Automated Penalty Institutions on Staked Blockchain Networks

Eric Alston, Bryce Bugge, and William Lehr[1]

August 15, 2025

Abstract: Blockchain networks provide a novel form of economic organizational coordination through the protocol's ability to facilitate joint production absent the traditional locus of centralized authority that characterizes the private firm. But the joint productive scope of this coordination is limited if it relies exclusively on rewards to incentivize good faith performance. This has led to the emergence of automated penalties, also called "slashing", to better approximate the balance of incentives that firms have long relied upon to motivate employees to exert more effort. Yet just as firms' specific contractual arrangements vary considerably depending on the joint production that they coordinate, staked blockchain networks' penalties vary considerably in practice. We survey the largest staked ("Proof-of-Stake") networks by market capitalization to develop a comparative institutional analysis of their penalties, identifying the actions that typically trigger penalties, common classes of penalties, and means by which penalties are assessed. We detail reasons for this variation including contextual tailoring, random variation, mimicry, and default rules' existence trumping specific rule choices. Finally, we consider the broader applicability of automated penalties in coordinative network contexts, including those on which numerous AI agents will be making risky decisions on behalf of a human principal.

JEL Codes: D21, J30, L22, M12, O33, P51

[1]Eric Alston Eric.Alston@Colorado.EDU, Bryce Bugge Bryce.Bugge@Colorado.edu and William Lehr wlehr@mit.edu. The authors gratefully acknowledge the financial support of the Ethereum Foundation whose views are not necessarily reflected herein.

1. Introduction

Blockchain networks have emerged as a new form of digital coordination, enabling independent users to trust in the distributed character of their governance, in which network validators cannot unilaterally steer outcomes to capture rents. As such, these networks are tractable to analysis through organizational economic theory. Just as corporate governance seeks to productively align the incentives of the especially empowered managers to act on behalf of shareholders and consumers, the design of blockchain networks aims to productively align the incentives of independent validators. But important differences also mean that the traditional institutions of private firms do not cleanly map to these network-coordinated contexts. In particular, the lack of centralized authority means discretionary monitoring of performance is nearly absent on blockchain networks. Instead, these networks must rely on a more complete set of ex-ante-specified rewards and penalties than traditional firms to coordinate joint productive activity. Herein we use organizational economic theory to assess the role of automated penalties on staked blockchain networks, networks whose programming logic enables more complex contractual relationships to emerge among network identities, which makes them more akin to joint productive organizations like the firm than the earliest proof-of-work blockchains such as the Bitcoin network.

Sustaining the integrity, viability, and efficient operation of a blockchain network requires that validators honor their obligations and adhere to the network rules. Validators' incentives to participate and honor their obligations are secured, in part, by the rewards that validators earn in updating the distributed ledger. To sustain those incentives, most proof-of-stake blockchain networks also adopt and enforce penalty policies, which result in validators having their staked rewards reduced or penalties imposed for failure to meet their performance obligations. On these staked blockchain networks, penalties are typically enforced by reducing or "slashing" the stake held by the validators for failure to comply with performance obligations. Although the proliferation of the use of these penalties suggests that they are an important mechanism for governance on staked blockchain networks, there is considerable diversity across these networks in the policies and procedures by which penalties are assessed. Most staked blockchain networks rely on automated penalty procedures that initiate an action based on performance metrics, with different levels of penalty being applicable for different performance infractions. In limited instances, staked blockchain networks may also have due-process and appeals processes and off-chain governance bodies with some level of discretion to modify penalty decisions and policies. Another distinction that we uncover is that some staked networks do not entail automated penalties.

In the following, we present results from an analysis of automated penalty institutions in three principal parts. First, we propose an organizational economic framework for situating the role of automated penalties within the larger context of blockchain governance. Second, we present the results of an empirical inquiry into the penalties among the largest proof-of-stake (PoS) blockchains (by transaction volume/market value) and selected other, smaller blockchains. Third, we offer our hypotheses and suggested avenues for future inquiry to further develop the theory and empirical research on automated penalty institutions and their emergent role in organizing, managing, and operating PoS blockchains and the evolving ecosystem of which they are part. We motivate our typology of penalties on staked blockchain networks with the recognition that penalty policies require specification of the actions that will trigger a penalty, the specification of the penalty itself, and the means by which the penalty is assessed. We survey the variance in triggering actions, penalties, and means of penalty application before considering the reasons for the level of variance in automated penalty institutions we identify on staked blockchain networks. We consider

multiple potential explanations for the variance, including that the variance is the result of deliberate organizational design choice, mimicry, or random variation.

Finally, we consider the broader relevance of automated penalty institutions in light of the growing significance of AI agents being engaged in a growing array of contexts for decision-making under uncertainty. It is in precisely such circumstances that ex ante rules are insufficient to control behavior and outcomes in furtherance of a joint productive purpose, and an allowance for discretionary judgement is needed. Ex post, it is often difficult to determine whether deviations from expected behavior or outcomes was the result of rule violations or the appropriate exercise of discretion. Therefore, we anticipate that the effective deployment of such AI agents will require incorporating reward and penalty policies to better align the incentives of automated agents making decisions on behalf of human principals. Thus, the design of reward and penalty policies that still afford an appropriate scope for discretion that is the focus of our inquiry here should also provide insights of relevance to other realms of digitally-augmented decision-making and production.

2. Economic Context for Understanding Automated Penalties

Rewards and penalties play an essential role in coordinating contributors' (often employees') efforts in the vast majority of joint productive contexts at impersonal scale (Jensen & Meckling 1976; Prendergast 1999). Given this, the field of organizational economics has long considered the appropriate balance of rewards and penalties[2] needed to reduce agency costs and align many agents' decisions with those of the organization (or its principals) (Fama & Jensen 1983; Holmström 1982). For joint productive purposes of sufficiently narrow scope, the act of economic exchange can serve to allocate resources between fully independent parties (Coase 1937; Fama 1980). This provides a lens to understand the incentive alignment created by proof-of-work blockchain networks like that which supports the issuance and transfer of units of bitcoin. But just as many joint productive purposes require a tighter contractual locus of coordination (Coase 1937), so too do many blockchain networks create incentive alignment among contributors by requiring intertemporal commitment to joint productive processes with pre-specified rewards and penalties playing an essential role in this incentive alignment. Motivated by the extensive foundation of rewards and penalties in organizational economics' consideration of incentive alignment, we consider the distinction between proof-of-work ("PoW") and proof-of-stake blockchain networks in this section.

2.1. Rewards and Penalties in Economic Organizational Coordination

Automated penalties appear to be an important component of PoS network protocols to secure them from malicious attacks and to incentivize good performance, given the prevalence of the emergence of these penalties on most networks that rely on staking for incentive alignment among independent validator nodes. Both deterring attacks and aligning productive incentives are potentially necessary to engender and sustain the trust necessary to attract the economic activity

[2] The legal–economic calibration problem—how large penalties should be relative to detection probability—has been studied in public enforcement and contract design; see, e.g., Polinsky & Shavell (2000) and Goetz & Scott (1977) on liquidated damages versus penalties.

the blockchain networks seek to service.[3] Ensuring such trust has been a central concern of protocol design for blockchain networks since their emergence with the advent of the Bitcoin network.

The promise of blockchain networks rests in their ability to offer an automated platform for decentralized semi-automated management of economic transactions. The combination of smart contracts, cryptocurrencies, and the automated digital markets thus enabled hold great promise for expanding economic surplus in digitally-intermediated contexts, but if abused due to poor design, can also result in significant problems including hacks and broader systemic instability.[4] A desirable feature of these networks is that the decentralization of "managerial" control (over decision-making, transaction management, governance, etc.) can reduce the risks of abuses of economic power while also supporting new models for collective governance (such as DAOs).[5] Realizing either of those promises depends on the ability of the protocols supporting blockchains to incentivize good performance, which includes robustness to the threat of malicious attacks. Yet perfect decentralization would give rise to an insufficient concentration of authority to apply discretionary ex-post penalties; in such a limit, the only option would be ex-ante commitment to automatically executing penalties among equally empowered independent contributors to joint production.[6] One key risk is that attackers might capture enough of the blockchain nodes to gain control over the blockchain, inducing a fork to a ledger state that does not conform to valid network transaction requests. The PoW consensus mechanism requires that an attacker capture more than 50% of the nodes to capture the blockchain, but even with less than 50%, an attacker could degrade network performance and thus harm the joint productive purposes the network is designed to coordinate. In the PoW context, the challenge is to capture the computing power supporting the protocol; whereas on staked networks, the challenge is different. On staked networks, the attacker would need to reliably control enough staked validator nodes such that the attacker could force acceptance of malicious blocks. The staking requirements to participate as a validator node and the number of validator nodes that would need to be captured varies across staked blockchain networks. Recent work by Roughgarden (2024) suggests that including penalty policies in protocols, as Ethereum did following its transition from PoW to a PoS consensus protocol in 2022[7] contributes to the robustness and resilience of the blockchain protocol in the face of sybil attacks.

The role of penalization has a long history in economics going back to Adam Smith and his discourse on how market exchange informs the governance of distributed economic activity, and the subsequent political economy literature Smith's seminal work spawned.[8] A more modern view of penalties can situate it in a mechanism design or game theoretic framing of how payoffs to alternative strategies in extensive form games might be altered by the game designer (principal) to induce social choice/welfare (rather than private welfare) enhancing behavior by players in the

[3] In agency cost settings, credible downside sanctions complement rewards to deter shirking and misconduct when monitoring is imperfect; see Shapiro & Stiglitz (1984) on the "discipline device" role of penalties and the effort–deterrence trade-off.

[4] When multiple tasks are weakly observed, pure high-powered rewards can distort effort; in response, calibrated punishments can re-balance incentives across tasks (Holmström & Milgrom 1991).

[5] For example, see Wright (2021).

[6] Formal versus real authority can diverge in organizations; penalties are one lever to discipline misuse of real authority under decentralized decision rights (Aghion & Tirole 1997).

[7] Ethereum merge from proof-of-work to proof-of stake was completed September 15, 2022 (link).

[8] Generally, the economics of industrial organization, regulatory economics, social welfare theory, and institutional design are relevant subfields that focus on how firms, labor contracts, markets, and other forms of economic activity are organized and coordinated to meet efficiency and distributional goals. The tools of game theory and mechanism design have also proved useful here.

game (Maskin 2007; Laffont & Martimort 2002). The institutional and organizational economics literature explores the various ways in which institutions, organizations, and markets may be governed to address this challenge. One powerful framework for exploring this and making the linkage to the law and economics literature is via recourse to the literature on property rights, contracts,[9] and the problem of incomplete contracts (markets) that inherently always exists in the face of systems evolving in the face of uncertainty and changeable goals, objectives, and shared perspectives on the benefits of coordination.

For example, Becker (1968) brought the tools of mathematical economics applied to private versus social decision-making to understand the role of penalties in economic systems, which arises because of externalities and the effort to manage those externalities. If markets (contracts) were complete then externalities could be brought within the neoclassical realm of efficient competitive equilibria, but that is inherently infeasible. Much of the literature on crime and punishment, and penalty regimes, focuses on the role of society and the governance of individual behavior (citizens). Although the structure of blockchain-coordinated governance, including DAOs, has relevance to questions of sovereign governance, much of the actual joint productive output and governance decisions that ensue therefrom are focused on governing the actions of smaller, less formal, and evolving coalitions of actors wherein voluntary participation is presumed. That is, the actors can freely exit in ways that those subject to criminal penalties cannot. Therefore, the governance potential for blockchains or DAOs may be equally informed via reference to the theory of the firm or multilateral contracting among private individuals (Alston 2019; Davidson 2024) as it is the theory of nation states and their sovereign-constraining constitutions (Rajagopalan 2018, Cowen 2019, Alston 2019).[10]

Therefore, we analogize the penalty policies of blockchains to the penalty policies used in private firm governance for managing the behavior of employees by structuring the penalty (and reward) systems to induce behavior that maximizes enterprise value (rather than private employee value) (Gillan and Nguyen 2016). In corporate governance, compensation regimes are designed with incentive schemes that offer a variety of rewards (Bradley 2007), but the range of penalties are more limited, with the most significant being termination of employment.[11] While efficiency in competitive production contexts would dictate that employees get compensated equivalent to the marginal productivity of their effort, effort's unobservability creates agency costs, which are partly tamed through ex-post discretionary rewards and termination (Gibbons and Roberts 2013).[12]

---

[9] For a discussion of how smart contracts relate to the economics literature on contracts, including property rights, see Lehr (2021).

