bitcoin at attack scale exceeds available derivatives liquidity, KYC at every fiat off-ramp makes the attacker's identity legible before they can position, and a pool revealed as malicious loses the hashrate it depends on.

The view of this paper, in line with Szabo[10], is the opposite of Budish: mining energy is the unforgeable costliness sustaining neutral settlement.[11] Three results triangulate the intertwined cryptographic and economic hardness Assumption 3 refers to. Biais, Bisière, Bouvard, and Casamatta[12] show that following the longest chain is a stable equilibrium in the mining subgame. Garay, Kiayias, and Leonardos[13] prove that under honest majority the protocol delivers common prefix, chain quality, and chain growth: the formal cryptographic guarantee that proof-of-work yields consensus. Chen[14] closes the price-security loop in general equilibrium. This literature models coordination within Bitcoin; the Exit Game models coordination within the wider sovereign monetary order.

3. Model

3.1 Maintained Assumptions

The analysis rests on four assumptions. All subsequent claims derive from these assumptions; rejecting any assumption invalidates the specific claims that depend on it.

Assumption 1 (Multipolarity)**.** Power is split across power centers where no center can exert unilateral control of the settlement layer.

Assumption 2 (Rational Self-Interest)**.** Actors optimize for self-interest. When defection from cooperative agreements is unpunished or unpunishable, actors defect.

Assumption 3 (Computational Hardness)**.** Certain mathematical problems cannot be solved in practical time by any known algorithm. Digital scarcity and cryptographic custody are implementable.

Assumption 4 (Network Effect Persistence)**.** Past critical mass, switching costs exceed marginal gains of alternatives. An established network attracts further adopters; its lead compounds (Katz and Shapiro[15]).

3.2 Definitions

Definition 1 (Neutral Settlement)**.** Bitcoin is the neutral settlement asset of the Exit-Stay game, satisfying Properties P1–P7. Each Pi closes one *attack channel* a coalition of Exit holders could use to profitably return to Stay. Cryptographic and consensus primitives grounding P1–P3 and P5–P6 trace to Nakamoto[1]. P1–P7 are assumed throughout, as is the miner-subgame result of §2.

Table 1. The seven attack channels Bitcoin closes. Each row names the property and the attack it blocks.

Closure Property	Attack Channel
P1: Informational security	Physical confiscation
P2: Protocol security	Protocol rule change
P3: Permissionless access	Transaction censorship
P4: Absolute scarcity	Supply inflation
P5: Cheap finality	Prohibitive cost/delay
P6: Neutrality	Governance takeover
P7: Adaptive resilience	Obsolescence

Gold as Neutral Settlement. Gold has been synonymous with neutral settlement across five millennia. Density, divisibility, recognizability, and aesthetic appeal sustained it as money in the market-selection account of monetary emergence (Menger[16]). Wampum, cowrie shells, and salt shared the surface properties but failed P4 once industrial production, oceanic shipping, and refined extraction broke their scarcity. Only gold held its scarcity across that span, the hardest proxy for P4 until Bitcoin. Arslanalp, Eichengreen, and Simpson-Bell[17] trace the post-2022 central-bank gold buying to financial-sanctions risk on the main reserve issuers, not generic flight-to-safety. Gold has four known structural failures: confiscation at handoff (P1), with Executive Order 6102 as precedent; assay risk (P5); settlement cost scaling with mass (P5); and supply inflation at roughly 1.5% annually (P4).

Definition 2 (Capturable System)**.** A monetary system in which at least one of P1–P7 fails for its underlying settlement asset. Fiat, sovereign bonds, and equities have no exit valve: equities admit no bearer instrument separable from the share, and fiat and bonds have no underlying outside the issuing system. Gold and Bitcoin offer redemption to a self-custodiable bearer asset. Gold terminates in physical gold, which retains the four failures above. Bitcoin ETFs can be redeemed in steps to self-custodied Bitcoin, which satisfies P1–P7.

Definition 3 (Exit)**.** The action of moving capital from a capturable system to a neutral settlement asset. The complement is *Stay*: holding capital in the prevailing capturable system.

3.3 Game

Reader's note: A mover changes total adoption

OWN-WEIGHT: With others at q, an allocator of weight a receives Exit utility at q+a and Stay utility at q. This is substitution into the existing aggregate definition. The finite-player version makes the BPW recursion applicable. BPW pp. 3, 6.

The Exit Game is a non-cooperative game in which each capital allocator chooses one of two actions: hold capital in the prevailing capturable system (Stay) or move it to a neutral settlement asset (Exit). Exit is not free. Early exit to Bitcoin requires a social cost to act against peer consensus. Depending on jurisdiction and employment, exit to Bitcoin can also cost real legal or reputational damage.

