Economics Bulletin

Volume 45, Issue 4

Proof of Honesty: Blockchain consensus with 99% Byzantine Fault Tolerance

John P. Conley
Vanderbilt University

Abstract

Proof of Stake (PoS) offers only 33% BFT in the sense that if a larger percentage of the validating network fails to follow protocol, there is no guarantee of honest blockchain validation. We show the PoS mechanism generates a coordination game. Thus, if nodes make self-interested choices over whether to follow protocol (honesty or dishonesty), anything can be supported as a Nash equilibrium. Nodes benefit from coordinating on dishonest validation, but honest validation, and failing to commit both honest and dishonest blocks, are also equilibria. This result persists regardless of how much nodes stake, or how diverse that staking pool might be. We define a simplified version of the Proof of Honesty (PoH) blockchain protocol. PoH leverages the fact that blockchains are deterministic, and so users are able to distinguished correct from incorrect chain states without the need for consensus. A single node presenting a correct chain view is therefore sufficient, since users can ignore false forks and continue to trade tokens of a fork they know is correct. As a result, Proof of Honesty is (N − 1)/N BFT (which we style "99% BFT"). We also show that PoH implements honest blockchain validation in Nash equilibrium with relatively modest stakes.

Proof of Honesty is an element of a more general protocol developed for Geeq, a layer 0 infrastructure blockchain project in which the author is a participant. The author takes sole responsibility for the contents of this paper. Editor`s Note: This paper was originally submitted under a different manuscript number on 11/18/2025 and accepted for publication on 12/29/2025. A system error required us to replace the original submission with a new manuscript with a later manuscript number.

Citation: John P. Conley, (2025) "Proof of Honesty: Blockchain consensus with 99% Byzantine Fault Tolerance", *Economics Bulletin*, Volume 45, Issue 4, pages 1672-1686
Contact: John P. Conley - j.p.conley@vanderbilt.edu.
Submitted: December 29, 2025.   **Published:** December 30, 2025.

1. Introduction

Castro and Liskov (1999) propose a mechanism for state machine replication in asynchronous networks with faulty nodes. In the context of blockchain, a fixed set of nodes who recognized one another's authority would work as follows: Each round, one node is selected as the leader and proposes a set of transactions to update the current ledger state. If two-thirds of the nodes in the network agree that the chosen transactions are valid, then each node uses them to identically update their copy of the ledger. A new leader is then selected for the next round, and the process repeats.

This **Proof of Authority** (**PoA**) consensus protocol is the foundation for the family of **Proof of Stake (PoS)** protocols in which the probability of being selected as leader and voting weights are proportional the amount of native token each node chooses to stake. Both PoA and PoS are 33% **Byzantine Fault Toleran**t (**BFT**) in the sense that if more than one-third of the nodes are dishonest or faulty, then correct transaction validation and ledger updates cannot be guaranteed.

Viewed as a mechanism, PoS consensus does poorly at incentivizing self-interested nodes to behave honestly. We show that PoS is in effect a coordination game with many Nash equilibria. Nodes prefer to coordinate on dishonest validation. This result is true regardless of how much is staked, or how widely distributed voting power might be.

We describe an alternative protocol called **Proof of Honesty** (**PoH**) and show that it is 99% BFT. We also show PoH implements honest validation in Nash equilibrium with staked **Good Behavior Bond**s (**GBB**) that are far smaller than are typical in PoS blockchains. The results in this paper are provided for one-shot versions of a blockchain game. However, these results generalize, and are in fact strengthened, in infinitely repeated versions of the game (see Conley 2018 for a full treatment).

1.1. A Short Discussion of Related Literature

Many papers have drawn attention to real-world, non-protocol forks in Bitcoin, Ethereum, and other blockchains that have ultimately been accepted as canonical by both users and the validating network (Chin *et al.* (2020), Zhang (2021), Ahn, *et al.* (2024) and Jiao, *et al.* (2024), for example), or detail specific sorts of attacks that selfish validators might employ to profit through violations of protocol (see Brown-Cohen, *et al.* (2019), Liu, *et al.* (2019), D'Amato *et al.* (2024), among many others).