[10] Applying the managerial incentive perspective to analyze the reward/penalty policies of blockchains rather than the criminal/civil perspective of sovereign states is preferable here not just because participation is voluntary (in blockchains and firms in ways that do not apply to nation states), but also because it better corresponds with the unit of analysis for our empirical inquiry. Just as the managerial policies of firms are constrained and shaped by the nation states and economies in which they operate, so too are the policies of individual blockchains constrained and shaped by the larger cryptocurrency economy and states in which those blockchains operate.

[11] It might be tempting to view a penalty as just a negative reward, and in some contexts, that may be appropriate. Here, however, we focus on penalties that involve imposing a loss on an agent in terms of something the agent already has a property right of ownership in rather than a reduction in a potential reward that might have been delivered for superior performance. In the corporate governance context, a lower bonus would not be regarded as a penalty, but having one's pay docked, being demoted, or losing one's job would be regarded as penalties. In staked blockchain network contexts, penalization involves the loss of staked tokens and/or the right to continue to participate in validator functions, penalties akin to being demoted or fired from one's job.

[12] A common solution in executive compensation is to forfeit certain rewards if the executive is terminated, which displays the role of both rewards and penalties in private employment contexts (Gillan and Nguyen 2016). Such

Another reason for incentive limitations of penalties in contrast to rewards is that penalties are typically constrained by lower bounds that may not be applicable in case of rewards. For example, a stock price cannot be negative and legal limits may bound what penalties may be inflicted, whereas rewards are only bounded by the joint productive capacity of the firm. That has implications for the estimation of expected penalty/reward tradeoffs. For example, if there is small chance of an extremely large ("near infinite") reward that can provide incentive impact that is not symmetrically feasible with penalties since those are bounded.[13] This goes to the notion that penalties are just negative rewards, which suggests that both positive and negative incentives may result in more complete contractual environments. If a penalty involves the termination of a human operator's validator role, then that penalty is bounded and its implications depend on whether termination is bundled with other actions such as loss-of-stake or rewards that may accompany termination of role as validator. In contrast, businesses may refer employees for criminal or civil prosecution of claims to augment the penalties that may be directly applied by the corporation. Understanding the role of penalties and the diversity of mechanisms employed by different enterprises provides a framework to analyze the role of automated penalties in creating a stable foundation for the development and execution of smart contracts.

What tends to predominate in private employment contexts is that of positive rewards, as the proverbial carrot to the stick of termination. Bonuses, raises (including with promotion), and other positive benefits are much more common (Gibbons 1998) than affirmative financial penalties for failure to perform a role satisfactorily in a private organizational context. Importantly, these carrots and sticks can and typically do rely heavily on discretionary monitoring on the part of the private organization – if routes to exceptional rewards can be well-defined ex-ante, then these rewards can simply be priced into the contract as expectations on the part of individual employees' job duties.[14]

In contrast, blockchain networks lack centralized discretionary authority by design (Nakamoto 2010; Buterin 2014). While this disintermediation has numerous benefits that are discussed elsewhere in the extant literature (Catalini & Gans 2020; Narayanan et al. 2016), this also creates an inability to provide rewards in real-time based upon discretionary judgment. Therefore, just as validator rewards are baked into the protocol as a function of the successful addition of proposed blocks, automated penalties need to be defined ex-ante as part of the rules for the staked blockchain network and with more explicit definition than is necessary in corporate governance policies in order to achieve their desired goals.

However, as we explain further below, discretion is often (and some would argue, always) desirable in the face of decision-making in the presence of uncertainty.[15] Therefore, a key dimension of our inquiry involves our efforts to assess the extent to which penalty institutions by

termination does not, however, eliminate the ex-employee's ability to continue to participate in the economic future of the enterprise (as a shareholder) or as the employee of a competitor (except as may be constrained by non-compete clauses).

[13] That is, it is not feasible to reduce a validator's stake below zero.

[14] Explicit rewards that are part of the employment contract are common for job tasks with clearly defined and easy to monitor performance based on measurable outputs (e.g., bonuses for salespeople, rates for piece-work, or overtime pay). However, for higher-level management employees or other jobs where precise monitoring of task performance is less feasible, discretionary rewards may be more common (e.g., size of the bonus) and complement ex ante-specified rewards (e.g., options tied to stock price performance).

[15] With Knightian uncertainty (Knight 1921; Ellsberg 1961), it is impossible to anticipate ex ante all possible future states. Prescriptive ex ante rules that eliminate discretion may result in worse decisions than if some level of discretion is allowed. However, imperfect ex post observability and enforcement may make it impossible to differentiate between failure to enforce ex ante rules or the use (or abuse) of appropriate discretion.

blockchains may afford scope for discretion and whether such allowances are welfare enhancing or not. Although we are inclined to believe that discretion in the design and application of penalties in blockchains is likely to emerge at a greater scale, even if less so than in traditional corporate governance contexts, we are not clear yet how that might best be detected or manifested. It is conceivable that poorly designed or implemented penalty institutions may give the appearance of allowing discretion when the reality is they merely reflect the inherent incompleteness of governance.

While protocol updates are a testament to the incompleteness of blockchain secondary rules (Alston et al 2022), and bugs and hacks indicate the incompleteness of the primary rules of these same networks, penalties can be understood as making these commitments more complete than otherwise. By analogy, removing a validator from the set of nodes eligible to validate transactions is akin to termination or demotion, but that does not capture the full dimension of the automated penalties and corporate governance comparison. We apply the corporate governance analogy to the design of our empirical strategy for analyzing automated penalties. We look at what those penalties are (in terms of how they affect the rewards, participation incentives and capabilities of validators) and what other off-chain elements/factors appear to be likely to characterize the penalty institutions of different blockchain networks. Most centrally, two features of staked blockchain networks' governance make semi to fully automated penalties essential: (i) the lack of centralized discretionary control of these networks; and (ii) the inability to disambiguate between faulty validator performance and malicious intent.[16]

2.2. The Organizational Economic Logic of Penalties on Staked Networks

A key feature characterizing blockchain networks is their consensus algorithm. The first blockchain network, Bitcoin (Nakamoto 2008), and many that were created following its introduction, rely on Proof of Work (PoW). Our focus here is on Proof of Stake (PoS) networks (Kiayias et al. 2017) that use a consensus algorithm whose incentive alignment derives from the fact that validators must stake network tokens to perform validation functions. Understanding the way in which the joint productive capacity of blockchain networks varies as a function of their choice of consensus algorithm helps explain why penalty institutions are emergent on staked blockchain networks.

2.2.1. Incentive Alignment on proof-of-work networks

The Bitcoin network's lack of penalties makes the role of such penalties on staked networks clearer by comparison. In Bitcoin's context, costly effort is expended prior to being able to obtain a reward, with the incentive alignment occurring through the cost of miners successfully competing to solve a cryptographic hash puzzle (Nakamoto 2008), a cost which increases in the value of Bitcoin, due to how increased value incentivizes additional independent miners to compete for the scarce rewards.[17] In the PoW case, the costs miners face in terms of incentive

[16] As we explain further below, although these problems also exist in PoW networks like Bitcoin, separating them proves less important and the nature of these networks' joint productive purpose and its simplicity render the need for nuanced penalty institutions irrelevant. That proves not to be the case for staked blockchain networks.

[17] Moreover, assigning the reward is simple because the miner that first solves the PoW puzzle receives the reward and in so doing helps ensure that the Bitcoin blockchain's performance is sustained. It is easy to evaluate whether the solution is valid and the value of the bitcoins are like gold, and easily assessed and fungible commodity asset. In contrast, the output and value of a smart contract depends critically on what the smart contract is about, and that value may depend on attributes of the output that are difficult to assess and monitor without context-specific information.

alignment are also fully borne prior to the validator receiving the (potential) reward for the costly effort expended. While this works in the context of blockchain networks coordinating in furtherance of easily verifiable tasks like the reliable processing of transaction requests drawing upon ledger balances in a fully digitized context, there are several limitations to this model of incentive alignment.

First, it is well understood that use of PoW for incentive alignment is very processor intensive (Agur et al 2022), to the point where single transactions net out to electricity consumption comparable to monthly household bills.[18] Although this electricity consumption creates a barrier that prevents bad faith miners from spamming the network with proposed fraudulent blocks, these costs precedent to block proposal also operate as barriers to open participation and incentive alignment by forcing miners to perform the necessary "work" before even qualifying to propose a block to be added to the blockchain. Moreover, miners that fail to win the hash function competition risk not being able to recover their expenditure in competing for rewards. Nevertheless, the PoW mechanism when coupled with the independent ability of other validators (and soft nodes) to accept or reject proposed blocks, results in a highly tamper resistant network, albeit at a high cost to each participant. A further criticism is that the resulting level of tamper resistance is potentially inefficient (relative to available design alternatives) since it comes at the expense of excessive resources being devoted to protecting the consensus mechanism from malicious attacks (i.e., ensuring network resilience) and too little on securing better performance (i.e., directing resources to improving the actual computation of conformity of proposed transactions with the existing ledger balances on the network, or forgoing the capacity to compute much more complex transactions than simply adjusting ledger balances among network addresses). To date, distributed network processes have largely been executing over network objects of trivial computational load, but as trusted execution environments and AI agents begin to be deployed, the computation required of validator nodes is likely to increase.[19]

There is thus a deeper concern associated with such a lean governance model. As the complexity of coordinative purposes of a given network increases, the viability of such an already electricity-intensive model decreases. That is because the incentive alignment of this system comes from front-loading the energy expenditure prior to the production of real value. In the case of Bitcoin, this joint production entails ensuring that proposed transactions conform to the last block's ledger balances, but if the network requires significant computation at this stage, then this requires higher rewards to compensate for calculations that are increasingly non-trivial. Therefore, while a proof-of-work consensus mechanism aligns incentives in a context where production can near costlessly be verified by other independent producers, it arguably cannot extend to more complex coordinative purposes. It is precisely to enable automated support for such more complex and diverse economic transaction contexts that provides the explicit motivation for creating smart contracting platforms and the blockchain networks that support them such as Ethereum, Solana, and others.[20]

---

[18] The Bitcoin network consumes more energy annually than many small nations (Cambridge CCAF CBECI; de Vries 2018).

[19] Even if scaling solutions like Layer-2s (L2s) on the Ethereum network mean the data which posts to the network is nowhere near as large as that generated by the off-chain processes the network is ultimately coordinating, this either operates to require trust in the off-chain scaling solution, or limits the computational intensity of that which can fully be proved "on-chain".

[20] The Ethereum white paper explicitly contemplates a more robust smart contracting environment to facilitate processes ranging from escrow to voting to prediction markets (Buterin 2014), a point which was followed by identification of the role of penalties in aligning independent contributors' incentives (Buterin & Griffith 2017).

The long-standing analogy of mining precious metals helps clarify this distinction – gold (and gold-bearing ore) can immediately be verified in ways that many complex intertemporal productive activities cannot. Thus, the miners of Bitcoin more closely approximate gold miners, whereas the validators on staked blockchain networks require more robust incentive schemes involving explicitly specified penalties alongside rewards. This comparison is deliberately agnostic as to which set of joint productive purposes distinct blockchain networks should be designed to facilitate, but merely serves to emphasize the distinction between "work" that can be nearly instantaneously verified and more complex joint production that cannot.

2.2.2. Incentive Alignment on staked blockchain networks

Proof-of-stake relies on a distinct process to minimize the likelihood that malicious actors will be able to propose fraudulent blocks (Ethereum.org 2025, PoS attack and defense).[21] The validator role of the "miners" on PoW networks is instead performed by "stakers" on staked blockchain networks. Staked validator nodes are responsible for proposing and updating the blocks that comprise the distributed ledger that the PoS blockchain network maintains. To participate as validators, each node commits stake to qualify to perform validation functions. In this consensus context, each staked validator node has a probability proportional to their stake that they will be chosen to propose the next block. However, this proposal function still entails a role for other validator nodes, a subset of which are randomly chosen to attest to the validity of each proposed block.