Players. ~~N is the~~ $N = \{1, \dots, n\}$ is the finite population of capital allocators, weighted by their share a_i of total capturable capital. $\kappa_i > 0$ is how much social cost allocator i pays for moving before peers. $\rho_i > 0$ is how exposed allocator i is to state penalty for the move itself. Fiduciaries answering to boards, committees, or constituencies sit at the high end of κ_i. Individuals and unilateral sovereigns sit lower. Both weights vary continuously across allocators.

Objective. Preserve real purchasing power. Preference between capturable and neutral assets follows from Exit Game payoffs alone.

Rules. Each allocator's per-period action is binary: Exit or Stay. The aggregate p_t ~~evolves continuously because allocators are heterogeneous in their thresholds and cross~~ changes as heterogeneous allocators cross their thresholds at different moments. Per-player actions are reversible: a player can re-Stay if ~~p_t~~ other allocators' Exit capital retreats below their personal threshold. Switching assets incurs friction; positions are sticky. As p rises, Exit pays more and Stay pays less (§3.4).

Dynamics. Recognition that capturable assets erode in real terms grows; debasement keeps accumulating. As both rise, ~~p_t trends upward~~ adoption thresholds shift downward (§3.5).

Payoffs. The game is $G = (N, S, u)$ in standard non-cooperative form (Fudenberg and Tirole[18]), with per-period action set $S_i = \{\text{Exit}, \text{Stay}\}$, capital weights $\sum_i a_i = 1$, and aggregate adoption $p_t = \sum_{i \in \text{Exit}_t} a_i$. ~~Each player's strategy is a threshold $p_i^* \in [0,1] \cup \{+\infty\}$: Exit when $p_t > p_i^*$~~Write $x_i = 1$ for Exit and $x_i = 0$ for Stay, $p(x) = \sum_i a_i x_i$, and $q_i(x) = \sum_{j \neq i} a_j x_j$. Player i receives $u_i^E(p(x))$ at Exit and $u_i^S(p(x))$ at Stay. With others fixed at adoption q, its unilateral Exit advantage is $F_i(q) = u_i^E(q + a_i) - u_i^S(q)$, for $0 \leq q \leq 1 - a_i$. It chooses Exit when $F_i(q) > 0$, Stay when ~~$p_t \leq p_i^*$. The per-player~~ $F_i(q) < 0$; either is a best response at equality. The action is binary and reversible~~at the player level; the~~; aggregate p_t ~~moves continuously~~ changes across the heterogeneous threshold distribution and may itself drawdown. The monotone object is the critical mass p_c, which shifts downward as ~~(K_t, D_t)~~ (K_t, D_t) (recognition; debasement) accumulate (§4.3).

$$u_i^E(p) := R_B(p) - \kappa_i K_A(p) - \rho_i R_A(p),$$

$$u_i^S(p) := R_F - K_N(p).$$

- $R_B(p)$: real return on the neutral asset.
- R_F: return on capturable assets.
- $K_A(p)$: conformity cost of acting before peers (Katz and Shapiro[15]).
- $R_A(p)$: realization and regulatory friction; gains tax, fees, *etc.*
- $K_N(p)$: non-adoption penalty.

Weights are strictly positive and differ across allocators. R_B, K_A, R_A, K_N are ~~continuous~~ continuously differentiable and bounded.

~~**Table 2.** Normal-form payoff matrix for player i, with u_i^E, u_i^S as defined above.~~

~~	Others: Stay	Others: Exit~~
~~Player i: Stay~~	~~$u_i^S(0)$~~	~~$u_i^S(p)$~~
~~Player i: Exit~~	~~$u_i^E(0)$~~	~~$u_i^E(p)$~~

Table 2. Payoff comparison for player i, with u_i^E, u_i^S as defined above.

	Others: Stay	Others: adoption q
Player i: Stay	$u_i^S(0)$	$u_i^S(q)$
Player i: Exit	$u_i^E(a_i)$	$u_i^E(q + a_i)$

3.4 Monotonicity Conditions

Milgrom and Shannon[19] showed that when the payoff gap between two options widens as some parameter rises, the player's best response moves toward the favored option and does not return. This paper applies that result with adoption p as the parameter and Exit as the favored option: as more actors exit, Exit's gap over Stay widens. Four conditions on the payoff components secure that widening.

(M1) $R'_B(p) > 0$: Bitcoin's return rises with adoption (network effects, Assumption 4)

(M2) $K'_A(p) < 0$: The cost of switching to Bitcoin falls as more allocators switch

(M3) $R'_A(p) < 0$: Friction eases as adoption normalizes

(M4) $K'_N(p) > 0$: The cost of staying in capturable assets rises as competitors exit

These ~~conditions hold on average over long horizons rather than at every instant~~are pointwise conditions on the long-horizon payoff functions used in the formal results.