The closest work to the current paper is Biais. *et al.* (2019) who explore a dynamic PoS blockchain game and find a kind of Folk Theorem for Markov Perfect Equilibrium. Their focus is on equilibria in which the validating network chooses between forks, especially hard forks, involving upgrades to protocol (which are *ipso facto* deviations from current protocol). Conley (2025) explores a similar game and finds a Folk Theorem in Subgame Perfect Equilibria that includes all possible deviations from protocol. Interestingly, the Folk Theorem holds independently of any expectations the network might have regarding stake devaluation due to a loss of user confidence in a dishonest chain.

The vast majority of work on blockchain is published in the information and computer science literatures. Most of the economics literature is either empirical or applied work in finance, banking, IO, and other fields. The purely game theoretic literature on blockchain in economics is limited.

2. The Model

Ledgers: $\mathcal{L}$, the set of permissible ledger states, $L \in \mathcal{L}$, containing records and accounts holding native tokens controlled by users with **Public-Private Key** (**PPK**) pairs.

Blocks: $\mathcal{B}$, the set of permissible blocks, $B \in \mathcal{B}$, containing transactions that move tokens from account to account in the ledger.

Protocol Compliance Indicator Function (PCIF): PCIF: $\mathcal{L} \times \mathcal{B} \Rightarrow \{\text{corr}, \text{incorr}\}$.

Blockchains are defined by protocol rules which determine whether a set of transactions in a block is correct given the current ledger state, details contained in the transactions, and how transactions relate to one another. For example, transactions must address an account that exists in the current ledger state, be correctly signed, and the collection of transactions addressing a given account must not jointly overspend its balance. Blocks that satisfy all protocol-defined criteria are deemed correct by the PCIF, and incorrect otherwise.

We model a one period blockchain consensus game as follows:

Agents: $n \in (1, \ldots N) \equiv \mathcal{N}$.

Nodes: $n \in (1, \ldots [N-1])$.

Block Leader: N: a single agent designated to propose the next block for a chain.

Action Space for Nodes: $\mathcal{A}_n \equiv \{\text{acc}, \text{rej}\}\ \forall\ n \in \{1, \ldots [N-1]\}$: nodes vote to accept or reject.

Action Space for Leader: $\mathcal{A}_N \equiv \mathcal{B}$: the set of blocks that the leader can choose to propose.

State-Transition Machine: STM : $\prod_{n \in \mathcal{N}} \mathcal{A}_n \times \mathcal{L} \Rightarrow \mathcal{L}$

Transactions update records in the current ledger to a new state following a deterministic, protocol defined, **State-Transition Machine**. Thus,

$$\text{STM}((a_1^*, \ldots a_N^*), L^0) = L^1$$

returns the updated state of the records in ledger L^0 given the block of transactions, $B = a_N^*$, and an action choice by each of the nodes, where the asterisk indicates an element that is cryptographically signed with the private key of the subscripted agent.

Note that the STM incorporates the consensus rules of a blockchain's protocol, and is well-defined for both correct and incorrect blocks. For example, incorrect or rejected blocks might result in an identity mapping and/or in penalties being assessed to certain agents.

2.1. Strategies

In general, a strategy for nodes would be a mapping from the space of initial ledger states and proposed blocks into the action space, $\mathcal{A}$. In the interest of simplicity, we will assume that agents choose one of two reduced form strategies: Honesty or Dishonesty.

Honesty and dishonesty for nodes, h_n, d_n: $\mathcal{L} \times \mathcal{B} \Rightarrow \mathcal{A}_n \equiv \{\text{acc, rej}\}$ are defined as follows:

$$h_n(L^0\ B) = \begin{cases} \text{acc} & \text{if PCIF}(L^0, B) = \text{corr} \\ \text{rej} & \text{if PCIF}(L^0, B) = \text{incorr} \end{cases}$$

$$d_n(L^0, B) = \begin{cases} \text{rej} & \text{if PCIF}(L^0, B) = \text{corr} \\ \text{acc} & \text{if PCIF}(L^0, B) = \text{incorr} \end{cases}$$

Similarly, honesty and dishonesty for the leader, h_N, d_N : $\mathcal{L} \Rightarrow \mathcal{A}_N \equiv \mathcal{B}$ are defined as follows:

$$h_N(L^0) = B \in \{\overline{B} \mid \text{PCIF}(L^0, B) = \text{corr}\}$$

$$d_N(L^0) = B \in \{\overline{B} \mid \text{PCIF}(L^0, B) = \text{incorr}\}$$

A **Strategy Profile** is a choice of either h_n or d_n for each agent:

Strategy Profile for Agents: $(s_1, \ldots s_N) \in (\mathcal{S}_1, \ldots \mathcal{S}_N) \equiv \mathcal{S}$, where $\forall$ n $\in \mathcal{N}$, $\mathcal{S}_n \equiv \{h_n(\ ), d_n(\ )\}$.