To take as a clarifying example, we focus on the operation of automated penalties on the Ethereum network since this is the largest and best-known staked blockchain network. Relatedly, Ethereum's decision to switch from PoW to PoS consensus also involved Ethereum's adoption of automated penalties which significantly raised the salience of such institutions within blockchain communities.[22] With Ethereum, a fraction ($1/32^{nd}$) of eligible validators are pseudo-randomly chosen from the universe of Ethereum validators to attest to a slot of transactions within each proposed block. Thus, the performance required for reliable block production involves two margins on which penalization can be assessed – block proposal and block attestation. Notably, block proposal is singular to a given validator node, whereas block attestation requires a large

[21] In this paper, we have focused on discussing the penalty policies observed in staked networks. In much of the academic and trade literature discussing staked networks those networks are referred to as Proof of Stake (PoS) networks. This makes some sense to exploit the nomenclature and terminology for characterizing alternative consensus mechanisms to PoW. The reason we resist the PoS terminology is because it obscures the essential role of the blockchain's consensus mechanism in governance. In the PoW mechanism, the incentive benefits of the proof are dependent on and consistent with the solution to the cryptographic puzzle that requires significant computational work. The solution is indeed proof that the work has been undertaken. By way of contrast, the incentive alignment that staked networks seek to provide is more complex. It depends on the action of validators to pledge stake to the blockchain conditional on the reward and penalty policies that are implemented and secured by the blockchain. The pledge of stake is not proof of anything other than a pledge may have been made. Without the reward and penalty policies that connect that pledge of that validator operator to the network, the desirable properties of incentive alignment that the consensus mechanism intends to ensure are not convergently likely to obtain.

[22] Ethereum switched its main chain to PoS in September 2022. Automated penalties were adopted in 2020 when it first launched its PoS chain in a preliminary capacity, with the chain reaching consensus without transactions proposed for addition. The code development that this transition required was a complex and massive undertaking, such that the public discussion of design elements was extensive. One early reference to the role of penalties was for the finality gadget for the Casper protocol update, with Casper FFG formalizing slashable faults as conditions that irreversibly destroy stake upon proof of misbehavior, anchoring economic finality (Buterin & Griffith 2017).

number of validators, which means penalties do not carry equal likelihood of assessment due to the comparative magnitude of attestation activity across validators relative to block proposal.[23]

The act of block proposal and attestation gives rise to three triggers that may give rise to validators being penalized: (i) proposing two blocks simultaneously; (ii) signing two attestations for the same span of proposed transactions; or (iii) signing an attestation that overlaps with a previously signed attestation. Proposed blocks are added to an available chain where they await finalization through other validators' attestation as to their validity – consensus is thus split between proposing a block to the "LMD-GHOST" chain, and finalizing proposed blocks through a separate attestation process that sits on top of the unfinalized sequence of blocks. This should make apparent that a minimum proportion of validator nodes (as a function of pledged stake) need to be online for the Ethereum network to finalize blocks. As we subsequently discuss, the exact conditions that trigger a need for a penalty vary significantly across staked blockchain networks (Cosmos SDK Docs 2024–2025; Polkadot Docs 2025; Tezos Docs 2025).

This makes the final class of penalties that can be assessed on this network clearer – downtime also results in penalization, with a slowly increasing penalty for being offline assessed for validators due to the (minor) diminution in network security that a single validator being offline entails.[24] Of course, on such a salient network with tokens that command significant exchange value, the number of validators consistently exceeds the minimum proportion necessary to attest to a particular slot of transactions. Nonetheless, the mechanism design is meant to drive the probability close to zero that the network will not finalize transactions or be subject to a malicious attack. These goals are secured through the incentive alignment benefits that come with being a validator as well as large number of validators participating in the Ethereum network. The large number of validators reduces the probability that an attacker will be selected to propose bad blocks in the first instance, as well as the increased cost in stake that would be at risk for an attacker that seeks to increase their probability of gaining the opportunity to propose a bad block. The large number of validators also reduces the probability of being selected as one of the proportion of attestation validators for any given slot, and even if selected, the need to be able to capture the behavior of a sufficient number of the other attestation validators to successfully capture the update process.[25]

Another major distinction warrants discussion, though, which is that these penalized activities can carry the additional penalty of being slashed from the network as a validator altogether. While conflicting proposals and attestations are a potential signal of malicious intent, downtime is not something that can be construed as malice, such that conflicting proposals and attestations also carry the penalty of being removed from the validator set on the Ethereum network, along with an incremental penalty that is applied while the slashed validator is in a pool of validators waiting to exit their roles. While validators that voluntarily choose to exit with their pledged stake are placed in an exit pool, slashed validators are involuntarily placed in this pool,

[23] Indeed, the record of penalties on the Ethereum network shows that the vast majority of penalties assessed are for attestation violations, consistent with the reality that a fraction of validators have to attest to each subset of transactions within every block (beaconchain 2024).

[24] In an extreme case of large numbers of validators offline, a quadratic penalty is applied to the offline stake, which suggests that in the limit, the consequences of prolonged downtime among enough validators are quite serious for performance and security (Edgington 2025).

[25] Not only would multiple attestation validators need to be compromised, but the conspiring attestation validators would need to be randomly selected from the pool of all eligible validators to participate in the conspiracy.

with additional penalties beyond the 1 ETH that is assessed for conflicting proposals and attestations.[26]

This shows how the institution of graduated penalties is also emergent on staked blockchain networks, with behavior that could signal malicious intent carrying a much stiffer penalty than simply failing to be reliably online for the purposes of transaction validation.[27] Importantly, the pseudonymous nature of most staked networks makes it impossible to prevent someone from rejoining, which creates a need for additional penalization to make rejoining costlier than the fixed penalty associated with conflicting transactions. As noted previously, though, this introduction of common penalties on staked networks is specific to those on the Ethereum network, even as these penalties vary significantly across different staked networks. This is the case with outright removal from the validator set (called "slashing" on the Ethereum network), as many networks instead render a validator ineligible to participate for a period, perhaps due to the reality that pseudonymity of validator nodes prevents effective permanent exclusion of bad performing validators. On the Ethereum network, the lockup period is intended to prevent an attack with a longer time horizon, in which a subset of validators controlled by a malicious actor tries to fork an earlier block when that validator set obtains the right slot attestation opportunity to do so. Desired consensus mechanism properties thus entail the need for specific penalties to achieve those properties,[28] an example of how a specific joint productive institutional design requires specific remedies to align the incentives of pseudonymous and automated participation in joint production.

Another element of the Ethereum network's penalties is necessary to understand the range of penalties we discuss in our comparative analysis. Given that the network has hundreds of thousands, and at times, approaching a million validators, the likelihood that a single validator with adversarial intent double signing or proposing could actually subvert outcomes is quite low; their ability to fork the network to a fraudulent block or sway attestations thereto is effectively nil. However, given the pseudonymous nature of validators, it is conceivable that one actor could control multiple validators, and mount a coordinated attempt to force a fraudulent transaction. Thus, penalties are assessed in practice as a function of how many other validators also have similar conflicting proposal or attestation behavior within a particular period. This can be seen as a bar to collusion among validators, as well as deterring the potential of one actor coordinating numerous validator nodes, an outcome that would be indistinguishable from collusion due to the pseudonymous nature of the network.

Absent the costly barrier that PoW consensus requires to propose new updates to the shared ledger, this reduces the good faith incentives to propose blocks, which creates a need for an affirmative signal of incentive alignment in which validators must stake a certain amount of network currency to participate in maintaining the network in terms of numbers of validators online and confirming each subsequent proposed block. However, this notion of staking valuable

[26] As we document more comprehensively in Section 3, the variance in penalties includes how the penalty level is calculated, which results in quantifiable variance in these automated penalty institutions in economic terms. For example, Cosmos' x/slashing specifies downtime and double-sign penalties with jailing/tombstoning; Polkadot employs proportional slashing for equivocation; Tezos provides denunciations for double-baking/endorsing with deposit forfeiture (Cosmos SDK Docs 2024–2025; Polkadot Docs 2025; Tezos Docs 2025).

[27] Fork-choice rules shape which validator actions threaten consensus; LMD-GHOST can be linked back to high-throughput analyses like GHOST (Sompolinsky & Zohar 2015), clarifying why equivocation is particularly hazardous and therefore slashable.

[28] On the Ethereum network, this is called "weak subjectivity", in that nodes that are always online can be assured of the objective validity of the ledger state, even as those who experience time offline have to rely on the block header from a trusted source in order to trust the ledger following a sufficient period of downtime.

network assets directly poses the question of when and how a given validator can exit with their pledged stake. This creates a need for a delay following an announced intent to exit during which the validator cannot withdraw their pledged stake.[29] If exit were instantaneous, this could theoretically be an attack vector itself, but also poses another risk associated with the volatility of token prices on cryptocurrency networks writ large – given a large enough price increase, validators might wish to cash out their stake to convert it to stablecoins to lock in exceptional gains, potentially creating fragility at a time of high network activity and demand. The need to enable validator entry and exit while still retaining the incentive penalty function associated with committed stake gave rise to an exit pool. Validators in the exit pool must wait a pre-specified period before exiting with their stake. This is the same pool in which slashed validators are placed, although validators voluntarily exiting are not subject to the aforementioned incremental penalties that those placed into the exit pool for conflicting proposals or attestations are.

The incentive alignment that automated penalties and pseudo-random rewards provide can be thought of in terms of asset-specific contracting (Williamson 2005). In contesting the belief that markets could produce universally standardized contract terms, Williamson argued that different forms of production require contractual tailoring, in some cases reaching to the entire organizational level itself. While Williamson focused on contracting and organizational forms, his seminal insights have more general purchase: certain forms of joint production rely on highly specific coordinative institutions for their realization.[30] The implications of this for cryptocurrency networks is that not only should fixed network parameters like block size vary according to the joint productive (i.e. coordinative) purpose to which the network is being applied, but so also should the entire governance regime itself, including the balance of rewards and penalties in place to productively align incentives among independent contributors.

An additional characteristic of larger staked networks is also revealed through the lens of organizational economics. Hybrid firms involve joint production among otherwise independent producers (Menard 2025), such as sharing trans-Atlantic flight routes among different national airlines, or syndicating risk among independent insurers in high-risk contexts. Certain joint productive outputs require the utilization of distinct firms' productive capacities (or property rights) to be more efficient. In most staked network contexts, the federated and randomized nature of "managerial" authority is likely insufficient to even call the network a singular private firm, such that blockchain networks themselves could be characterized as a unique type of hybrid firm. Equally interestingly, though, the lightly permissioned nature of network processes has led to additional diffusion of specialization of the process of validation of transactions. Networks with a public mempool in which proposed transactions are visible, coupled with the capability to pre-commit using smart contract logic, have developed ancillary market for block ordering. This capacity for validators to order the transactions within each proposed block is not limited to staked blockchain networks but has become significant on some of the largest staked networks, including Ethereum, Solana, and Avalanche. This involves the separation of block ordering from the actual proposal of the block by a staked validator, with finding profitable transaction bundles completed

[29] If the required term for staking is either too long (in the extreme, permanent) or too short, staked tokens will fail to have their desired incentive penalty effects. If too long, then the requirement to stake will be regarded more like a sunk entry fee that is unrelated to the validator's performance once paid since the value associated with its recovery is so low (i.e., contingent cash payments in the distant future have a small present value). If too short, then the risk that a penalty might be assessed and the stake forfeited is too low to deter bad behavior.

[30] A coal mine and a railroad spur fundamentally dependent on one another have sufficiently adversarial incentives that their efficient joint production must be organized within a firm to overcome these otherwise rent-consuming adversarial incentives; the organizations are less competitive, and therefore less profitable, if separately owned.

by "searchers" and maximizing a block's value completed by "builders" that submit these built blocks to validators. Each stage shares in the gains from the demand for block ordering, although the proportional gains to each can vary depending on the network.[31] This has come to be known as maximal extractible value ("MEV"), and is not without its own controversy in terms of the strategic behavior that a market for block ordering can create.[32] Setting aside the controversy, the independence of stages of block production carries a network level benefit associated with mitigating the risk of malicious attack: searchers and builders are motivated by the revenue from successfully proposed blocks, such that their incentives are well-aligned to presenting validators with valid blocks for proposal. The more that validators come to rely on MEV revenue as part of their incentive function, the less likely that any proposed block will contain bad transactions. The independent and randomized contribution to block validation also applies to the construction of blocks precedent to proposal for inclusion in the distributed ledger, additionally enhancing the security of the process.