3.5 Threshold Drift

Reader's note: The cutoff and its direction

DRIFT: The minimum of ordered individual thresholds is ordered too. That does not imply that actual adoption crosses it. The cutoff is measured from others' adoption and the coalitional restriction is stated separately. Milgrom–Shannon pp. 162, 175.

Sudden seizure (capital controls, asset freezes, hyperinflation) is the visible failure mode of the capturable system. The actual failure mode is slower: continuous extraction of real wealth. Debasement raises the price level. Equities, real estate, and similar assets inflate nominally as they absorb a share of the newly created money. Holders are then taxed on the nominal gain. After tax, real returns are negative (Feldstein[20]). Public debt keeps growing; private claims bear the hidden inflation tax to inflate away the real debt burden. Multipolarity (Assumption 1) leaves debasement unchecked, so $R_F < 0$ in real terms persists. Allocators under money illusion (Fisher[21]) read nominal stability as safety and lumpy real outperformance as risk, so the erosion goes uncorrected.

The macro state evolves on two slowly-varying indices:

- $K_t \in [0, 1]$ (*recognition*): the share of capital allocators who recognize capturable assets as financially repressed. Recognition accumulates gradually rather than arriving with a discrete capture event.
- $D_t \in [0, \infty)$ (*debasement*): a latent debasement-intensity index summarizing three post-1945 financial-repression channels: debt-to-GDP overhang (IMF[22]), the inflation–capital-gains tax wedge (Feldstein[20]), and the rate-cap and reserve-requirement toolkit (Reinhart and Sbrancia[23]).

Both K_t and D_t are monotonic over the structural horizon. Short-horizon reversals are admitted as fluctuations around the trend: generational forgetting in K_t, hawkish tightening episodes in D_t.

Let $\Delta_i(p; K, D) := u_i^E(p) - u_i^S(p)$ denote player i's Exit minus Stay payoff gap at adoption p and macro state (K, D). Exact best responses use $F_i(q; K, D) = u_i^E(q + a_i; K, D) - u_i^S(q; K, D)$. Each allocator's threshold depends on ~~(K, D):~~

$$p_i^*(K, D) := \inf\{p \in [0, 1] : \Delta_i(p; K, D) > 0\}.$$

(K, D):

$$p_i^*(K, D) := \inf\{q \in [0, 1 - a_i] : F_i(q; K, D) > 0\}, \qquad \inf \emptyset = +\infty.$$

~~Assume the sign conditions~~

$$\frac{\partial p_i^*}{\partial K} \leq 0, \qquad \frac{\partial p_i^*}{\partial D} \leq 0.$$

Maintain the threshold-order condition

$$p_i^*(K', D') \leq p_i^*(K, D) \quad \text{whenever } K' \geq K,\ D' \geq D.$$

These hold strictly for at least some capital allocators. Higher D_t lowers R_F in real terms. Higher K_t erodes $\kappa_i K_A$ and $\rho_i R_A$. Both effects raise ~~$\Delta_i(0; K, D)$~~ $F_i(0; K, D)$ and push p_i^* down.

The ~~critical mass invoked~~ critical-mass cutoff used in Assumption 4~~is the smallest adoption level at which at least one allocator's threshold has been crossed:~~

$$p_c(K, D) := \inf\Big\{p \in [0, 1] : \sum_{i\,:\,\Delta_i(p; K, D) > 0} a_i \ > \ 0\Big\}.$$

's formal restriction (NR) is the minimum individual threshold, measured in others' adoption:

$$p_c(K, D) := \min_{i \in N} p_i^*(K, D).$$

Above p_c, the ~~M1–M4 feedbacks (Katz and Shapiro[15]) prevent Stay coalitions from reforming~~ (formal coalitional-persistence condition (NR) in §4~~.2~~).3 prevents profitable returns from Nash profiles. The endpoint is dominance in the network-effects sense, not full adoption ~~$p = 1$. For each (K, D), $\{p > p_c\}$~~ $p = 1$. If a Nash profile lies above p_c, the greatest Nash profile in that region is coalition-proof~~against return-coalitions. As (K_t, D_t)~~. As (K_t, D_t) accumulate, p_c shifts downward by monotone comparative statics (Milgrom and Shannon[19]):

~~$p > p_c$ holds even when p_t is flat, and the coalition-proof region expands~~ at fixed adoption, $p > p_c$, once satisfied, remains satisfied as p_c weakly decreases (§4.3).

The model's claims are structural and directional: existence of thresholds $\{p_i^*\}$, sign of comparative statics, and direction of shift in p_c under accumulating (K_t, D_t). It does not predict specific values, tipping times, or saturation levels.

An ~~adoption level p is an equilibrium of the Exit Game~~ action profile is Nash when no allocator ~~, alone or in any group that holds together,~~ gains by switching alone. Coalition-proofness additionally uses the recursive self-enforcement test of Bernheim, Peleg and Whinston[25], with outsiders' actions fixed. Coalitional gains require strict improvement for every member; deviations are pure and exclude transfers.