2.2. Payoff and Equilibrium

A **Payoff Function** is a mapping from strategies into the real numbers.

Payoff Function: $F \equiv (F_1, \ldots F_N)$ where $\forall$ n $\in \mathcal{N}$, $F_n : \mathcal{S} \Rightarrow \mathbb{R}$.

and **Nash Equilibrium** is defined in the usual way:

Nash Equilibrium: A strategy profile, $s \in \mathcal{S}$, is a Nash equilibrium if there does not exist an agent, $m \in \mathcal{N}$, and alternative strategy, $\hat{s}_m \in \mathcal{S}_m$, such that $F(s_{-m}, \hat{s}_m) > F_m(s)$.

We assume the following:

No Equivocation: The block leader proposes only one block. That is, we assume that there is never an equivocation where a leader proposes, signs, and then sends, two different blocks to the nodes in the validating network. Likewise, nodes vote once, and only once, to either accept or reject a block proposal.

Non-Partitioned Network: All nodes see the proposed block, are able to vote, and see the votes of all other nodes in the network.

Deterministic Protocol: While there may be many correct blocks that might be proposed by the leader, the protocol rules as defined by the PCIF unambiguously determine if a block is correct or incorrect given the initial chain state. Similarly, the STM defines the unique, correct, updated ledger state given the initial ledger state and an action profile for agents (which includes the proposed block).

Leader Votes to Accept: By proposing a block, the leader is assumed to cast his vote to accept it.

We will use these rules and the following notation to calculate the specific payoff functions for PoS and PoH below:

H – The number of agents who choose **Honesty** as a strategy.

D – The number of agents who choose **Dishonesty** as a strategy: $D + H = N$.

F – The **Fees** paid collectively to the validation network in native token.

L – The number of native tokens that agents can collectively **Loot** through dishonest behavior.

G – The amount staked as a **Good Behavior Bond (GBB)** by each agent in native token.

δ – The per period **Discount Factor**.

In order to participate in the validating network, agents must stake G in native tokens. This has an opportunity cost regardless of their choice of actions, or the outcome of the game, equal to:

$$(1 - \delta)G$$

3. Proof of Stake

Proof of Stake and its variations, are the most common consensus rule in use by blockchains today. We describe a simple version of PoS in which all nodes stake the same GBB. This special case is closer to the original PoA approach, but we will argue below that the differences are immaterial.

In PoS, the leader proposes a block, signs it, and circulates it to all the nodes in the validating network. Nodes vote to accept or reject the block, and then send their signed votes to all other agents. Protocol rules are as follows:

- If two-thirds of the votes from agents, $(a_1^*, \ldots a_N^*)$, are to accept the block then:

  $$STM((a_1^*, \ldots a_N^*), L^0) = L^1$$

  where

  - All transactions in $B = a_N^*$ are deemed correct and used to update the ledger, even if $PCIF(L^0, B) = incorr$.

  - All nodes who vote to reject B have their GBBs confiscated,
  - All nodes who vote to accept B get an equal share of the fees, loot, and the sum of any confiscated GBBs. Note that assuming equal division of fees and confiscated GBBs for correct blocks can be enforced by protocol. Assuming equal division, or any specific division, of payoffs for incorrect blocks is a simplification. We will offer a justification below.
- If less than two-thirds of the votes from agents, $(a_1^*, \ldots a_N^*)$, are to accept the block then:

$$STM((a_1^*, \ldots a_N^*), L^0) = L^0$$

and the block is discarded. No transactions are executed, and no penalties are assessed.