Intertemporal rewards and penalties thus characterize significantly more complex productive output than individual delivery of gold for a fixed price. Because of this, the incentive alignment that automated rewards (with a convergent expected probability given long enough duration of contribution to network validation) and penalties plausibly extends to a broader range of coordinative purposes than individual producers bringing affirmative proof of output prior to receiving a discretized set of rewards. Our comparative institutional analysis herein is therefore of more general interest than merely assessing the comparative viability of different blockchain governance mechanisms. Instead, a world of increasingly automated governance, and potentially even semi to fully automated autonomous agents acting on individuals' behalf, means that incentive alignment according to automated rewards and penalties have relevance beyond the issuance of cryptocurrencies and the derivative coordinative purposes to which they are applied, a point to which we return in our final section. Moreover, in distributed digital governance contexts, the appropriate blend of automated rewards and penalties may be a precondition to their effective coordinative application. If blockchain networks constrain the exercise of network authority in productive ways, then this "constitutional" (Alston 2019) character is likely to have broader application in other distributed network governance contexts, such that our comparative analysis also applies to a subclass of institutions precedent to more constitutionally constrained digital governance writ large. Put more directly, the design and implementation of penalties on staked blockchain networks is likely to inform automated execution of penalties in other joint productive contexts governed by network protocols.

[31] One 2023 estimate of MEV fee-sharing involved Ethereum validators obtaining over 70 percent, with builders getting roughly 10 and 17 percent respectively (EigenPhi 2023).

[32] Among the opportunities that such a market creates is the ability to get in ahead of large orders to capitalize on the some of the price effect that large transactions have; in traditional capital markets, this is called "frontrunning" and is prohibited by exchange policies in the vast majority of such contexts. Nonetheless, arguments in favor of such a market include lower fees (ceteris paribus) for transactions due to the gains that go to validators, the fact that sophisticated actors can themselves "price" in the possibility for MEV into their behavior, and that technical strategies (like private mempools, shielded RPC endpoints and splitting large transactions) exist for transaction proposers to avoid the incidence of MEV (CoW DAO 2025), although this will typically entail higher fees or longer wait times. In theory, given sufficient network concern over block order manipulation due to its intensification of adversarial incentives (Daian et al 2019), automatic penalties could be created to govern block ordering that has certain undesirable characteristics or ex-post effects.

Pre-specified penalties are thus an important component for the considerable majority of blockchain networks that align contributor incentives through pledged stake requirements. Performance assurances ("service level agreements") are a critical element of all contracting environments for coordinating the intertemporal and distributed coordination of quasi-independent economic agents. Joint producers need to have agreement on expected paths so they know what they can count on, what the relevant risks are, and what happens if deviations from the expected or agreed upon trajectory occurs. This applies to economic organization of institutions and markets generally, and therefore, also to blockchain networks.

Penalties and rewards are key design elements of any such system. Yet not all penalties are created equal. This means Williamsonian tailoring of commitment mechanisms also plausibly applies within staked blockchain networks, in that variation in penalties may themselves reflect context-specific characteristics of these systems.[33] Importantly for the purposes of our analysis, the coordinative purposes of different blockchain networks are not fully defined in protocol choice, such that contractual tailoring likely also reflects characteristics external to a given blockchain network itself. For instance, the economic magnitude of a given network is in part a function of choice among users, which is governed by competitive forces, yet likely also affects incidence of penalties; validators with thousands on the line with respect to any penalty have much higher-powered incentives to invest in redundancy measures to prevent downtime.

Another margin to consider surrounds how one institutional choice also interacts with other institutional choices, such that the need for penalties is arguably greater as withdrawal holding periods on a given staked network decrease. In the limit, if network validators could never withdraw their stake, this stake could be viewed as a payment to access the revenue that becoming a validator affords, but misaligns incentives associated with being subject to penalization. For sufficiently lengthy lockup periods of stake, this makes the need for real-time detection of inadequate validation diminish, as the stake is securely held for a long enough period that violations can be detected and penalized.

Another way to understand the emergence of well-specified penalties for failure to perform the validator role surrounds the professionalization of expected performance in a sufficiently defined joint productive role. A given number of validators is necessary for network security and performance, and this set needs to be online with sufficient reliability to ensure those outcomes. Other errors can replicate behavior consistent with malicious intent, such that it also merits sanction, like signing blocks for the same slot or signing two separate attestations to the same target. The formalization of penalties has directly led to operators of stake pools to devise systems to reduce errors that can be conflated with malicious intent. Due to the natural risk of network issues leading to downtime or double attestations that can be conflated with malicious activity, automated penalties have clearly led to the emergence of practices to reduce the incidence for accidental violations, with organizations specializing in aggregating stake from many individual token holders advertising the fact that they have never been slashed, alongside mechanisms they

[33] In the economics of antitrust, these arguments provide a justification for applying a rule-of-reason to vertical constraints of practices that previously and more generally for horizontal constraints among firms are regarded as per-se violations. The added requirements of coordinating vertical relations among firms to manage transaction costs may give rise to contractual arrangements like retail maintenance agreements, most-favored-nation clauses, and assorted franchisee rules that may be justified as procompetitive when examined under a rule-of-reason standard. See for example, Williamson (1979a,b).

have created to minimize the likelihood of the type of accidental errors that lead to the incidence of automated penalties.[34]

This all means that distributed networks that coordinate joint production among pseudonymous nodes that represent human (and organizational) operators have chosen to automate penalty enforcement to a widespread level, whether measured by economic magnitude (market capitalization) or numerical proliferation (number of staked blockchain networks). The emergence of these penalties can be understood through the economic organizational logic of contractual choice, even as validators' contractual choice can be understood as a form of smart (i.e. automated) contracting with respect to the rules of the network at any given moment. This recognition begets an avenue of analytical inquiry, though, which is to say that observed variation in contractual choice, in this case, automated penalty institutions, helps to more generally understand the range of joint productive purposes to which staked blockchain networks are applicable. It is in furtherance of this objective to which the following empirical component of our comparative institutional analysis is devoted.

3. Empirical Assessment of Current Penalty Practices

While the penalties contemplated by proof-of-stake networks vary considerably, they nonetheless display certain common features. At an overarching descriptive level, we classify automated penalties into two categories – "harmful behavior" and "non-participation." Harmful behavior requires affirmative actions by the target validator, whereas non-participation relates to the failure by a validator to act. The affirmative actions include signing two blocks, signing bad attestations, and omitting transactions. Non-participation is associated with validators being offline for some period. This classification schema is related more to the trigger that initiates a penalty rather than the motivation of the validator, although harmful actions can be perceived as more severe threats to trust in the blockchain's integrity than downtime or a failure to participate. A blockchain ledger that contains bad blocks with invalid transactions (very bad) or missing valid transactions (maybe less bad) can fork the entire blockchain, whereas a validator being offline for a period is far less likely to do more than delay or marginally degrade performance of the blockchain network. Moreover, staked blockchain networks are designed to be robust and resilient to fluctuations in the performance of individual or small numbers of validators. Consequently, as we shall explain, penalties for non-participation increase with the duration of a validator outage, and potentially, with the number of validators that are off-line. That is reasonable since concurrent outages by validators may be suggestive of malicious collusion, and in any case, poses a greater threat to performance.

For users ("consumers") of a given blockchain network's joint productive output, a sufficient possibility for bad recordation of transactions because they are inaccurate or are posted too late may each be separately sufficiently damaging to the user's perception of and trust in the blockchain network's validators to cause the user to look elsewhere for comparable benefits. Moreover, it may matter little to a user whether the failure to perform as expected was the result of malicious or simply careless behavior.[35] It is often the case that harmful behaviors result from

[34] One stake pool, Everstake, advertises the fact that its Ethereum validators have never been penalized as part of its call to pledge stake with their service provider, and Consensys similarly notes how they have never been subject to penalty assessment on the Ethereum network.

[35] Presumably, one might view discussions on social media applications like Discord and Telegram and the trade press to be primary sources of information about the prevalence of performance problems or malicious attacks on blockchains that would factor into users' trust perceptions about blockchains generally and the relative trustworthiness

unintended, non-malicious validator performance (e.g., software glitches in validator nodes or human error), and it is also conceivable that non-participation may be intentional (rather than the result of, say, a power outage that may not be attributable to affirmative intent on the part of the operator of the validator node producing the error).[36] In any case, the motivations of validators is harder to assess and seeking to embed a determination of motivation too explicitly in the trigger conditions for a penalty may be contrary to its effectiveness, as we shall discuss further later.[37]

Identifying the effects of different automated penalty institutions across blockchain networks is exacerbated by multiple confounding factors that make it difficult to even determine what the penalties are for all staked networks, to identify what is common across those policies, and ultimately, how and why they might differ across these staked networks. As we shall describe, there is considerable heterogeneity across the networks we examined in terms of penalties, but our ability to explain the sources of that heterogeneity are limited by the available data. In the following three sub-sections, we first explain our data collection strategy and methods; offer a qualitative assessment of the different types of triggers and penalty mechanisms (Table 1); and some preliminary analyses of the prevalence of different features of note in our sample.

---

of different blockchains. One might suspect that a lower threshold of validation for signals of performance problems would be required than for malicious attacks. The reasoning for this hypothesis is as follows: a successful (here, equate with trustworthy) blockchain is likely to have a larger market capitalization and attract more attackers (due to the greater rewards from a successful attack on a better capitalized network) but have few, ideally zero, evidence of successful attacks. The evidence of successfully repelling attacks is a signal of good security, and therefore an affirmative metric of performance. On the other hand, even rumors of poor performance that may be hard to verify may suggest a problem and weaker signals of performance that are more numerous may be expected to have a bigger effect on user trust. Note, by assumption, easily verified evidence of either significant performance failures (e.g., major blockchain outages) or successful malicious attacks (e.g., successful theft or damage inflicted on blockchain users) would both be expected to be extremely damaging to the blockchain network's continued success. Exploring this hypothesis further is of relevance because it might shed light on the impact of numerous, but minor activity of validators warranting penalties, compared to much rarer but more significant penalty events that are more important for promoting user trust. On one hand, observation of penalty events that are verified for a blockchain which also has good performance and is free of successful malicious attacks may demonstrate active and effective enforcement. On the other hand, failure to observe penalty activity may be taken as either evidence of sufficiently superior governance to have effectively neutralized the threats to trust posed by either malicious activity or poor performance; or, as good luck on part of blockchain in dodging bullets and poor information about the blockchain's performance, which may imply the absence of meaningful policies. A goal of our research is to understand the role of automated penalties in governance, but the difficulty in collecting reliable information about the penalties specific to all the predominant staked blockchain networks is a key challenge our work confronts.

[36] Ensuring good performance is costly so unscrupulous validators may seek to scrimp on performance-ensuring expenditures and continue to garner rewards from participating in the blockchain. This is a common threat that gives rise to many of the sorts of vertical agreements used to address principal agent problems (e.g., retail price maintenance agreements and other franchisee terms and conditions that may be used to protect brand image, prevent free-riding, and better coordinate value chain activity), and can also be understood to be mitigated by MEV revenue incentives, because validators that perform less effectively should expect less revenue from MEV-optimized blocks, all else equal.

[37] If penalties were only incurred for trigger actions associated with malicious intent, then that would raise significant concerns of Type I and II errors. Presumably, most real malicious attackers endeavor to obscure their actions and motivations to circumvent ex ante or ex post enforcement actions. The pseudo-randomization of validator identities that is a key feature of most of the staked blockchain networks makes it hard to identify specific validators and create a reliable evidentiary trail that would be needed to adjudicate a determination of intent. That is the substance of criminal and civil adjudication processes that are imbued with significant scope for discretion that is, by design, mostly lacking from blockchain governance.

3.1. Data Collection and Research Methods

To better understand what automated penalty institutions look like in the staked blockchain ecosystem, we focused on a sample of networks that included the 100 largest (by market capitalization) cryptocurrency networks and added a few other notable examples of variation that we identified as having penalty institutions that are not present in the largest staked blockchain networks by market capitalization. We readily acknowledge that in a still newly emerging, rapidly growing, and evolving ecosystem like that characterizing the world of staked blockchain networks and their governance practices that today's market capitalization leaders may not provide the best indicators of future outcomes, or in this case, penalty institutions that come to predominate. Nevertheless, we expect path dependencies to play an important role and so looking at market leaders identified as we have is a defensible approach as a function of the revealed preference that use at actual and comparative costs is larger on networks with larger market capitalization.

