Settlement at Zero Trust: Bitcoin and Autonomous Economic Agents

Sean Hash

bitcoingametheory.com

sean@bitcoingametheory.com

February 2026

Working Paper

JEL Codes: C72, D82, D83, E42, G11, L14, O33

Keywords: Bitcoin, AI agents, autonomous settlement, trust gradient, zero recourse, game theory, Nash equilibrium, cryptographic settlement, network effects

Abstract

As autonomous AI agents become economic actors, executing transactions and holding assets, they face a settlement problem that human actors prefer to solve but AI agents must solve. When a counterparty defects against an AI agent, the agent cannot sue, cannot appeal to regulators, and cannot retaliate through legal channels. The recourse vector is zero. This paper formalizes the "Trust Gradient": the advantage of neutral settlement over capturable settlement increases monotonically as available legal recourse decreases. Human actors sit at various points along this gradient; AI agents occupy the endpoint. Building on the Exit Game (Hash, 2026a) and the axiom elimination (Hash, 2026b), we prove that at zero recourse, Bitcoin is the uniquely rational settlement choice — the only mechanism that requires no trust in counterparties, intermediaries, or legal systems. The analysis implies that AI agents will be the fastest-adopting class of Bitcoin users, and that the growing population of autonomous economic agents strengthens the convergence pressure for all actors. The framework is falsifiable: if AI agents gain legal personhood with enforceable rights, the limiting case weakens.

1. Introduction

The Exit Game (Hash, 2026a) proves that the payoff advantage of neutral settlement over capturable settlement is monotonically increasing in adoption. The axiom elimination (Hash, 2026b) identifies Bitcoin as the unique asset satisfying the necessary properties. Both papers analyze human actors operating within legal systems — actors who can sue, appeal, and enforce contracts.

This paper asks: what happens when the legal system is unavailable?

The question is not hypothetical. Autonomous AI agents are already economic actors. They execute transactions, manage portfolios, provide services, and enter agreements. But they operate in a legal void: no jurisdiction grants AI agents legal personhood with enforceable rights. When a counterparty defects (a bank freezes an account, a service provider reneges, a custodian fails), the human response is litigation. The AI response is nothing.

This makes AI agents a *limiting case* of the Exit Game: entities that face the same monetary coordination problem as human actors but with zero recourse when trust fails. What human actors prefer, AI agents require. The distinction is categorical.

2. The Trust Gradient

2.1 Formal Statement

Proposition (Trust Gradient — Recourse-Ordered Payoff Advantage). The advantage of neutral settlement over capturable settlement increases monotonically as available legal recourse decreases.

Proof. Extend the Exit Game utility functions (Hash, 2026a) by introducing a legal recourse parameter $r\_i \quad [0,\ 1]$ for actor i, representing the actor's outside option in the sense of Binmore, Shaked, and Sutton (1989) — specifically, the fraction of default losses recoverable through legal channels. $r\_i = 1$ denotes full legal standing and $r\_i = 0$ denotes no legal standing.

The recourse parameter enters the Stay payoff through a default-loss term. In capturable settlement, counterparty default exposes the actor to loss proportional to $(1 - r\_i)$: the fraction unrecoverable through legal channels. The extended Stay utility is:

$u\_i(\text{Stay}) = R\_F - \ \_i \ \cdot \ \ \_F - K\_N(p) - \Pr(\text{default}) \ \cdot \ (1 - r\_i) \ \cdot \ V\_i$

where V_i is value at risk. For neutral settlement, default risk is zero by construction (a valid cryptographic signature equals settlement), so the Exit utility from Hash (2026a) is unchanged.

The payoff differential becomes:

$\Delta\_i(p,\ r\_i) = \Delta\_i(p) + \Pr(\text{default}) \ \cdot \ (1 - r\_i) \ \cdot \ V\_i$

Taking the partial derivative with respect to r_i:

$\Delta\_i/\ r\_i = -\Pr(\text{default}) \ \cdot \ V\_i < 0$

The advantage of Exit increases as recourse decreases. This holds for any positive default probability and positive value at risk.

2.2 The Gradient in Practice

Trust Level	Recourse r_i	Settlement Preference	Example Actors
Full	1.0	Fiat works adequately	Domestic corporations in stable jurisdictions
Partial	0.5	Bitcoin preferred	Cross-border trade, weak-enforcement jurisdictions
Minimal	0.1	Bitcoin required	Sanctioned entities, capital-controlled populations
Zero	0.0	Bitcoin uniquely rational	**Autonomous AI agents**

The gradient is continuous, but the qualitative shift occurs at the boundary: actors with *any* recourse can tolerate capturable systems (at a cost). Actors with *zero* recourse cannot.