3.1. Payoff Function

Table 1: PoS Payoff Function for Agents: $F : \mathcal{A} \Rightarrow \mathbb{R}$

Outcome	Vote	
	Accept	Reject
Correct Block Committed	(1) $\frac{F+DG}{N-D}-(1-\delta)G$ Honest Node Honest Leader	(2) $-G-(1-\delta)G$ Dishonest Node Honest Leader
Correct Block Not Committed	(3) $(1-\delta)G$ Honest Node Honest Leader	(4) $(1-\delta)G$ Dishonest Node Honest Leader
Incorrect Block Committed	(5) $\frac{F+L+HG}{N-H}-(1-\delta)G$ Dishonest Node Dishonest Leader	(6) $-G-(1-\delta)G$ Honest Node Dishonest Leader
Incorrect Block Not Committed	(7) $(1-\delta)G$ Dishonest Node Dishonest Leader	(8) $(1-\delta)G$ Honest Node Dishonest Leader

Explanation:

- If a block is not committed, then all agents pay the opportunity cost of staking their GBB, but receive no other payoffs or penalties: (3), (4), (7), and (8).
- Agents who vote to reject a block which is ultimately committed lose their GBBs in addition to paying the opportunity cost of staking: (2) and (3)
- Agents who vote to accept a correct block which is ultimately committed share equally in all fees and the sum of the GBBs of any nodes who vote to reject the block: (1)
- Agents who vote to accept an incorrect block which is ultimately committed share equally in all fees, the sum of the GBBs of any nodes who vote to reject the block, and the loot extracted from the chain by their dishonest behavior: (5).

Leaders are always assumed to vote in favor of the block they propose and so receive the payoffs in the gray boxes.

3.2. Nash Equilibrium:

Theorem 1: Under PoS, committing a correct or incorrect block, and failing to reach positive consensus to commit a correct or incorrect block, are all Nash equilibria.

Proof: See Appendix.

■

Proof of Stake is essentially a coordination game. All agents benefit (at least weakly) from voting the same way regardless of the correctness of the block. Worse still, it is in the interest of the all agents to coordinate on an incorrect block, if they can, since this gives them strictly more payoff than a correct block.

Both correct and incorrect blocks can appear in equilibrium, and both types can be rejected, resulting is no block being committed. These results are independent of the number of nodes, or the size of the GBBs (or to any asymmetric staking profile arising from more complex PoS protocols). The conclusion is that PoS protocols offer extremely weak security guarantees for users.

In practice, there are probably two factors that support coordination on honest validation.

First, nodes may speculate that the external value of tokens will be negatively impacted if users observe dishonest behavior. This could result in the *external* value of (F + L) on a dishonest chain being smaller that F on an honest one (and cause additional loses to the value of staked tokens).

Second, nodes may speculate that significant dishonesty will result in a reorganization of the chain by an external agent (foundations, founders, exchanges, or large token-holders, for example). Other users may come to a social consensus to accept this protocol-violating fork in preference to the incorrect one offered by the validating network. Note that this obviates the consensus mechanism entirely. The security guarantee ends up relying on nodes' expectations, the actions and motivations of external agents, and an uncertain potential social consensus, instead of the mechanisms define by PoS.

4. Proof of Honesty

Proof of Honesty takes advantage of the fact that all well-defined blockchain protocols are deterministic. While a mechanism must be implemented to choose a canonical block of candidate transaction for consideration, there is no real need to vote on its correctness. It makes no sense to vote about the truth of something that can be proven. Just because a majority tells you red is blue does not make it so. Your own two eyes are perfectly capable of making this determination their own.

PoH is actually **non-consensus protocol** implemented as follows: The leader distributes his chosen, signed, block to all nodes. Each node receives the block proposal, votes to accept or reject the block, signs his vote with his private key, circulates his vote to all other agents in the validating network, and then creates his own **Chain State View (CSV)** which he also signs with his private key:

$$CSV_n{}^* \equiv ((a_1{}^*, \dots a_N{}^*), L^0, L^1)_n{}^*$$

A CSV includes the agent's view of the all votes and the block proposed by the leader. The key observation is that users can verify if a $CSV_n{}^*$ is correct by checking whether:

- All actions are correctly signed by the correct agent: $(a_1{}^*, \dots a_N{}^*)$.
- $PCIF(L^0, B) = corr$, where $a_N{}^* = B$ and L^0 is common knowledge.
- $STM((a_1{}^*, \dots a_N{}^*), L^0) = L^1$.

The main innovation in PoH is that each agent separately and independently relies on his own opinion about the correctness of the block to arrive at a new ledger state, and then presents his own CSV to users. This means that the PoH does not require a two-thirds majority, or any majority at all, to commit a block or update the ledger.