An additional motivation for this approach was to maximize the likely sources of publicly available information on the penalty institutions of staked blockchain networks. The largest blockchain networks attract the most online attention and are more likely to have the community resources to provide documentation of their policies and rules. That information includes developer produced white papers that are a critical source of information on penalty institutions, but there is no rigorous standard practice or format for what those white papers include. Consequently, and not surprisingly, those white papers and the level of online documentation of their rules and practices vary significantly in quality and the level of detail. The blockchain developer-provided information was supplemented by scouring publicly available information in the thriving and expanding digital media that community participants rely on to track information about the networks' joint productive output (and to communicate with one another more generally). We also scoured the general media and tech industry press that provides analyses to inform interested parties outside of the active blockchain community.[38]

In addition to scouring the usual publicly available sources, tracking down relevant details also necessitated active engagement with developers associated with particular staked blockchain networks and with relevant knowledge about the codebase and penalty institutions of the blockchain. That active engagement was pursued, in part, via direct communications over social media channels like Discord and Telegram by the co-authors of this study due to the extent to which many communities that coordinate their activities on blockchain networks rely on these channels for public communication either primarily or exclusively.

To build the automated penalty institutions database, we began by identifying all the staked blockchain networks from the top 100 layer-1 blockchains by market capitalization. Information regarding these networks was collected through CoinMarketCap's categories feature,[39] which puts coins into certain categories based on the defining characteristics of their protocol, including their consensus mechanism. For our purposes, we are only focused on coins that are Layer 1 networks, as opposed to those that are subsidiary to many of the major networks. From there, we tabulated all networks that utilize proof of stake as their consensus mechanism.

To collect the data for the penalty characteristics for these networks, we began by using https://staking-explorer.com/, which details the penalty parameters for all Cosmos-based chains.

[38] Ideally, we would have liked to have found academic or other authoritative analyses of the current status of penalty institutions across staked blockchain networks. Unfortunately, we did not find any such studies which provided a key motivation for our current effort.

[39] Coinmarketcap.com

This is helpful because those blockchains share an originating codebase promulgated by the Cosmos Software Development Kit (SDKs), a tool which allows developers to spin up new applications from an originating codebase.[40] This greatly facilitated our ability to identify the common features relating to penalty institutions on staked networks that derive from Cosmos' codebase. For the networks that were not based on the Cosmos template, identifying the penalty policies was more challenging. We had to manually review each network's developer documentation, and in some cases had to query developers on Discord or Telegram to establish the exact nature of the penalties active on their network. The list of blockchains included in our analysis are in Appendix I. The detailed classification data is still under development, but will be shared once this work is finalized for publication.

3.2. Qualitative Taxonomy of Penalty Policies

As already noted, we partition the triggers that can initiate penalty application into harmful behaviors and non-participation. These are summarized in Table 1, which is discussed further here. This discussion is further sub-divided into the three sections of the table into Penalty Conditions (triggering events), Penalty Types, and Penalty Applications.

Category	Description of Typology
PENALTY CONDITIONS	Validator behaviors that result in the application of penalties
- Malicious Signal	On-chain validation actions with possible malicious intent (sending or validating conflicting information)
-- Double Signing	When a validator attempts to sign two blocks/transactions simultaneously for the same height/position
-- Double Attestation	Validator equivocation as to the validity of a proposed set of transactions: "double attestation" or "double voting"
-- Transaction Omission	Omission of transactions in a proposed block
- Non-Participation	Failure to participate in validating transactions on the blockchain due to "downtime" or "missed blocks"
PENALTY TYPES	Possible penalties predicated on the preceding condition(s) being met
- Economic Penalties	Penalty affecting a validator's stake and/or rewards being received for validating transaction on a network
-- Percentage Penalty	Validator's stake is reduced by a percentage, a penalty which tends to be uniformly present for all infractions in cases where applied
-- Fixed Penalty	Validator's stake is reduced by a fixed amount, regardless of the amount that they have staked to the network
-- Forgone Rewards	Loss of potentially earned rewards due to temporary or permanent removal
- Enhanced Penalties	Additional penalties applied as a function of other harmful behaviors in a contemporaneous period as one validator's infraction

[40] SDKs are a common tool for software-based businesses seeking to encourage the growth and activity of third-party developers and providers of complementary applications and services. Providers of SDKs include documentation to explain the architecture and its key functionality and guidance for the use and integration of software modules so that the SDK tools can be used to seamlessly integrate third-party software and allow it to interact with other software components in the ecosystem.

- Validator Removal	A validator is involuntarily removed from the validator set
-- Permanent Removal	Validator is involuntarily removed from the validator set permanently - often seen with conflicting actions (i.e. "Tombstoning" or "Slashing")
-- Temporary Removal	Validator is involuntarily removed from the validator set temporarily - often seen with downtime infractions (i.e. "Jailing")
- Forgone Governance	Ability of validator to participate in protocol update decisions is absent or reduced in cases of temporary or permanent removal
PENALTY APPLICATION	The process by which behaviors are assessed and penalties are applied
- Automatic	Behavior is automatically identified and penalty is automatically applied
- Conditional	Additional penalties applied conditional on the automatically assessed contemporaneous behavior of other validators
- Discretionary	Assessment of infractions can be human-initiated, or penalty levels can be defined by a human authority

Table 1: Penalty Conditions, Types, and Application Methods on Staked Blockchain Networks

3.2.1. Penalty Conditions

For the Ethereum network, the largest staked blockchain network and one with extensive documentary evidence of its policies, harmful behaviors are associated with validators sending conflicting information at the block proposal or attestation stage as detailed in our earlier discussion of these penalties as an illustrative example. Nonetheless, the class of validator actions that can entail penalties are generalizable from this specific (and prominent) example of a staked blockchain network. These observable validator actions entail double signing, double attestation, and transaction omission. Double signing violations are cases where a validator attempts to sign two blocks or transactions simultaneously for the same block height or position. Double attestations occur when independent validators equivocate as to the validity of a proposed set of transactions, either through signing two different attestations in a single attestation period, or through signing attestations whose source and target votes overlap. Finally, omission of transactions within otherwise valid blocks can also trigger penalties. We distinguish all these categories by the fact that they involve an affirmative action associated with a validator's pre-defined network role that triggers a penalty. Although penalties are not triggered by the motivation of the validator, it is easier to associate malicious intent with affirmative actions, rather than passive.[41]

Although less problematic because being offline is not an affirmative behavior consistent with malicious intent and outages by a single validator are unlikely to cause harm to the blockchain, too much downtime can reduce network performance. Additionally, if regular outages are tolerated that can pose a threat to network security, this could operate to reduce the likelihood that enough good faith validators will be online in a given period. That may increase the probability

[41] An agent intending to cause harm does not actually cause harm unless the agent acts or others are aware of those intentions. Whether a governance body may penalize an agent on the basis of intentions or thoughts rather than actual action depends on the context but is generally precluded for reasons of observability and morality. For example, prosecution for murder or theft depends on evidence of the crime having occurred, although admission to a club may be conditioned on other member's expectations of future behavior. As we explained, penalizing validators for the non-action of being offline need not be justified with reference to malicious intent and so may avoid this complication.

of malicious attack and the likelihood that an attack might be successful. The resiliency benefits of validator redundancy is weakened by such outages.

Despite the variance we note in specific penalties, the Ethereum categories of validator (mis)behavior broadly correspond to the most common forms of observable validator activity that trigger penalties in practice across the staked blockchain networks we have reviewed.

3.2.2. Penalty Types

Unsurprisingly, different triggers result in different levels of penalty. We classify penalties into three categories: "Economic penalties," "Validator removal," and "Forgone Governance." "Economic penalties" refer to the penalties that result in validators losing tokens that are staked as part of their role on the network. Such tokens have a monetary value that is often assessable via reference to exchanges, where tokens that are not staked may be exchanged for stablecoins or other cryptocurrencies. The staking requirements of blockchains are key elements of how the governance of different blockchains ensures incentive alignment and good behavior from the validators that are fundamental elements for securing the blockchain's consensus mechanism and performance guarantees.

Economic penalties thus involve the loss of some of the misbehaving validator's staked tokens, which tends to involve either a fixed amount of tokens (as seen on the Ethereum network, for example) or a percentage of stake (as seen on the Polkadot network, for example). It is worth noting that if all validators must pledge an identical amount, a fixed and percentage penalty are equivalent to one another.[42] For staked blockchain networks where the validator staked-tokens subject to automated penalties may vary across validators, a fixed or percentage penalty can have very different monetary implications: a percentage penalty scales with size, whereas a fixed penalty does not. Which is more appropriate may depend on the context and the likelihood of penalties being incurred and their severity.[43]

The economic penalties may also vary with the severity of the threat to the blockchain the penalties seek to address. For example, some networks, like Ethereum, observe other validator behavior during a period contemporaneous to a given validator's triggering behavior to apply enhanced penalties. This is because triggering behavior by multiple validators has a multiplier effect on the risk of increased harms being realized; and may also serve to provide a signal of collusion among validators, which in itself is indicative of malicious intent. If just one validator attests incorrectly, this is much less pernicious for network security than if numerous validators attest incorrectly with respect to the same proposed block.

Beyond penalties that reduce a validator's staked tokens, additional penalties may adversely impact that future stream of tokens that are part of the compensation that validators receive for their role as active validators. Triggering an economic penalty can also result in a validator being removed temporarily from the set of active validators that are eligible for on-going rewards. This class of economic penalty directly indicates the next set of penalties. We classify

[42] On the Ethereum network, to become a validator requires the staking of 32 ETH, so a penalty of 1 ETH or 1/32nd have the same economic impact.

[43] For example, as discussed in an earlier footnote, penalties that impose small, fixed token losses for even slight deviations in performance might be expected to be incurred more often due to outages or even small but more common software or human errors; and such normal deviations from desired performance might be incurred more commonly by larger networks. Imposing a larger per-event penalty based on percentage of stake for a larger validator may be counter-productive and perceived as unfair. On the other hand, if automated penalties are only imposed for events that pose a risk of significant harm, then the added deterrence value of a penalty that increases with the stake of the validator may be needed for it to deter effectively.

these as validator removal penalties that involve removing a validator from the eligible validator set. This can take two forms, with removal either being permanent or temporary. The permanent removal option makes it such that a given validator identity is blocked from ever again performing validator services for the network. This is akin to termination of an employee in a more traditional private firm. As noted previously, this comes with the cost of forgoing rewards that a validator would have otherwise expected, like how an employee loses their ongoing wages or salary upon termination. While the threat of termination serves as a powerful deterrent to employees who might otherwise shirk or misbehave, this permanent removal is less stringent than it might seem in pseudonymous validator contexts, precisely because validators can create a new network identity with their remaining stake (plus some measure of new tokens). This permanent removal is called either "slashing", to evoke being irrevocably cut from the eligible validator set, or "tombstoning", to evoke the effective death of the validator identity's eligibility to validate transactions.

The fact that permanent validator exclusion is not readily feasible[44] means the termination penalty is, in effect, always temporary if the human operator of a validator node wishes to rejoin the network. The duration for being placed in the inactive validator set may be fixed and relatively short, or may be longer or open-ended (e.g., conditional on meeting certain performance hurdles to regain eligibility to be part of the active validator set). Because permanent exclusion and long-term or uncertain duration exclusion may be indistinguishable in practice, this is a challenging feature to evaluate with respect to its impact on penalties. Temporary relegation to the inactive validator set is often called "jailing." This terminology indicates that the validator will be released from its temporary purgatory and allowed to rejoin the validator set after a set period, or sometimes after certain affirmative actions or performance criteria are met. At a minimum, in cases where validators have had a portion of their required stake removed via an automatic penalty, the operator of such a validator node will need to manually rejoin the eligible validator set, which will require the addition of tokens to reach the minimum stake level in most instances.