4. Results

4.1 Monotone Exit Advantage

Reader's note: The original derivative result survives

BOUNDARY / POINTWISE: The derivative signs apply to the modeled long-horizon utilities. The exact own-weight gain also increases. At a zero threshold the gain may already be positive, so the old claim that every finite threshold has zero gain is corrected.

First, each actor's payoff gap from Stay to Exit widens as more actors exit. Every actor therefore has a threshold above which Exit becomes the only rational move. Hold the slowly varying state (K, D) fixed throughout this subsection. By convention, ~~$p_i^* = +\infty$ when Δ_i~~ $p_i^* = +\infty$ when F_i stays non-positive on ~~$[0,1]$~~$[0, 1 - a_i]$.

Proposition 1 (Monotone Exit Advantage)**.** *Under Assumptions 1–4 and maintained conditions (M1)–(M4):*

(i) The payoff differential $\Delta_i(p)$ is strictly increasing in p for all player types.

(ii) For each player i there exists a threshold ~~$p_i^* \in [0,1] \cup \{+\infty\}$~~ $p_i^* \in [0, 1 - a_i] \cup \{+\infty\}$ such that Exit is the unique best response for ~~$p > p_i^*$~~ $q_i > p_i^*$. If $F_i(0) > 0$, Exit is uniquely optimal also at $q_i = 0$.

Proof. Define the payoff differential and take its derivative in p:

$$\begin{aligned} \Delta_i(p) &= u_i(\text{Exit}) - u_i(\text{Stay}) \\ &= [R_B(p) - R_F] - \kappa_i K_A(p) - \rho_i R_A(p) + K_N(p), \\ \frac{d\Delta_i}{dp} &= R_B'(p) - \kappa_i K_A'(p) - \rho_i R_A'(p) + K_N'(p). \end{aligned}$$

Under (M1)–(M4), each term is signed: $R_B'(p) > 0$ and $K_N'(p) > 0$ strictly, while $-\kappa_i K_A'(p) \geq 0$ and $-\rho_i R_A'(p) \geq 0$ for $\kappa_i, \rho_i > 0$. Therefore $d\Delta_i/dp > 0$: the advantage of Exit rises strictly with p. For exact unilateral incentives,

$$F_i'(q) = (u_i^E)'(q + a_i) - (u_i^S)'(q) > 0.$$

Thus the positive set of F_i is an upper interval, possibly empty or the whole domain. Each player's threshold ~~$p_i^* := \inf\{p \in [0,1] : \Delta_i(p) > 0\}$ specializes~~ is the §3.5 ~~definition with (K, D) fixed. For actors with $p_i^* \in [0,1]$, the intermediate value theorem applied to the continuous Δ_i delivers $\Delta_i(p_i^*) = 0$, and for $p > p_i^*$ Exit strictly dominates~~infimum with (K, D) fixed. Continuity gives $F_i(p_i^*) = 0$ at a finite interior threshold. If $F_i(0) > 0$, the threshold is zero and Exit is uniquely optimal throughout; at an actual zero of F_i, either action is a best response.

4.2 Coalition Instability

> **Reader's note: Test the proposed outcome**
>
> POST-COALITION: A Stay member destabilizes a proposed coalition only when exiting is profitable at that proposed outcome, including its own weight. Profitability at the starting allocation is a different comparison. BPW p. 6.

§4.1 gave each actor a threshold. The group implication is sharper: no Stay coalition is self-enforcing once any one member has crossed. The first defector profits unilaterally; the actors left behind face a higher exit advantage than before.

Proposition 2 (Coalition Instability)**.** *Under Assumption 2 and Property P3, no Stay coalition C containing at least one actor with a crossed threshold (~~$p > p_j^*$ for some $j \in C$~~$F_j(p) > 0$ for some $j \in C$ at the proposed coalition outcome) is Nash-stable.*

Proof. Consider a coalition $C \subseteq N$ maintaining Stay. For C to be Nash-stable, no member $j \in C$ can profit from unilateral defection. ~~By Proposition 1, u_j(Exit) $> u_j$(Stay) when p exceeds p_j^*~~At that outcome, $F_j(p) > 0$ means $u_j^E(p + a_j) > u_j^S(p)$. Property P3 (permissionless access) closes the institutional defection channel: ~~no counterparty can refuse the trade~~ Exit is feasible in the modeled action set S_i, and no clearing intermediary can foreclose settlement, so the only remaining friction is the regulatory cost already baked into each actor's threshold via R_A. Under Assumption 2, the crossed-threshold member defects: for at least one $j \in C$, u_j(defect from C) $> u_j$(remain in C). The coalition therefore fails Nash.