PoH protocol rules are as follows:

- All nodes who vote differently than the node offering its view of the ledger have their GBBs confiscated on that ledger state.
- All nodes who vote the same as the node offering its view of the ledger get an equal share of fees and any GBBs that are confiscated on that ledger state.

We will also maintain the assumption that all nodes who vote to accept an incorrect block get an equal share of the loot, although as above, this is not a protocol rule.

If all nodes are honest, then all will present users with identical and correct CSVs.

If all nodes are dishonest, then no node will present users with a correct CSV. These false CSVs may or may not be identical, but all are verifiably false. Users, however, do not have a correct chain on which to trade tokens in this event. They are left with the choice of abandoning their tokens, or accepting the dishonest behavior of the validating network and continuing to trade tokens on an incorrect ledger.

If, however, at least one node is honest, then he presents a correct CSV to users independently of the choices made by other nodes. Since users can verify that the honest node's CSV is correct, they have an honest, protocol-compliant, chain on which to trade their tokens. They can therefore ignore CSVs offered by any dishonest nodes. This implies the following Theorem:

Theorem 2: PoH is $(N - 1)/N$ Byzantine Fault Tolerant.

Proof: Suppose one agent always chooses honesty without regard to self-interest or strategic considerations. Then the agent presents a correct CSV to users, and users will always choose this uniquely correct CSV regardless of the actions or views of the other agents in the validating network.

■

We call this “99% BFT” to make the comparison to 33% BFT for PoA and PoS, and 50% BFT for PoW protocols, clear. Of course, $(N - 1)/N$ BFT may be more or less than 99% depending on the size of the validating network.

4.1. Payoff Function

Table 2 : PoH Payoff Function for Agents: $F : \mathcal{A} \Rightarrow \mathbb{R}$

Outcome	Vote	
	Accept	Reject
Correct Block	(1) $\frac{F + DG}{N - D} - (1 - \delta)G$ Honest Node Honest Leader	(2) $-G - (1 - \delta)G$ Dishonest Node Honest Leader
Incorrect Block At least one other Honest Node	(3) $-G - (1 - \delta)G$ Dishonest Node Dishonest Leader	(4) $\frac{F + DG}{N - D} - (1 - \delta)G$ Honest Node Dishonest Leader
Incorrect Block All other Nodes Dishonest	(5) $\frac{F + L}{N} - (1 - \delta)G$ Dishonest Node Dishonest Leader	(6) $F + (N - 1)G - (1 - \delta)G$ Honest Node Dishonest Leader

Explanation:

- If the leader proposes a correct block, then users automatically have access to the correct chain state, and so ignore any CSVs that differ. Thus, nodes that vote to accept the correct block confiscate the GBBs of all dishonest nodes who vote to reject, and along with fees, share them equally with all other nodes who vote to accept, which results in the payoff in box (1). Voting to reject only results in a node’s GBB being confiscated on the only non-orphaned fork, and the payoff in box (2).
- If the leader proposes an incorrect block, but there is at least one honest node who votes to reject, then again, users are presented with a correct CSV. Thus, nodes that vote to accept the incorrect block get their GBBs confiscated on the chain that users choose, which results in the payoff in box (4), while any node that votes to reject it shares equally in fees and confiscated GBBs and gets the payoff in box (3).
- If the leader proposes an incorrect block, and all notes vote to accept it, then users are only presented with an incorrect CSV. In this case, all agents share equally in fees and the loot

taken as a result of their dishonesty giving them the payoff in box (5). If a single node chooses to defect from this conspiracy, however, he gets all the fees and the GBBs confiscated from the $(N - 1)$ dishonest agents on the verifiably correct chain he creates, and which is chosen by all users, and gets the payoff in box (6). As a result, dishonest agents find themselves getting the payoff in box (3) instead of box (5).

4.2. Nash Equilibrium

Theorem 3: Under PoH, if $G > \overline{GBB}(F, L, N) \equiv \frac{L - (N-1)F}{(N^2 - N)}$, then universal honesty by all agents is the only Nash equilibrium.

Proof: See Appendix.

■

Under PoH, honest validation is the unique Nash equilibrium given the correct combination of GBB stakes and network size. Significantly, for any given potential of loot from dishonest behavior, the GBB required to secure honest validation *decreases with the square* of the number of nodes.