The final class of penalty – foregone governance penalties – are analogous to the forgone rewards associated with validator removal. On many networks, being a validator also entails choosing whether to accept proposed protocol updates. This governance function (Alston 2019; Alston et al 2022) directly involves validators having a perspective on whether a particular update is desirable or not, yet if a validator is offline during a particular protocol update period, then they cannot influence the outcome one way or another. While many updates are broadly supported, the ones over which there is significant disagreement are those for which each validator's vote may be more likely to be salient and impactful. Thus, foregone governance penalties have additional significance beyond just the stream of rewards that may accrue to normally (good) performing validators. Because protocol updates can fundamentally influence the trajectory of the blockchain, posing an increased risk of harm.

[44] In theory, one could trace the unique set of tokens that the validator pledged to a new wallet attempting to become a new validator, but the existence of mixers and tumblers and exchange on-ramps and off-ramps would stymy this as an effective deterrent. While there are means of tracing coins to and from exchanges, and even through tumblers and mixers themselves, these require considerable technical skill and effort, such that automating doing so is implausible. This is especially the case given that these networks are designed to minimize the need for discretionary enforcement of the kind that would be required to track validator operator identity across wallets. What is nearly certain is that an automated protocol could not be encoded to do so reliably to where it would have its intended deterrent effect.

3.2.3. Penalty Application

Having considered the behavior that triggers a given penalty, and the specific nature of the penalties themselves, the final institutional condition of penalties that merits assessment relates to the process and procedures by which penalties are implemented.

Our analysis of automated penalty institutions across staked blockchain networks finds the least documented variation with respect to the application of penalties. This is not surprising since the available data and our analysis focuses first and foremost on the rules-based application of penalties. We identify the trigger and different penalties different staked blockchain networks impose on account of those triggers based on the available documentation of these networks' protocols and practices. Most of the focus on blockchain governance has been the ability and role of blockchain networks to automate collective decisions and action and by doing so to eliminate the need for discretionary control by intermediaries. Blockchain networks structure the role and operation of validators to protect against the risk of malicious capture of validators or performance failures by individual or groups of validators from resulting in unacceptable behavior of the blockchain. Ex post, deviations from expected or guaranteed performance may be indistinguishable from the abuse of unauthorized discretion. It may not matter if the apparent abuse of discretion was the result of software or human error, for liability attaches to the identity legible to the network and thus subject to the application of penalties, the validator node.

A key way in which such protections and decentralization of discretionary decision-making power is achieved is by adopting procedures so that validators are randomly chosen to validate and attest, and the pool from which validators assemble a proposed block is also determined exogenously to the validators themselves.[45] Thus, the predominant way in which penalties are applied is automatic and follows directly upon a proscribed trigger being observed. The events that will be observed to trigger a penalty and the resulting penalty are pre-specified. This has the advantage of reducing ex ante expectations of discretionary and potentially unfair and arbitrary enforcement of penalties. The ex-ante behavior of both potential malicious and well-intentioned good validators should be influenced by the clear specification of penalty triggers and their effects, and verification that those rules are applied automatically as specified in the network protocol. Reducing the uncertainty associated with discretionary decision-making processes facilitates efforts for efficient contracting among validators that is critical to building trust among stakeholder communities participating in validating (and thus securing) the staked blockchain network. This can be understood as a specific example of the way in which sufficiently complete rules specified ex ante and reliably enforced (through automation) can coordinate human node operators who have never met one another, do not speak the same language, and do not necessarily share values or preferences beyond those sufficient to motivate node operation as a function of the rewards and penalties that node operation entails.

Although the goal of automated governance is to reduce discretion and uncertainty, that goal can only imperfectly be realized, and as with other elements of blockchain governance, some role for discretionary decision-making is both unavoidable, and ultimately, we believe desirable. This is also true in the case of seemingly automated penalties. For example, even when the penalty is applied automatically, there is often a delay and process that separates the identification of the trigger and the imposition of the penalty. It is often the case that validators may be placed in an

[45] On larger staked networks, a market for block ordering has emerged, which makes block ordering (out of the set of available transactions in the mempool) at least partly independent of validators as well, for searchers and builders order blocks to maximize revenue.

exit pool once a trigger has been confirmed as part of the automated penalty process.[46] This is at a minimum necessary to allow consideration of whether other contemporaneous violations warrant the assessment of additional penalties. These process delays and complexities may introduce a degree of apparent discretion (uncertainty with respect to the predicted outcome) to what is otherwise regarded as predictable, automated process.

The discretionary penalty application proceeds through two distinct means among the networks we considered. The first surrounds the ability for validators to effectively accuse other validators of misbehavior, in which case the misbehavior is considered via human judgment in a way akin to how validator operators consider proposed updates to the network protocol. In one notable case with which the authors are familiar, the conditions for penalization are fully specified ex-ante in the network's constitution, yet the network in question (Q) reserves a human layer of discretion in terms of proposing a slashing, and the exact magnitude of penalty being applied.[47] This occurs through a class of soft nodes ("root nodes") that have applied (often by revealing their human identity and judgment expertise) for the role then choosing what penalty to apply in response to a slashing proposal initiated by one of the root nodes. The existence of discretionary mechanisms is consistent with the large role that discretion has played historically in determining rewards (bonuses) and penalties (forgone rewards, termination, and criminal sanctions). Given how uncommon this feature is among the networks we survey, it remains to be seen whether this will be adopted sufficiently to facilitate comparison across network outcomes where it is and is not utilized among the set of otherwise fully automated penalty (and reward) regimes.

3.3. Analysis of Common Penalty Features

In this section, we present our analysis of common penalty features and the variability we observe in the available data. In Table 2, we provide counts of the number of blockchains in our sample that impose penalties denoted in staked-token-percentages with different levels of severity for two of the triggering events, harmful and non-participation triggers. The data shows a wide dispersion in the magnitude of the penalties, ranging from less than 2% to 100% for harmful triggers, but mostly significantly below 1% for outage triggers.[48] As noted earlier, translating this diversity into its implications on the behavior of validators and the economic market relevance of their activity is complex. A sufficient reason is that the penalties may differ in economic effect depending on whether the validator stakes vary and whether penalties are fixed numbers or percentages of staked tokens.

% of Stake						# Networks
Double Sign Penalty	0-2%	2-4%	4-6%	6-8%	8-100%	
# Networks	6	2	63	0	4	75
Downtime Penalty	0-0.25%	0.25-0.5%	0.5-0.75%	0.75-1%	>1%	
# Networks	65	5	5	3	1	79

[46] The action (or non-action) by the validator that justifies the penalty typically precedes detection of the triggering event and there may also be further delays to allow for detection and penalty process initiation. The likelihood of detection and the lag between detection and imposition of the penalty may all impact the strategic calculus of validators (i.e., for those intending to engage in triggering actions and those who are concerned about accidentally being subject to penalties) and therefore impact validator behavior.

[47] More information on the application of this hybrid model of automated penalty triggers coupled with discretionary penalty levels can be found at hq.q.org and stats.q.org.

[48] Note that the number of networks included in the table varies due to not all networks having penalties for each class.

Table 2 – Distribution of Percentage Stake Penalties for Double-signing and Downtime

Table 3 reproduces the classification of Table 2 by translating the penalties into dollars based on total market capitalization of the tokens included in the relevant staked blockchain networks.[49] To further understand the penalties on these networks, it is important to consider the economic value of either a double sign or downtime penalty that validators bear on their networks. The economic value of these penalties can be calculated simply by taking the price of the native token of the network and multiplying it by the respective penalty. For fixed penalties, we simply multiply the coin's price by the fixed penalty amount. For percentage penalties, we multiplied the minimum stake amount to be a validator by the percentage penalty. Then we multiplied the product by the price of the native token for the network which yields the economic value of the penalty. The proof of stake networks included were those in the top market cap and had either a clearly outlined double sign or downtime penalty. After the economic magnitude of these penalties was calculated, we compared these values across these by creating quintiles which were calculated through a standard percentile function. All the individual economic values described in the previous paragraph were used in the percentile function to yield the results in the table.

Penalty Magnitude Quintile	0.2	0.4	0.6	0.8	Max
Double sign	$1.04	$11.36	$52.25	$5,428.80	$375,600.00
Downtime	-	$0.01	$0.08	$73.65	$6,180.00

Table 3 – Economic Magnitude (US$) of Penalties by Network Quintiles

While both are revelatory in terms of the institutional variation between staked blockchain networks, it is important to note that cryptocurrencies vary significantly in value so the same penalty percentage may have very different dollar value implications across blockchains, rendering assessment of the incentive implications for validators complex. The business models of validators have not been extensively studied, and many have business interests across multiple blockchains. This is even more the case given the advent of liquid staking derivative services, where those pledging stake can rehypothecate their pledged stake for derivative tokens, which they can sell or exchange for other tokens, including to obtain stake for pledging to validate on other networks. Moreover, considering the pseudonymous-by-design nature of most blockchain networks, it is hard to identify and evaluate potential business relationships among validators on a particular blockchain or across different blockchains within the industry, beyond staking pools that are explicitly dedicated to staking-as-a-service, a professionalization of the role that is consistent with the way in which institutional definition tends to precede professionalization of joint production at impersonal scale in more traditional private firm contexts.

Because the number of networks that utilize fixed penalties as opposed to percentage penalties is small, we are unable to create summary statistics to show how these penalties differ across networks that also have fixed percentage penalties. From the small dataset though, the fixed penalties seem to be similar once we take the fixed penalty and divide it by the minimum stake amount to get a pseudo penalty. The downtime fixed penalties range from 0.03% to upwards of

[49] This comparison is necessarily from a single moment in time; as the market capitalization and volume of these networks varies daily, the comparative magnitude of penalties should be understood to vary according to these and other forces.

0.5%. When compared to percentage penalties we can see they are very similar with 65 networks having a percentage penalty for downtime between 0% and 0.25%.

It is noteworthy that the penalties triggered by double signing are significantly higher than for outages. This is consistent with the view that double signing is consistent with malicious intent and is thus a more important threat to deter than are outages (at least relatively brief outages). Like an employee in a secure area that is not authorized to be there, the evidence is not dispositive of improper motives, but it is sufficiently consistent with such motives that deterring it with punishment may be one way to deter those with actual malicious intent. In contrast, downtime is like not showing up for work – if enough employees do not show up, the ability for those with improper motives to act on those motives is eventually enhanced. Similarly, if enough validators are offline simultaneously, it becomes harder for the network to jointly produce its output, just as enough employees not showing up stymies the productive capacity of a given firm.[50]

One thing to notice is the herding in percentage penalties for both downtime and double sign penalties. Networks seem to herd around a percentage range for their network which signals a level of consensus or at least the focal nature of specified parameter settings in design choice. This herding seems to show that double sign penalties tend to herd in the middle of the range while downtime penalties herd in the bottom of the range. This may be evidence of considerable mimicry in penalty institution design, although the similar joint productive purposes of most staked blockchain networks could also indicate some tailoring to this shared purpose as the causal driver. We explore these potential channels of institutional variation in greater detail in Section IV.

The existing slashboards are insufficient to test whether economic magnitude has a causal relationship with lower frequency of penalty assessment, although this stands as an obvious avenue for future research. At a minimum, the table displays the extent to which economic magnitude of penalties varies across staked blockchain networks. Obviously, the comparative purchasing power associated with the different penalties in Table 3 has implications for the strength of the incentives that penalties may impose for the behavior of validators and induce them to ensure good performance. Blockchain networks that opt for low-powered penalties may be more attractive to validators that are less able or willing to commit to providing high-levels of performance or are less concerned about the potential harms that malicious attackers may inflict. Validators that make substantial, long-term and potentially partially sunk investments in a blockchain network may be expected to have larger private incentives to ensure the network has sound and robust penalty policies. On the other hand, validators that are not vested in the long-term value proposition of a blockchain may care much less about the threat to future value posed either by poor future performance or malicious attacks. It is also possible that heavier penalty policies that set more rigorous performance standards and protections against attacks may also impose higher entry costs that are beyond the resources of many would-be validators. Obtaining the 32 ETH stake needed to become a validator on the Ethereum blockchain is vastly more expensive and presents a larger entry barrier than becoming a validator on many newer and smaller staked blockchain networks with much lower token values and staking requirements.

[50] On the networks with the largest market capitalization, enough validators are online such that an inability to produce blocks as a function of enough validators being down is a vanishingly small possibility. Nonetheless, making the analogy more comprehensive helps ground in organizational economic logic why downtime is penalized less than affirmative proposal and attestation violations, even as it is still a category of validator behavior that is frequently subject to a non-trivial level of penalty. Stated more generally, penalties play an important part in creating the conditions under which a sufficiently low probability of a bad outcome is obtained in a given joint productive context.