Proposition 2 covers coalitions with at least one crossed-threshold member. Permanent holdouts ($p_j^* = +\infty$), admitted by Proposition 1, lie outside its scope: no private reason to defect. The aggregate share of capital held for ideological rather than rational reasons (Assumption 2) is not fixed: preferences adjust as new information arrives. The real-world case is regulatory coercion holding the coalition together: capital controls, fiduciary mandates, jurisdictional penalties. P1 limits what coercion can do. An asset controlled by secret information, not physical possession, can be seized only by forcing each holder to disclose it one at a time. Cost grows with the number of holders, not the size of any one holding. Preference falsification (Kuran[24]) then separates what holders say from what they do. Aggregate p_t tracks what they do. Once enough private exits register, public Stay breaks.

4.3 Ratchet

Reader's note: Coordinated return and the replacement proof

A4-NR / PROFILE-SCOPE: If every returning member gains, any subgroup re-entering receives no more than its original Exit utility and therefore loses. Such a profitable return is self-enforcing. NR excludes it directly. The greatest Nash equilibrium additionally excludes profitable collective entry; departures mixed with entries are ruled out by NR. This yields strong Nash and BPW. The removed finite-plateau inference does not follow from downward threshold drift. BPW pp. 6–7; Milgrom–Shannon p. 175.

§4.1 gave each actor a threshold; §4.2 showed coalitions cannot hold once any one member crosses. §4.3 lets the slowly varying state (K_t, D_t) move. Per-period actions stay binary and reversible at the player level. What changes over time is the macro state, and with it the threshold profile $\{p_i^*(K_t, D_t)\}$ and the critical mass $p_c(K_t, D_t)$. Recognition K_t ratchets rather than oscillates: once private allocations reveal previously concealed Exit preferences, the social cost structure sustaining prior public Stay no longer holds (Kuran[24]). Debasement D_t does not reverse on the structural horizon (Reinhart and Sbrancia[23]). Both push p_c downward. The result is not that any specific player is locked in; it is that the floor a returning Stay coalition would have to clear keeps dropping. Two claims follow under the explicit persistence condition below. First, Nash profiles above p_c ~~no returning Stay coalition can form: a candidate with a crossed-threshold member is broken by Proposition 2; a candidate without one contains only allocators not moving to Exit~~resist returning Stay coalitions; if such a profile exists, the greatest Nash profile is coalition-proof. Second, p_c is non-increasing in t.

Formal coalitional persistence. Keep Assumption 4 and make its coalitional scope explicit by maintaining the following additional payoff restriction. Write $E(x) = \{i : x_i = 1\}$

and $a(C) = \sum_{i \in C} a_i$. At every Nash profile x under consideration with $p = p(x) > p_c$, require

$$\text{for every } \emptyset \neq C \subseteq E(x), \quad \text{some } i \in C \text{ has } u_i^E(p) \geq u_i^S(p - a(C)). \tag{NR}$$

This compares current Exit utility with the utility after the whole returning group moves. It is maintained as the formal persistence restriction, not inferred from the derivative signs. The proof below shows why it is the exact return-coalition condition for this game.

Theorem 1 (Ratchet)**.** *Under Assumptions 1–4, maintained conditions (M1)–(M4), (NR), and the §3.5 ~~sign conditions~~threshold-order condition: (a) ~~$\{p > p_c\}$ is coalition-proof[25,26] against~~ every Nash profile with $p > p_c$ resists profitable returning Stay coalitions. If any Nash profile lies above p_c, the greatest Nash profile there is strong Nash and hence coalition-proof[25,26]. (b) Under monotonic accumulation of ~~(K_t, D_t), the critical mass $p_c(K_t, D_t)$ shifts monotonically downward~~ (K_t, D_t), $p_c(K_t, D_t)$ is non-increasing in t.*

Proof. (a) Coalition-proofness~~of $\{p > p_c\}$~~. ~~If $p > p_c(K, D)$, the crossed-threshold set~~

$$H(p) := \{i : \Delta_i(p; K, D) > 0\}$$

~~has positive capital weight. Any perturbation attempting to coordinate a return~~ First consider any strictly profitable return by $C \subseteq E(x)$ from adoption p to ~~high-Stay above p_c is a candidate Stay coalition C and falls into one of two cases. If $C \cap H(p)$ is nonempty, Proposition 2 applies: the crossed-threshold member defects under P3, so~~ $r = p - a(C)$. Every member then has $u_i^S(r) > u_i^E(p)$. In the game induced on C ~~fails Nash and therefore fails BPW self-enforcement. If $C \cap H(p)$ is empty,~~ with outsiders fixed, any nonempty subgroup $B \subseteq C$ re-entering Exit receives

$$u_i^E(r + a(B)) \leq u_i^E(p) < u_i^S(r) \qquad (i \in B).$$

Every actual re-entrant loses. The return outcome is therefore strong Nash in the induced game, hence BPW coalition-proof. Consequently (NR) is necessary and sufficient for resistance to profitable BPW return coalitions.