Sybiling is of no direct benefit under PoH. What matters is the number of independent agents. If one of these agents created a number of Sybils, he would have to deposit a GBB for each. Since one honest agent offering a correct CSV undermines a dishonest coalition, generating Sybils to join such a conspiracy does nothing to impair or disincentivize other independent agents from acting honestly. In fact, it only increases the reward to other agents who choose honesty. Similarly, larger fees increase the reward for honesty, and so decrease the GBB needed to penalize dishonesty.

Table 3: Threshold GBB assuming $F = 0$.
$\overline{GBB}(F, L, N)$

L	N	GBB
10,000	100	~ 1
1,000,000	100	~ 100
1,000,000,000	1000	~ 1000

Table 3 gives a sense of the power of PoH. With 1000 nodes, a GBB of only $1000 worth of native token is sufficient to secure a value of $1 billion in equilibrium. In contrast, Ethereum, the leading PoS blockchain, has approximately 14,000 nodes, each staking about $90,000. With this configuration, a PoH chain could secure $\$1.8 \times 10^{13}$ or about 20 trillion dollars of value (approximately 50 times Ethereum's total market cap as of December 2025).

5. Conclusion

Under PoH, if one node is honest as a matter of type, as opposed to strategic calculation, then it does not matter what the other nodes do. If users have access to at least one honest view of the chain state, then all will choose it. If there is a dishonest fork, even if this view of the chain state is identically held by all other nodes, users will ignore it in favor of the correct view. Thus, one honest node is required to get honest blockchain validation ("99% BFT"). Fortunately, PoH does not rely on arational, altruistic, node behavior. Instead, it actually implements honest validation by all nodes in Nash equilibrium with very modest staking requirements.

In contrast, PoS requires that two-thirds of nodes to be honest (33% BFT). Honest validation is one of many Nash equilibria since PoS is effectively a coordination game. Nodes would actually prefer to coordinate on dishonest equilibria, which suggest that elements external to PoS as a mechanism are responsible for its relative success. Unfortunately, large stakes, and large, diversified, validator pools, do not change these negative results.

6. Appendix 1: Proofs of Theorems

Theorem 1: Under PoS, committing a correct or incorrect block, and failing to reach positive consensus to commit a correct or incorrect block, are all Nash equilibria.

Proof:

1. Correct Block – Committed: Suppose that all nodes choose honesty. A single node deviating and voting dishonestly instead would not change the outcome, but would receive the strictly smaller payoff in box (2) instead of box (1). If the leader deviates and chooses dishonesty instead, the honest nodes would unanimously reject the incorrect block, giving the leader the strictly smaller payoff in box (8). Thus, all agents choosing honesty and committing a correct block is a Nash equilibrium.
2. Correct Block – Not Committed: Suppose that half the nodes choose honesty. A single node deviating to either honesty or dishonesty is not enough to change the outcome. The block would still not be committed, and the node would get the payoff in box (4) which is identical to box (3). If the leader deviates to dishonesty in this case, the honest half of the nodes would reject the incorrect block which is enough to prevent it from being committed, giving the leader the payoff in block (8) which is identical to the payoff in box (3). Thus, failing to reach positive consensus and rejecting a correct block is a Nash equilibrium.
3. Incorrect Block – Committed: Suppose that all nodes choose dishonesty. A single node deviating to honesty would not change the outcome, but would result in his receiving the strictly smaller payoff in box (6) instead of box (5). If the leader deviates and chooses honesty instead, the dishonest nodes would unanimously to reject the correct block, giving the leader the strictly smaller payoff in box (4). Thus, all agents choosing dishonesty and committing an incorrect block is a Nash equilibrium.
4. Incorrect Block – Not Committed: Suppose that half the nodes choose honesty. A single node deviating to either honest or dishonesty is not enough to change the outcome. The block would still not be committed, and the node would get the payoff in box (8) which is identical to box (7). If the leader deviates to honesty in this case, the dishonest half of the nodes would reject the correct block which is enough to prevent it from being committed,

giving the leader the payoff in box (3) which is identical to the payoff in box (7). Thus, failing to reach positive consensus and rejecting an incorrect block is a Nash equilibrium.

■

Theorem 3: Under PoH, if $G > \overline{GBB}(F, L, N) = \frac{L - (N - 1) F}{(N^2 - N)}$, then universal honesty by all agents is the only Nash equilibrium.