A context-dependent rationale for different penalty proposals may relate to the relevance of the importance of ensuring that all validators have the latest client software. The necessity of ensuring updates are homogeneously deployed across validators will depend both on the distribution of user traffic and the importance of client software changes. If only a small share of a given network's users require the new functionality available only on the most recent client updates, then more lax policies for updating clients may be sufficient. On the other hand, if ensuring uniform adoption of client updates is very important that may slow the pace of protocol updates to ensure better coordinated management of the update process. Such differences could account for either laxer penalties (under the assumption that penalties affect validator software upgrade rates) or ones that tend to lag adoption of otherwise similar blockchains. Alternatively, this could require software compatibility across update versions if the incentives to update validator client nodes' software are sufficiently low powered to where validators only do so infrequently. To the extent that varying client versions presents a risk to security, we would expect more stringent requirements for client updates, which would result in a greater likelihood of downtime for validators not running the latest version, with the corresponding penalties for downtime operating to incentivize client updates.

Despite the clustering that we observe in terms of the distribution of the penalties assessed, it remains the case that roughly one sixth of the networks surveyed are outside the predominant penalty brackets we describe above. Even within those brackets, the networks also vary in terms of the exact level of their penalty in a considerable number of cases.

The final category of institutional variation in this context that is tractable to quantitative comparison is that of the withdrawal timelines that validators face when "slashed" or "tombstoned" from the validator set. As noted previously, withdrawal timelines are intended to prevent a large simultaneous drop in the number of validators, but withdrawal timelines also enable the assessment of additional penalties conditional on the behavior of validators in a specific period. In Table 4, we present the quintile distribution of withdrawal timelines that validators face, although not all of the 54 networks with withdrawal delays assess additional penalties during the specified period.

Withdrawal queue Quintile	Min	0.2	0.4	0.6	0.8	Max
Delay (# of days)	0	2	7	14	21	28

Table 4 –Withdrawal Delay Times (in days) by Network Quintiles

Thus, a common property of most staked blockchain networks is a withdrawal timeline for a validator's stake after a voluntary exit. The withdrawal time can fluctuate depending on the network's state, as it depends on a certain number of completed epochs. Thus, if the network is congested, withdrawal timelines will be extended, which makes the comparison of withdrawal delays more context-specific than the above table could be taken to indicate. Nonetheless, for purposes of comparison, we have computed each under the network conditions observed at the time of this writing.

However, networks do exhibit variance in how these withdrawal queues are determined. For example, the wait time on Ethereum's withdrawal queue is variable as there can only be 16 withdrawals processed per block. If the number of validators joining the exit queue increases

rapidly and in large amounts, this makes the queue time also increase.[51] Another blockchain network named Aptos takes a different approach where the withdrawal time is calculated by how many days are left in the 14-day lockup period. If a node stakes on day one of the lockup period and then attempts to withdraw, that node will have 13 days remaining before staked funds are once more available. Another example of variance in this design is Axie infinity, which allows for instant withdrawal if a given validator has staked for three days. With this variance in mind, it is still the case that most proof-of-stake blockchain networks will involve wait time for validators. A final consideration to note here is the popularity of liquid staking, which involves someone pledging tokens to be staked by a validator service in exchange for derivative tokens (LSTs) which are a claim against the staking service for staked tokens, although these derivative tokens' value can fluctuate relative to that of the staked token over which they are a claim. Thus, for staked validator nodes that want to be certain that they will receive their full staked amount (in cases where they voluntarily join the withdrawal queue), they will commonly face some delay (as seen in Table 4) before being able to fully control their formerly staked tokens. This variance in the waiting times imposed on validator stakes reflects the desire to balance the needs for blockchain network stability and appropriately balancing the private risk imposed on validators by having the liquidity of their stake reduced. Variations across blockchain networks attest to how these conflicting challenges are addressed differently.

In sum, this section considers the margins of automated penalty institutions that are tractable to quantitative comparison. As a minimum baseline for understanding the incentive effects that automated penalties are likely to have on a given staked blockchain network, one should understand the behaviors that trigger a penalty, the economic and governance penalties that result, and how penalties are assessed. These penalties (including their magnitude in terms of network-specific tokens) entail a comparable economic consequence when the network-specific tokens are compared to purchasing power at the time of assessment, a comparison we get by comparing to USD at the time of writing. Finally, we also compare withdrawal delays that each network imposes on validators, both because this forms part of the incentive alignment that undergirds staked blockchain networks generally, but also because withdrawal delays enable the assessment of conditional penalties. In the following section, we consider the potential reasons for the level of institutional variation that we identify.

4. Sources of Variation in Staked Blockchain Network Penalties

The emergent, rapidly growing and rapidly evolving nature of the cryptocurrency industry renders it infeasible to offer reliable predictions as to where the penalty policies will settle as equilibria, or dare we say, some version of "best practice" consensus approaches. To the extent that network effects play a significant role in which staked networks will survive (Katz & Shapiro 1985; Farrell & Saloner 1985), this could indicate that the penalties on the largest networks may be more likely to persist, although all penalties are (theoretically) mutable through network consensus if a better mechanism emerges, including on other competing blockchain networks. The challenge of understanding the forces shaping these policies[52] is rendered significantly more difficult by the paucity of high quality data with which to characterize the diversity of institutional

[51] This slowing of withdrawals may be seen as helping to stabilize the blockchain, while also increasing the penalty for any validators that are being penalized which is consistent with a desire to impose penalties which are larger when the threat to the blockchain network is greater.

[52] Comparative institutional analysis has long emphasized that rules and enforcement co-evolve with actor behavior and information conditions, making cross-system variation persistent (Aoki 2001).

arrangements we observe. This lack of data and the exigencies confronting new enterprises surviving in such a tumultuous marketplace further complicate the challenges facing industry participants seeking to make informed institutional design choices.

Although we cannot forecast where things will settle, we can appeal to institutional and organizational economic theory to better understand the forces at work that we expect may account for institutional variation in any future equilibria that will emerge. Fundamentally, a blockchain network architecture provides a new framework for coordinating joint production across distributed agents, while aspiring to decentralized control (i.e., with capability of disintermediating and thereby rendering ineffective any centralization of control by a subset of participants). Organizational economic theory holds that the specific coordinative purposes and characteristics of a given network determine the specific rewards and penalties that best align incentives among independent joint producers of that network's output. Yet this can be an over-determinative read of organizational economic efficiency given that the definition of the organizational forms of business partnerships and corporations evolved over centuries of incremental adjustments among the contracting parties, constrained in part by the courts interpreting their arrangements in cases of disputes and failure (Langlois 2023, Lamoreaux 1998).

There is thus an incremental and emergent nature (Alchian and Demsetz 1973; Alston 2025) to the definition of coordinative institutions facilitating joint production; participants' collective definition of contractual terms that maximize the surplus to such production responds to changing relative prices, technological change, and demographic change, to name just a few forces exogenous to these participants' control. This means institutional definition is in part a discovery process, where changes to existing institutions are risky yet offer greater potential value than more certain institutional arrangements, which makes institutional change at a grand scale subject to persistence and sticky equilibria (Arthur 1989; David 1985; North 1990). This implies that institutional variation is also partly characterized by an evolutionary process (Alston et al 2025), albeit one governed by certainty in existing institutional choices relative to uncertain novel forms. This leads to both a random walk component to institutional evolution, as well as benefits from following existing institutional choices among sufficiently similar joint productive contexts (Devereaux 2025), a process we call institutional mimicry. Finally, the study of default rules also indicates that rules have value even when not enforced (Wallis 2022) through better aligning the incentives of those subject to those rules' enforcement.

In light of this institutional and organizational theory, we explore four distinct processes contributing to the evolution of staked blockchain networks' penalty institutions that arguably give rise to the observed variation.[53] These include some blend of (1) coordinative efficiencies, (2) random walk, (3) mimicry, and (4) default rule existence. We consider each of these explanations in turn considering the variation we have documented among penalty conditions and magnitudes on the various proof-of-stake networks we have analyzed thus far. Each of those processes are affected by considerations of resource constraints (cost minimization) and industry economics (scale/scope/learning economies, path dependencies etc.). The latter may lead to expectations that the industry is subject to winner-take-all economic forces and such beliefs when coupled with resource constraints may explain the behavior of industry participants governed by each of the four prototypical processes. For example, coordinative efficiencies, rational cost-minimization, and belief in winner-take-all economics might lead to follow-the-leader mimicry. Alternatively, severe resource constraints might justify behavior that may drive what appear to be random institutional design behaviors.

[53]

4.1. Deliberate Choice for Coordinative Efficiencies

First, there is an important overarching sense in which the penalties on staked blockchain networks are the product of deliberate institutional choice.[54] Indeed, as we explained earlier, there are compelling economic rationales for staked networks to have explicit penalty institutions. The emergence of the proof-of-stake consensus mechanism was partly a response to the limitations of proof-of-work consensus mechanisms, and partly due to the beneficial properties of such staking systems in their own right (Saleh 2021; Kiayias et al. 2017; Bentov et al. 2016). This is not to say that this design is superior when it comes to all coordinative purposes, but simply that it was the product of deliberate institutional choice on the part of designers, perceiving comparative and absolute benefits to governing a cryptocurrency network this way. Considering the cost and uncertainty associated with launching a blockchain network, competing for users and capital with other networks, and facing an uncertain regulatory environment, it is clearly the case that the choice of a staked network was intended to facilitate a particular joint productive purpose. This is especially likely when it comes to cases of significant variation – our methodology was to first take all the staked networks among the top one hundred by market capitalization, but some major staked networks, such as Cardano, do not have automated validator penalties.[55] To posit that such variance is random simply defies logic, such that our first generalization is that sufficiently large variance in choice of penalties is more likely to be the result of deliberate choice than minor variation. This is also consistent with the economic rationale for why explicit reward *and penalty* rules are expected to be beneficial for staked blockchain networks in any future sustained equilibrium, while recognizing that putting such rules in place may not be the most important or likely first or next step for different networks.

4.2. Random Walk

Another explanation for variation may be that the design of the penalty policies is the result of intentional randomization. That might result from simply a lack of attention or resources devoted to specifying the penalty policies, and that lack of resources might be due to cost constraints or a belief that whatever automated penalties that a staked blockchain network might adopt could be at least partly irrelevant to future outcomes. A belief in the irrelevance of automated penalties might be rationale for a blockchain network that was adopting a wait-and-see attitude with a plan to follow once more information accumulates and a belief that whatever is done in the short-run has limited long-run implications for the network or the industry more broadly.

Another obvious explanation for an appearance of random variation in the penalty institutions may simply be the result of measurement error due to the paucity of available information in institutional design choices. The less important a blockchain community and its analysts think penalties are as a governance mechanism, the less resources are likely to be devoted

[54] At a general comparative institutional level, arguments for convergence and variation both have strong scholarly support. While institutional isomorphism involves the recognized tendency for organizations in certain contexts to become more alike because of regulatory/coercive demands, professional norms, and imitation under uncertainty (DiMaggio & Powell 1983), Ostrom's fieldwork showed that long-lived commons repeatedly use graduated sanctions and diverse rules-in-use, implying common functional reasons for heterogeneous penalty designs (Ostrom 1990).

[55] The fact Cardano does not yet have automated penalty policies does not imply that this is an equilibrium outcome that will be sustained in the long-run that is either in Cardano participants' best interests or future intention, but it might be, due to the arguments made by developers and informed users on this network that Cardano's protocol achieves incentive alignment through alternative means (Brünjes et al. 2020; David et al. 2018; Cardano n.d.).

to documenting those penalties, which means that some of the randomness may be simply measurement error.