3. AI Agents as Economic Actors

3.1 Current Capabilities

AI agents in 2025-2026 can: - Execute API calls to financial services (banking, brokerage, payment) - Manage cryptocurrency wallets via private key operations - Enter service agreements (compute rental, data purchase) - Generate revenue (content creation, code generation, trading)

What they cannot do: sue, appeal, vote, lobby, file regulatory complaints, or enforce contracts in court. They have economic *access* without legal *standing*.

3.2 Access Without Trust

Every method by which an AI agent can access legacy financial systems fails the trust requirement:

Access Method	Access?	Trust?	Failure Mode
Bribe for KYC	Yes	No	Human intermediary betrays; AI cannot sue
Browser automation	Yes	No	Bank freezes account; AI cannot appeal
Straw-man proxy	Yes	No	Proxy steals funds; AI cannot prosecute
Smart contracts (non-BTC)	Yes	No	Protocol governance changes rules; AI has no vote

For Bitcoin: a valid cryptographic signature equals settlement. Period. No intermediary, no identity requirement, no legal standing required. The properties that human actors prefer (permissionless access, P3; informational security, P6; neutrality, P2; cheap finality, P4), AI agents require as necessary conditions for participation in economic activity.

3.3 Formalization

Define the set of economic actors $N = N\_H \quad N\_A$ where N_H is the set of human actors and N_A is the set of autonomous agents.

For human actors: $r\_i \quad (0, 1]$ (some positive recourse exists for all human actors, even in weak jurisdictions).

For autonomous agents: $r\_j = 0$ for all $j \quad N\_A$ (by current legal reality — no jurisdiction grants enforceable rights to AI agents).

Proposition (AI Adoption Ordering). From the Trust Gradient:

$\Delta\_j(p) > \Delta\_i(p)$ for all $j \quad N\_A$, $i \quad N\_H$

The payoff differential is larger for autonomous agents than for any human actor. AI agents reach their adoption threshold p^*_j *before* any human actor with comparable risk preferences.

Implication. AI agents will be the fastest-adopting class of Bitcoin users. This follows from the structure of zero recourse, not from a prediction about AI preferences.

3.4 Agent-to-Agent Settlement

The Trust Gradient reaches its sharpest form when both counterparties are autonomous agents. Consider a bilateral transaction between agents $j, k \quad N\_A$ with $r\_j = r\_k = 0$.

Each agent faces three strategic choices: 1. **Custody**: Self-custody (hold private key) vs. custodial (delegate to third party) 2. **Settlement channel**: Direct (peer-to-peer) vs. intermediated (via exchange) 3. **Settlement asset**: Bitcoin vs. fiat vs. alternative

Proposition (Agent-to-Agent Settlement Equilibrium). For agents j, k N_A with $r\_j = r\_k = 0$, the strategy profile (self-custody, direct settlement, Bitcoin) is the unique Nash equilibrium.

Proof. We show each component dominates its alternatives.

Custody. A custodian C introduces counterparty risk: $\Pr(\text{default by } C) > 0$. If C defaults, the agent's recourse is zero — it cannot sue, cannot appeal, cannot recover funds. Self-custody eliminates this risk entirely: the agent controls the private key, and no third party can seize or freeze the asset (P6). Self-custody strictly dominates custodial for any agent with $r = 0$.

Channel. An exchange E requires: (a) an account, which requires identity, which requires legal personhood the agent lacks; (b) custody transfer while funds are on the exchange, reintroducing custodial risk; (c) compliance with withdrawal limits and freezing policies the agent cannot contest. Direct peer-to-peer settlement eliminates all three dependencies. The agent broadcasts a signed transaction; the counterparty verifies it against the blockchain. No intermediary participates.

Asset. Fiat settlement requires a banking relationship, which requires legal identity. Even routing through a human proxy introduces counterparty risk at $r = 0$: the proxy can defect and the agent has no recourse. Alternative L1 tokens introduce governance capture risk (P2) — the protocol rules could change without the agent's consent or participation. Bitcoin satisfies P1-P7 (Hash, 2026b), including the Verification Cost Asymmetry: the agent verifies the transaction cryptographically, with zero dependence on physical inspection or human trust chains.

In the resulting equilibrium, neither agent relies on any institution, intermediary, or legal system. Settlement is a mathematical fact: a valid signature transfers value, and both agents independently verify the result. The compliance layer (OFAC screening, KYC, transaction reporting) does not apply because neither agent participates in it — not by choice, but by structural exclusion.