Proof: Note that:

$$G > \overline{GBB}(F, L, N) = \frac{L - (N - 1) F}{(N^2 - N)}$$

if and only if:

$$F + (N - 1)G - (1 - \delta)G > \frac{F + S}{N} - (1 - \delta)G$$

That is, if the GBB is above the threshold value, then the payoff given in box (6) is greater than the payoff given in box (5) in the payoff function matrix for PoH (Table 2). Given this:

1. Suppose that the leader proposes an incorrect block, and all nodes choose dishonesty. Then the payoff to a single node for deviating and choosing honesty in given in box (6) which is larger than the payoff in box (5) by hypothesis. Thus, proposing an incorrect block, and all nodes choosing dishonesty, is not a Nash equilibrium.
2. Suppose that the leader proposes an incorrect block, and at least one node chooses honesty. If the leader deviates and proposes a correct block, he gets the payoff given in box (1) which is larger than the payoff in box (3). Thus, proposing an incorrect block when at least one node chooses honesty is not a Nash equilibrium.
3. Finally, suppose that the leader proposes a correct block, and all nodes choose honesty. Then the payoff to a single node for deviating and choosing dishonesty in given in box (2) which is smaller than the payoff in box (1). If the leader deviates and proposes an incorrect block, he gets the payoff given in box (3) which is smaller than the payoff in box (1). Thus, proposing a correct block, and all nodes choosing honesty and voting to accept it, is a Nash equilibrium.

It follows that universal honesty is the only Nash equilibrium of the PoH game.

■

7. Appendix 2: Extensions

Stake Weighted Voting: We only consider equally weighted voting. If agents in PoS staked differently, the results would not change since PoS remains a coordination game.

Note that the GBB in PoH is not like the stake in PoS. GBBs affect neither the odds of being selected as block leader, nor the relative voting power of nodes. More importantly, its level affects the PoS security guarantee of the protocol only at threshold level rather than continuously. Staked tokens in PoS, on the other hand, affect both block leader selection and voting weight. It is widely believed that the security guarantee is proportional to stake. This paper shows that the total stake has no effect of PoS security. PoS is a coordination game, and this remains the case at any staking level.

Unequal Division of Rewards: Equal division of loot and confiscated GBBs cannot be enforced by protocol for incorrect blocks by definition. The block leader is allowed to propose anything he wishes, and so can apportion rewards arbitrarily. Again, it makes no difference in PoS since it remains a coordination game. Security is unaffected by the amount of stake, or the penalties or rewards to dishonest behavior, since deviation by individual nodes does not affect the outcome of the vote over accepting a block.

In PoH, giving one node a larger share of rewards necessarily requires giving another node a smaller share. This weakens the incentive of poorly treated to nodes to participate in a conspiracy to unanimously accept a dishonest block. In other words, it raises the over all cost of satisfying the GBB threshold value for all nodes since some nodes receive smaller shares. Thus, Nash implementation of honest validation is supported at a lower threshold GBB in this case.

Sequential Blockchain Games: Writing an infinitely repeated version of PoS and PoH games is a bit complicated. The apparatus for remembering histories and making reasonable refinements to forward expectations regarding actions by users other nodes is notationally heavy. Nevertheless, the essential results remain.

In PoS, it becomes obvious that a kind of Folk Theorem for subgame perfect equilibria obtains. In itself, this is may not be not surprising. However, it can be shown that the Folk Theorem remains even if nodes expect a catastrophic reaction from users to dishonest behavior. Thus, the story that is often told that the threat of devaluing staked token is what keeps nodes honest is simply not true. The other surprising result is that while a diverse pool of validating nodes does not improve security, concentrated voting power can. If one node holds more than two-thirds of the stake, it fully internalizes the loss of staked value from dishonest behavior. PoS is no longer a coordination game. Thus, depending on the decisive node's expectations regarding user reaction, he may, in fact, be incentivized to behave honestly (Conley 2025)

In PoH, the results in the current paper are strengthened: PoH Implements honest validation in coalition-proof equilibria, given a similar threshold condition on the relative size of the GBB (Conley 2018).