For the full equilibrium result, define the greatest-best-response map

$$T_i(x) = \mathbf{1}\{F_i(q_i(x)) \geq 0\}.$$

It is increasing by Proposition 1. Iteration from all-Exit decreases and, since N is finite, terminates at a Nash profile x^+. For any Nash profile w, $w \leq T(w)$; induction gives $w \leq x^+$, so x^+ is greatest, including equilibria choosing Stay at ties. This is the finite extremal-

equilibrium construction of Milgrom and Shannon.[19] If any Nash profile exceeds p_c, so does x^+. Put $p^+ = p(x^+)$.

A profitable entry-only coalition A would produce adoption $s = p^+ + a(A)$, with each entrant satisfying

$$u_i^E(s) > u_i^S(p^+) \geq u_i^S(s - a_i).$$

Old Exit holders remain best responding by monotonicity. Thus the new profile y satisfies $y \leq T(y)$; upward iteration yields a Nash profile above x^+, a contradiction.

For a deviation containing returners C ~~consists entirely of uncrossed-threshold capital. These actors already prefer Stay at the current p and do not move toward Exit ; no candidate Stay coalition expands above p_c. The system is coalition-proof above p_c.As (K_t, D_t) accumulate, the finite-threshold Stay pool shrinks and the~~ and entrants A, resulting adoption is $s = p^+ - a(C) + a(A)$. If every member strictly gains, each returner satisfies

$$u_i^S(p^+ - a(C)) \geq u_i^S(s) > u_i^E(p^+),$$

contradicting (NR). This covers coalitions containing unchanged members as well. No strictly profitable pure coalition deviation exists, so x^+ is strong Nash. Each restriction of a strong Nash profile to an induced game is strong Nash; induction on coalition size gives BPW self-enforcement, and strong Nash excludes a strict Pareto improvement by the whole coalition. Hence x^+ is BPW coalition-proof~~region expands. ~~.[25]

(b) Drift of p_c. For ordered states $z' \succeq z$, the §3.5 condition gives $p_i^*(z') \leq p_i^*(z)$ for every player. Taking the minimum gives

$$p_c(z') = \min_i p_i^*(z') \leq \min_i p_i^*(z) = p_c(z).$$

This order argument includes infinite thresholds and needs no derivative of the minimum. Under monotonic accumulation of (K_t, D_t), p_c is non-increasing.

Drawdowns and cascade pauses along p_t are admitted; price drawdowns are not equivalent to drawdowns in p_t, since allocators below target weight can rebalance during decline. ~~A permanent stall would require two conditions to hold jointly: p_t plateauing below p_c indefinitely, and p_c failing to fall. The second condition is ruled out by part (b), so any p_t plateau is finite-duration.~~

~~*(b) Drift of p_c.* Because p_c is defined from the individual thresholds, the .5 sign conditions transfer directly to p_c:~~

$$\frac{\partial p_i^*}{\partial K} \leq 0, \quad \frac{\partial p_i^*}{\partial D} \leq 0, \quad \frac{\partial p_c}{\partial K} \leq 0, \quad \frac{\partial p_c}{\partial D} \leq 0.$$

~~Under monotonic accumulation of (K_t, D_t), $p_c(K_t, D_t)$ is non-increasing in t.~~

4.4 Bitcoin Exit Dominance

Reader's note: Keep the summary no stronger than its proof

PROFILE-SCOPE: The full BPW result applies to the greatest high-Exit Nash profile if one exists. The conjunction retains that existence qualification and calls the object a game, not an unspecified adoption dynamic.

The three preceding results all hold simultaneously in the Exit Game under the same hypotheses.

Theorem 2 (Bitcoin Exit Dominance)**.** *Under the four maintained assumptions A1–A4 of §3.1 (multipolarity, rational self-interest, computational hardness, network-effect persistence), the seven settlement properties P1–P7 of §3.2, the Exit Game G of §3.3, the monotonicity conditions (M1)–(M4) of §3.4 governing payoff derivatives in p, the formal persistence condition (NR), and the §3.5 ~~sign conditions~~threshold-order condition, the capital-weighted ~~adoption process~~ Exit Game satisfies all three of:*

(i) the payoff advantage of Exit over Stay is strictly increasing in p for every player type (Proposition 1);

(ii) no Stay coalition with ~~at least one crossed-threshold member~~ a member whose exact Exit gain is positive at the proposed outcome is self-enforcing (Proposition 2);

(iii) ~~$\{p > p_c\}$~~ Nash profiles above p_c resist returning Stay coalitions; if such profiles exist, the greatest is coalition-proof~~, and~~. The critical mass p_c shifts monotonically downward under accumulating ~~(K_t, D_t)~~ (K_t, D_t) (Theorem 1).