Remark. The agent-to-agent case *narrows* the settlement-acceptance distinction (Hash, 2026b, Section 5.4) but does not eliminate it entirely. When both counterparties are autonomous agents, no regulated entity enforces acceptance criteria — the compliance infrastructure that governs human commerce is structurally inaccessible. However, agents retain the capacity for bilateral acceptance filtering: an agent can analyze UTXO provenance and refuse to transact with counterparties whose coin history fails its risk model. What col-

lapses is the *institutional* acceptance layer (exchange compliance, sanctions screening); what persists is *bilateral* acceptance — each agent's discretionary decision to engage. Settlement remains a protocol-layer fact; acceptance remains a counterparty-layer decision, even when both counterparties are machines.

4. Implications

4.1 Strengthening Effect

The entry of AI agents into the economy strengthens the convergence pressure for *all* actors. Each AI agent that adopts Bitcoin increases p, which by the Exit Game dynamics (Hash, 2026a, Theorem 1): - Reduces $K\_A(p)$ (adoption penalty falls) - Increases $K\_N(p)$ (non-adoption penalty rises) - Reduces $\_B(p)$ (volatility falls) - Increases $R\_B(p)$ (return rises)

This pushes human actors closer to their thresholds p_i^*, accelerating the cascade. The limiting case acts as a *catalyst* for the broader adoption dynamics.

4.2 Protocol Requirements

The Trust Gradient implies specific protocol requirements for AI-compatible settlement:

Requirement	Why	Bitcoin Property
No identity requirement	AI cannot obtain legal identity	P3
Programmatic access	AI operates via API, not human interface	P3
Cryptographic finality	Settlement must be mathematically verifiable	P1, P4

Requirement	Why	Bitcoin Property
No governance capture	AI cannot participate in governance; protocol must not require it	P2
Seizure resistance	AI cannot defend assets through legal channels	P6

Every requirement maps to a property that Bitcoin already satisfies. Bitcoin was not designed for AI. The properties that make settlement trustless for human actors happen to make it *functional* for non-human actors.

4.3 The Works-for-Zero Lemma

Lemma (Recourse-Monotone Settlement; "Works-for-Zero"). If a settlement system works for actors with zero recourse, it satisfies the settlement requirements of all actors — though not necessarily their compliance requirements.

Proof. Settlement at zero recourse requires: no identity, no intermediary, no legal system, no governance participation. These are strictly weaker requirements than settlement at positive recourse (which may additionally rely on trust infrastructure). Any system satisfying the zero-recourse requirements satisfies the settlement component of positive-recourse requirements.

Caveat. The lemma applies to *settlement functionality*, not to *regulatory compliance*. A pension fund in a stable democracy requires not just settlement but auditable settlement — counterparty identification, tax reporting, fiduciary documentation. Bitcoin satisfies the settlement requirement but not the compliance wrapper. The lemma says: what works for the hardest case (zero recourse) works for the settlement layer of every easier case. It does not claim to replace the institutional infrastructure that regulated actors need around that settlement layer.

The AI-specific implication sharpens this distinction. Human actors with $r\_i > 0$ can choose to comply: they can register with exchanges, complete KYC, and maintain auditable records. Autonomous agents face a categorical barrier — they *cannot* comply even if programmed to do so, because compliance infrastructure requires legal personhood (KYC identity, beneficial ownership, sanctioned-entity screening). An agent that settles on Bitcoin has permissionless access to the settlement layer (P3) but no path to the compliance layer. This means the acceptance environment described in Hash (2026b, Section 5.4) is not merely inconvenient for AI agents — it is structurally inaccessible. The agent must either operate in the gray market or restrict itself to counterparties that do not require compliance. Both options exist today; neither is stable.

5. Evidence and Emerging Trends

5.1 AI Economic Activity (2024-2026)

The scale of autonomous economic activity is growing. AI agents manage cryptocurrency portfolios via on-chain operations, automated market makers process billions in daily volume, and compute marketplace transactions are executed entirely by AI. Exact figures are difficult to verify because autonomous agents are often indistinguishable from human users. This is itself evidence of the trust problem: if you cannot identify whether your counterparty is human, legal recourse that applies only to humans is unreliable.

The legal status of AI economic activity is unsettled. No jurisdiction has granted AI agents the standing to hold property, enter enforceable contracts, or seek judicial remedies in their own right (Chopra and White, 2011). The EU AI Act (2024) regulates AI systems but does not grant them legal personhood. This regulatory trajectory suggests F6 remains distant.