Equivocation: Majoritarian protocols like PoA, PoS have a difficult time dealing with equivocation. If two-thirds of the nodes are dishonest, then they can propose and commit blocks on multiple, incompatible, forks. This would allow an eclipse attack where one fork was made public, and after several blocks had been committed, a second fork would be revealed to replace it. This would allow tokens spent on the first fork to returned to the original accounts, and spent again on the new fork.

A slightly more complicated version of this attack is possible when only one-third of the nodes are dishonest. Protocols typically do not address equivocation, instead simply accepting that if one-third of the nodes are dishonest, then there is nothing to be done on a 33% BFT blockchain.

PoH, on the other hand, is not majoritarian. Equivocation is just another way to extract loot through dishonest behavior. Any node who becomes aware of an equivocation has the same incentives to report it. Thus, eclipse attacks cannot be profitable even for a coalition consisting of a single node. PoH includes a mechanism for choosing forks in this case, but the details are not discussed here.

Partitioned Networks: If one-third of agents are silent in PoS or PoA, then block-writing halts. It is impossible to muster a two-thirds vote for anything in this case. Silence might be a result of a partitioned network, failed or faulty nodes, or a strategic choice by agents. It may also be that a node is not silent, but a coalition of malicious nodes claim not to receive any messages that happen to arrive.

If it were possible to prove why a network is apparently out of full communication, then protocol might be able to offer a response. For example, if it could be proved that node has failed, then it could simply be removed from list of eligible voters. If it could be proved that a partition had occurred, then a fallback rule could be established that chose one component of the partition to continue block-writing. Unfortunately, it cannot be proved why a message is apparently not received.

Failures in communication such as these are foundation for the CAP Theorem (Gilbert and Nancy 2002) which shows that is impossible a distributed system to simultaneously satisfy Consistency, Availability, and Partition Tolerance. Blockchain is a special case of a distributed that uses various protocols to achieve a shared view of the chain state. Thus, we avoid this unsolved problem and simply assume that the network is in full communication.

8. References

Ahn, J., Yi, E., & Kim, M. (2024). Ethereum 2.0 hard fork: Consensus change and market efficiency. *The Journal of The British Blockchain Association*.

Biais, B., Bisiere, C., Bouvard, M., and Casamatta, C. (2019). The blockchain folk theorem. *The Review of Financial Studies*, 32(5), 1662-1715.

Brown-Cohen, J., Narayanan, A., Psomas, A., and Weinberg. S. (2019). Formal barriers to longest-chain proof-of-stake protocols. In Proceedings of the 2019 ACM *Conference on Economics and Computation*, pages 459–473. ACM.

Castro, M. and Liskov, B., (1999), February. Practical Byzantine fault tolerance. In *Proceedings of the Third Symposium on Operating Systems Design and Implementation* Vol. 99, No. 1999, pp. 173-186.

Chen, H., Pendleton, M., Njilla, L., & Xu, S. (2020). A survey on ethereum systems security: Vulnerabilities, attacks, and defenses. *ACM Computing Surveys (CSUR)*, 53(3), 1-43.

Conley, J. (2018). Proof of Honesty: Coalition-Proof Blockchain Validation without Proof of Work or Stake. Manuscript. https://www.geeq.io/wp-content/uploads/2018/08/technical-paper.pdf.

Conley, J. (2025). *A folk theorem for majoritarian blockchain consensus protocols,* Manuscript.

D'Amato, F., Neu, J., Tas, E. N., and Tse, D. "Goldfish: No more attacks on Ethereum?!", In: *Financial Cryptography*, 2024, https://eprint.iacr.org/2022/1171

Gilbert S., Lynch N., (2002). Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services. SIGACT ACM SIGACT News, Volume 33, Issue 2 Pages 51 – 59, https://doi.org/10.1145/564585.564601

Jiao, T., Xu, Z., Qi, M., Wen, S., Xiang, Y., & Nan, G. (2024). A survey of ethereum smart contract security: Attacks and detection. *Distributed Ledger Technologies: Research and Practice,* 3(3), 1-28.

Liu, Z., Luong, N. C., Wang, W., Niyato, D., Wang, P., Liang, Y. C., & Kim, D. I. (2019). A survey on blockchain: A game theoretical perspective. *IEEE Acces*s, 7, 47615-47643.

Zhang, J. (2021). An Anatomy of the Volatility of Cryptocurrency: Evidence from Ethereum and DAO Hack Event. Available at SSRN 3877086.