Proof. Proposition 1 gives (i). Proposition 2, with (i) supplying the per-defector incentive, gives (ii). Theorem 1 gives (iii). The conjunction of (i)–(iii) is the Bitcoin Exit Dominance claim.

5. Conclusion

The Exit Game treats capital allocators, not sovereign issuers, as the agents who decide what counts as settlement. Bitcoin is an asset that no government issues, and to date has satisfied P1–P7. The question is no longer whether allocators can leave the capturable system, but whether continued debasement pushes adoption past the critical mass p_c.

All capital allocators choose reserve assets on perceived risk and ability to store value. Since the Global Financial Crisis central banks have been accumulating gold. Smaller allocators, with higher risk tolerance and lower position size, can shift first into an asset that

holds gold's monetary properties and satisfies P1, P4, and P5, which gold fails. §3.5 predicts that continued debasement keeps ~~moving more allocatorsout of fiat~~lowering allocators' Exit thresholds.

A1–A4 are each independently load-bearing. A1 (persistent multipolarity) fails only if rival regimes coordinate and preference falsification no longer holds; rival seignorage is what holds the multipolar floor. A2 (rational self-interest) fails only if actors systematically forgo positive expected returns from unpunished defection. A3 (computational hardness) fails if either the underlying cryptography breaks and the protocol cannot replace it, or the economic incentive sustaining the mining subgame falls below the level §2 requires. ~~A4 (persistence of network effects) fails only if the incumbency account of network goods is disproved~~The formal coalitional-persistence restriction (NR) fails when a returning coalition strictly improves every member's payoff at its resulting adoption level.

The Bitcoin Exit Dominance claim is new and depends on two conditions outside this paper's scope. First, no other consensus ledger holds all of P1–P7: alternatives must either fail one of the properties or fail to draw adoption. Second, autonomous AI agents that need a settlement asset choose Bitcoin rather than invent a new settlement asset. A companion paper in preparation addresses the first across consensus ledgers; a second extends neutral settlement to autonomous agents.

Acknowledgements. The author thanks the game-theoretic tradition this paper rests on, from Von Neumann and Morgenstern through Nash and Selten to the equilibrium refinement program of Bernheim, Peleg and Whinston, which the Exit Game uses to reduce candidate equilibria to a provable coalition-proof state. The author also acknowledges Thorstein Veblen for the institutional-psychological tradition this paper draws on. Thanks to the readers who flagged the Theorem 1 ratchet-form gap and the redundancy of the volatility apparatus in the preprint. Finally, the author thanks Lyn Alden for popularizing the debasement concept to a general audience; her work in accelerating the broader recognition of fiat debasement aids the Exit Game. The author also acknowledges Arthur Hayes, Jeff Booth, Fred Krueger, Lawrence Lepard, and others whose public writing has accelerated the bitcoin-specific debasement discourse this paper formalizes.

Author Contributions. S.H. is the sole author and conducted all aspects of the work, including conceptualization, formal analysis, and writing.

Conflict of Interest. The author holds bitcoin and operates bitcoingametheory.com.

Table A1. Notation used throughout.

Symbol	Definition
G	Settlement game (N, S, u)
N	Set of capital allocators
S_i	Strategy set: $\{\text{Exit}, \text{Stay}\}$
u_i	Utility function
a_i	Actor i's share of total capturable capital, $\sum_i a_i = 1$
p_t	Aggregate capital share at Exit at time t: $\sum_{i \in \text{Exit}_t} a_i$
$p_i^*(K, D)$	Threshold for player i at slowly varying state (K, D)
$p_c(K, D)$	Critical-mass threshold at (K, D)
K_t	Financial-repression recognition share, $K_t \in [0, 1]$
D_t	Debasement-intensity index, $D_t \in [0, \infty)$
$\Delta_i(p; K, D)$	Payoff differential: $u_i(\text{Exit}) - u_i(\text{Stay})$ at (K, D)
$R_B(p)$	Expected real return on neutral settlement asset
R_F	Expected real return on capturable assets
$K_A(p)$	Conformity cost of acting before peers
$R_A(p)$	Realization and regulatory friction (decreases as adoption normalizes)
$K_N(p)$	Non-adoption penalty
κ_i, ρ_i	Conformity weight, friction weight (all > 0 strictly)
(M1)–(M4)	Monotonicity conditions (§3.4)

6. Notes and References

1. Nakamoto, S. "Bitcoin: a peer-to-peer electronic cash system." *White paper* (2008) https://bitcoin.org/bitcoin.pdf

2. U.S. Department of the Treasury, Office of Foreign Assets Control. "Iran-related Designations; Counter Terrorism and Iran-related Designation Update; Issuance of Iran-related General License." *OFAC Recent Actions* (24 April 2026) https://ofac.treasury.gov/recent-actions/20260424

3. Richardson, L. F. *Arms and Insecurity: A Mathematical Study of the Causes and Origins of War*. Pittsburgh: Boxwood Press / Chicago: Quadrangle Books 12–35 (1960).