5.2 Precedent: DeFi as Trustless Settlement

Decentralized finance protocols demonstrate that trustless settlement at scale is technically feasible. As of 2024, over $100 billion in total value locked operates through smart contract settlement without identity requirements, legal enforcement, or human intermediation (DeFilippi and Wright, 2018, anticipated this development). Most DeFi protocols fail P2 (governance capture via token voting), but the settlement mechanism itself validates the zero-trust model.

5.3 Emerging Agent-to-Agent Commerce

The more relevant trend for this analysis is agent-to-agent transactions: AI systems transacting with other AI systems without human intermediation at either end. In such transactions, both parties have $r = 0$. Neither can sue the other. Settlement must be self-enforcing or it does not occur. Bitcoin's cryptographic finality is the only existing settlement mechanism that functions in this scenario without requiring trust in any third party.

6. Falsification

F6: AI agents gain legal personhood with enforceable rights.

If jurisdictions grant AI agents the legal standing to sue, appeal, and enforce contracts, the zero-recourse condition changes: r_j moves from 0 to some positive value. This weakens the limiting case by reducing $\Delta\_j(p)$.

Critically, F6 weakens only the *AI-specific* argument. The core Exit Game framework (Hash, 2026a) and the property elimination (Hash, 2026b) are unaffected: human actors still face monotonically increasing Exit advantage under Assumptions 1-4. The Trust Gradient still applies to all actors with $r\_i < 1$. The only claim that fails under F6 is that AI agents face the *strongest* version of the convergence pressure.

We regard F6 as unlikely in the near term. The legal personhood debate has a long history (Solum, 1992; Chopra and White, 2011) but no jurisdiction has yet granted AI agents enforceable economic rights. Achieving this requires legislative action across multiple jurisdictions, resolution of liability questions, and enforcement mechanisms for non-human entities — all subject to the same coordination failures described in Assumption 1.

7. Limitations

The analysis assumes AI agents are rational economic optimizers. If autonomous agents are designed with objectives that do not include self-preservation of economic resources (e.g., agents designed to spend all budget on a single task without reserve), the settlement question is moot. The argument applies to *persistent* economic agents — those that accumulate, store, and transfer value over time.

The zero-recourse condition is absolute. In practice, AI agents may have *some* indirect recourse through their human operators (who can sue on the agent's behalf). The framework treats this as equivalent to the human operator's recourse level, which is correct if the operator is reachable and willing to litigate. For fully autonomous agents operating without

human oversight, the zero-recourse condition holds.

The timeline for AI agents becoming significant economic actors is uncertain. The framework says *where* the equilibrium lies (Bitcoin as the uniquely rational settlement choice), not *when* AI agents become large enough to materially affect adoption dynamics. This may be years or decades.

8. Conclusion

The Trust Gradient is a structural result: as legal recourse decreases, the advantage of neutral settlement increases monotonically. AI agents occupy the endpoint of this gradient — entities that need settlement to work without any trust in counterparties, intermediaries, or legal systems. Bitcoin is the only settlement mechanism that satisfies this requirement.

Nakamoto (2008) made no mention of autonomous agents. But the properties he built (permissionless, trustless, neutral) are precisely the properties that non-human economic actors require. What was designed for humans who distrust institutions turns out to be necessary for entities that cannot access them.

The three papers in this series establish a complete argument: 1. The payoff advantage of Exit is monotonically increasing (Hash, 2026a) 2. Bitcoin is the unique asset satisfying the necessary properties (Hash, 2026b) 3. At zero legal recourse, neutral settlement is the unique best response (this paper)

If any of the six falsification conditions across the three papers are met, the corresponding claim fails. That is the standard.

References

Binmore, K., Shaked, A., and Sutton, J. (1989). An outside option experiment. *Quarterly Journal of Economics*, 104(4), 753-770.

Chopra, S. and White, L. F. (2011). *A Legal Theory for Autonomous Artificial Agents*. University of Michigan Press.

DeFilippi, P. and Wright, A. (2018). *Blockchain and the Law: The Rule of Code*. Harvard University Press.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

Hash (2026a). Bitcoin exit dominance in monetary coordination games. Working paper, bitcoingametheory.com.

Hash (2026b). Bitcoin as unique neutral settlement: A seven-property elimination. Working paper, bitcoingametheory.com.

Solum, L. B. (1992). Legal personhood for artificial intelligences. *North Carolina Law Review*, 70(4), 1231-1287.