4. Powell, R. "Guns, butter, and anarchy." *American Political Science Review* **87.1** 115–132 (1993) https://doi.org/10.2307/2938960

5. Armstrong, S., Bostrom, N., Shulman, C. "Racing to the precipice: a model of artificial intelligence development." *AI & Society* **31.2** 201–206 (2016) https://doi.org/10.1007/s00146-015-0590-y

6. Cantillon, R. *Essai sur la nature du commerce en général* (c. 1755). English translation: *Essay on the Nature of Trade in General*, trans. Higgs, H. London: Macmillan 161–173 (1931).

7. Brunnermeier, M. K., James, H., Landau, J.-P. "The digitalization of money." *NBER Working Paper* **26300** (2019) https://doi.org/10.3386/w26300

8. Budish, E. "The economic limits of Bitcoin and the blockchain." *NBER Working Paper* **24717** (2018) https://doi.org/10.3386/w24717

9. Eyal, I., Sirer, E. G. "Majority is not enough: Bitcoin mining is vulnerable." *Communications of the ACM* **61.7** 95–102 (2018) https://doi.org/10.1145/3212998

10. Szabo, N. "Shelling Out: The Origins of Money." *Unenumerated* (2002) https://nakamotoinstitute.org/library/shelling-out/

11. *Note on Budish (§2).* The mining subgame imposes real costs that are a legitimate object of inefficiency analysis. The viewpoint of this paper is that those costs are not inefficiency in the rent-seeking sense but unforgeable costliness (Szabo[10]) sustaining neutral settlement: the specific property Bitcoin holders are paying to obtain. A full reconciliation with Budish's bound is beyond the scope of one paper and is left to subsequent work.

12. Biais, B., Bisière, C., Bouvard, M., Casamatta, C. "The blockchain folk theorem." *Review of Financial Studies* **32.5** 1662–1715 (2019) https://doi.org/10.1093/rfs/hhy095

13. Garay, J., Kiayias, A., Leonardos, N. "The Bitcoin backbone protocol: analysis and applications." In *Advances in Cryptology – EUROCRYPT 2015*, Lecture Notes in Computer Science **9057**, Springer 281–310 (2015) https://doi.org/10.1007/978-3-662-46803-6_10

14. Chen, L. "A game-theoretic foundation for Bitcoin's price: a security-utility equilibrium." arXiv:2508.06071 *Working Paper* (2025) https://arxiv.org/abs/2508.06071

15. Katz, M. L., Shapiro, C. "Network externalities, competition, and compatibility." *American Economic Review* **75.3** 424–440 (1985) https://www.jstor.org/stable/1814809

16. Menger, C. "On the Origin of Money." *Economic Journal* **2.6** 239–255 (1892) https://www.jstor.org/stable/2956146

17. Arslanalp, S., Eichengreen, B., Simpson-Bell, C. "Gold as international reserves: a barbarous relic no more?" *Journal of International Economics* **145** 103822 (2023) https://doi.org/10.1016/j.jinteco.2023.103822

18. Fudenberg, D., Tirole, J. *Game Theory*. Cambridge: MIT Press 3–7 (1991).

19. Milgrom, P., Shannon, C. "Monotone comparative statics." *Econometrica* **62.1** 157–180 (1994) https://doi.org/10.2307/2951479

20. Feldstein, M. "Inflation and the stock market." *American Economic Review* **70.5** 839–847 (1980) https://www.jstor.org/stable/1805772

21. Fisher, I. *The Money Illusion*. New York: Adelphi Company 3–22 (1928).

22. International Monetary Fund. "World Economic Outlook Database." *IMF* (October 2024) https://www.imf.org/en/Publications/WEO

23. Reinhart, C. M., Sbrancia, M. B. "The liquidation of government debt." *Economic Policy* **30.82** 291–333 (2015) https://doi.org/10.1093/epolic/eiv003

24. Kuran, T. *Private Truths, Public Lies: The Social Consequences of Preference Falsification*. Cambridge: Harvard University Press 3–21 (1995).

25. Bernheim, B. D., Peleg, B., Whinston, M. D. "Coalition-proof Nash equilibria I. Concepts." *Journal of Economic Theory* **42.1** 1–12 (1987) https://doi.org/10.1016/0022-0531(87)90099-8

26. *Note on coalition-proofness vocabulary (§4.3).* The ~~claim is~~ full coalition-proofness claim concerns the greatest Nash profile in the Bernheim-Peleg-Whinston sense~~,~~; resistance to returning coalitions is stated separately. It is not stochastic stability in the Young or Kandori-Mailath-Rob sense; the paper does not specify a perturbed best-response dynamic.