Intentional sources of random walk variation may be to facilitate experimentation and/or may be attributed to the desire for blockchains to differentiate themselves to better attract participants.[56] Thus, from the perspective of analysis of complex adaptive systems (Alston et al 2025), not all variation therein serves a singular purpose for any given varying element within that system, even as such variation itself may lead to broader systemic benefits. The storied example of this is that of genetic mutations, many of which have no effect, even as the beneficial variation that is revealed from behind the veil of uncertainty operates to facilitate survival of those that experience the randomly beneficial variation. While systems subject to deliberate choice mean that variation is not completely outside of the control of those subject to it, this also means that protocol designers versed in the study of complex systems may choose minor variance deliberately even as they do not have a clear understanding of what the effects of that choice of variance will be. A related perspective on this surrounds the fact that in a competitive environment, an identical competitor to an already existing option is unlikely to attract new users or capital, such that there are benefits to choosing minor margins of variation to distinguish oneself, even as these margins of variation are not tightly tethered to an expected benefit from the choice in rule variation. Of course, these sources of observed rule variation are not mutually exclusive, such that a given protocol designer could have a weak expectation of the benefits of changing a penalty mechanism can also be motivated by the desire to display sufficient distinction from existing networks.

4.3. Mimicry

The third process we see at work is mimicry when a staked blockchain network follows what others have done – either other successful networks or copying design features that appear to have successfully addressed analogous challenges in the off-chain, pre-chain world. It is collectively and individually costly and uncertain to design new coordinative institutions. Being a follower rather than an explorer may prove the better strategy for many. The history of institutions on frontiers is one that displays considerable mimicry (Alston et al 2012). Human groups along new frontiers do not emerge from broadcloth but instead are coming to the frontier from an environment with which they are already familiar (Anderson & Hill 1979; Umbeck 1977). Such groups tend to take the institutions with which they are familiar that are closest to the novel context they are confronting, and apply those institutions until the need for adjustment becomes apparent.[57] This serves to motivate our third explanation for the observed variation in penalties on staked blockchain networks, in that institutional designers mimic the contexts they know best and then make minor adjustments in furtherance of a particular coordinative purpose,[58] another case of complementarity between our forces of variation.

[56] If exercising choice within a sufficiently rugged fitness landscape, local experimentation can trap systems in locally optimal but path-dependent regimes; diversity in rule sets may improve robustness (Levinthal 1997; Page 2011).

[57] This was true in the case of mining institutions in the United States, where mining rules that were initially applied in the east were adopted in the west with the exception of rules pertaining to how minerals ran underground due to the distinct geology in each location leading to a need for distinct institutional treatment (Gates 1968; Umbeck 1977)

[58] Institutional change often proceeds via layering, conversion, and drift rather than wholesale replacement, consistent with incremental parameter updates to penalty regimes (Mahoney & Thelen 2010).

Evidence of mimicry surrounds the fact that networks sharing the same originating code base tend to have more similar penalties than those with distinct code bases. Forking the code of an open source project is very common for software developers, such that we expect mimicry to be most present in the case of staked blockchain networks that have the same original source code. This can also be understood as a means of reducing the uncertainty that institutional design within complex systems entails – by going with a proven codebase that has already successfully coordinated network validators and users, the risk of new parameter choices is limited to those that the designers deliberately tweak. Conversely, the choice to design and deploy a staked blockchain network with a completely novel codebase is quite likely to indicate a significantly different intended joint productive purpose to the network's activities.

Additional forces motivating mimicry may include cost-minimization (efficient resource optimization) since it is wasteful to duplicate good code that can be shared assuming that adoption does not forego strategic flexibility. Mimicry may also be justified by participants' belief that winner-take-all economics will dictate the future equilibrium, and followers may be motivated by the belief that they have identified the likely winning design choice. However, that belief need not be consistent with the belief that the expected winner will be a good choice for the ecosystem. It may be that path dependencies result in expectations of a lowest-denominator or lock-in to an early leader's design choice. The potential lock-in effects that bandwagon expectations may play provides a key reason why the economics of industry standardization may fail to select the best technologies.[59] Our early-stage research on blockchain penalty institutions is motivated, in part, by our concern that potential lock-in effects might manifest in the blockchain ecosystem and early awareness may facilitate avoiding such ecosystem pitfalls.

4.4. Default Rules in Spirit, if not in Fact

A fourth lens through which to consider the variation in penalties is taken from the study of default rules, and is also weakly complementary to the other rationales we proposed with respect to the observed variation. Just as the threat of termination operates to incentivize employees that are never terminated, the threat of penalties can have significant behavior and structural implications even if penalties are rarely assessed and policy is more notional than actual from the perspective of its empirical incidence. That is: the statement that a staked blockchain network has automated penalties even when they may not actually be enforced with predictable regularity can still have an impact in an evolutionary system (i.e., cheap talk can matter!).

Relatedly, the need for certainty in rules can be understood as partly independent from the specific choice in rule itself. Put differently, the existence of a credibly enforced rule may be as important as the precise level at which the rule is specified – some neighborhoods choose a speed limit of twenty miles per hour and others twenty five. However, what may be more important is that drivers expect a sufficient probability of enforcement than the marginal difference between the two speed limits. Having a rule that will be enforced in cases where things go wrong may in certain instances be more important than the exact rule in question, because independent parties coordinating under conditions of uncertainty can sufficiently adjust their behavior whether the rule is twenty or twenty-five miles an hour, and in each case (given sufficient enforcement probability) the intent of the institutional designers is reliably obtained. Such variation in the specifics of the rules introduces a level of intentional discretion to blockchain protocol designs to what might otherwise aspire to be fully algorithmic and free of discretion. Clearly, this argument does not hold

[59] See, for example, Lehr (1992) or David (1986).

in cases of radical variation in rules, as all neighborhoods do not have speed limits that mirror those of highways and interstates, a facile insight which emphasizes our original argument that sufficient variation in penalties is a likely indication of different joint productive purposes for one network relative to another.

The beneficial incentive alignment that reward and penalty regimes can produce relies on another element: those subject to the penalties (as well those investing in or consuming the output of the joint productive context being governed by the penalties) need to be sufficiently aware of the penalties for their alignment effects to occur.[60] But the lack of discretionary monitoring has a corollary implication for the observability of penalty regimes on staked networks. While any node can observe validators' actions on the network, the incidence of penalties is not easily observable to the average user of a given network. This has led to the emergence of "slashboards" for networks above a particular market capitalization or common originating codebase, with Ethereum being a prominent example of the former (beaconcha.in "Slashings"; BeaconScan "Validators that were slashed"), and staked networks built upon the Cosmos SDK being a clear example of the latter (Mintscan; Cosmos SDK "Slashing" module docs).

These slashboards allow users and other network stakeholders a means by which to observe the occurrence of penalties on networks without themselves running the client software through a node to observe the actual state of the ledger. While the emergence of these on some of the largest staked networks can be understood as filling a gap otherwise created by a lack of centralized monitoring function, this very need can also be understood as a weakness in the inherently decentralized nature of these systems – they provide no way for users to monitor until a sufficiently interested or altruistic ecosystem participant decides to set one up. This also serves as a limitation for comparing the incidence of automated penalties across different networks, arguably a signal of the comparative effectiveness of a given penalty regime. However, the emergence of these independently produced monitoring tools also stands as an indication that centralized monitoring within the organizational confines of joint production (Alchian & Demsetz 1972; Holmström 1982) is not a necessary condition for such a monitoring function to obtain in institutional practice. Yet given the fact that these publicly-legible slashboards are uncommon across the networks we considered (all of which coordinate significant economic value at current market capitalizations), this suggests that a primary audience for penalty choice is more likely the validators subject to these penalties than it is the individual users of a given network, although once the economic magnitude of a given network reaches a certain point, it appears that there are emergent benefits to making this information available publicly.

In summary, the penalty policies themselves are expected to have an organic evolutionary aspect that will allow them to evolve in response to and in conjunction to the evolution of the industry and the blockchain network's role within it. Some networks may be waiting to specify penalty policies, others may see benefits in specifying policies that are not enforceable, while still others may ultimately rely on the complementary emergence of enforcement capabilities and resources that are related to but not directly part of the network's formal institutional design.

[60] To take but one storied example, investors need to believe sufficiently in well-aligned corporate governance mechanisms to invest at the impersonal scale that modern capital markets entail (La Porta et al. 1998; Shleifer & Vishny 1997).

5. Automated Penalty Institutions in Digital Governance Contexts

Ultimately, the penalties that apply to human actors in private organizations (firms) are constrained by the range of social norms and institutional controls that apply to humans (e.g., basic "human" rights, but also a growing number of economic rights such as copyright). For example, employment law constrains what organizations can do to employees – e.g., it is illegal to kill you for failure to do your job.[61] This is surely a major reason for asymmetry between penalties and reward structures in corporate governance. It is generally acceptable to allow rewards to be open-ended and unbounded, but penalties are limited (e.g., bankruptcy law constrains what can be clawed back from result of bad decisions). While we are unlikely to accept unlimited penalties and their enforceability would be challenged (just as one proverbially cannot kill a person twice, the total stake of a validator presents an upper limit as to the penalties to which that network identity can be subjected), there are also limits to rewards, limits which are brought into stark relief when the nature of joint production lacks the centralized authority that makes discretionary monitoring and real-time rewards possible.

What an ecosystem composed of smart contracting enables (see Lehr 2021, 2022) is potential to automate a wide range of economic tasks, where the agents undertaking tasks have some discretion, which for humans tends to require mechanisms to improve incentive alignment given that the agent is acting on behalf of a principal (Shoham & Leyton-Brown 2009; Dafoe et al. 2020). As AI increasingly makes smart agents with increased autonomy (i.e., "smart contracts enable automation of automation") feasible, it will become necessary to confront the need to recognize or treat AI agents as holders of limited property rights, because it is only through the exercise of property or decision rights that agents will be able to exercise discretion under conditions of uncertainty. This makes algorithmic processes (like pre-scripted customer service menus that are fully automated) distinguishable from more complex decisions executed by an agent operating on a principal's behalf; investment in furtherance of return maximization conditional on a particular risk tolerance exemplifies this distinction, for the range of possible investments cannot be fully executed upon by the agent, and the complete outcomes of each strategic path is unknowable ex-ante. This algorithmic "agency" is related to ongoing regulatory concerns like the rights of corporations (Citizen's United) and re-examination of copyright laws (Cooper et al., 2024), and AI policy (Lehr, 2024a,b). In short, the focus of corporate governance focuses on human-employee and options for incentivizing behavior aligned with enterprise and broader societal goals by mix of financial and other incentive mechanisms.

While questions of AI's legal personhood are well afield of this analysis, the deployment of AI agents in investment contexts is already occurring (Hendershott, Jones & Menkveld 2011; LeBaron 2001; Farmer & Foley 2009), including but not limited to those acting on blockchain networks (Ante 2024; Karim et al. 2025). In fact, as of this writing, an AI agent has accrued enough ETH to become a validator on the Ethereum network, such that our argument that the effect on penalties on AI agents' incentive functions is no longer theoretical.[62] As the activities of AI agents increasingly becomes a repeated game, this could entail definition of their objective functions to

[61] Additionally, the world of off-chain corporate governance is circumscribed by the larger regulatory and economic context. There are significant regulatory protections for workers that are more likely to circumscribe the scope for corporate punishments (e.g., illegal firing, discrimination, breach of labor contracts which may include collective bargaining, etc.) than the rewards that corporate governance may enable (although there are legal protections constraining reward systems also, including shareholder interest protections).

[62] For an extended discussion, see "Zerebro, the Agent that Bootstrapped Itself to Become an Ethereum Validator." Dec. 12, 2024.

involve both rewards and penalties. To be clear, until an AI agent can be said to have something approaching a human cognitive response to unexpected outcomes, such an agent cannot be said to intrinsically care about being penalized – it will simply seek to better satisfy the objective which it has been encoded to pursue. Thus, a much more sophisticated game theoretic approach will increasingly become necessary to embed an objective function in an AI agent that approaches the rewards and discretionary monitoring that better align human employees' incentives when acting in furtherance of the interests of a principal. To give a motivating example, an AI agent could perceive a highly lucrative short position that would be highly likely to maximize the expected returns for the principal. However, if the short position could harm the market overall (affecting the principal's positions elsewhere), create legal liability for the principal who deployed it (or the individuals who provided the code in the first place (Khalilli 2025), or transgress deeper moral values of the principal, then embedding a simplistic objective function is likely insufficient. More generally, common AI failure modes including reward hacking, negative side effects, unsafe exploration arguably motivate institutionalized penalties and constraints to shape agent behavior at scale (Amodei et al. 2016).

Thus, the ex-ante definition of automatically executing penalties is an emergent institution whose purchase is unlikely to be limited to aligning validator incentives on blockchain networks. Indeed, a way to understand validator nodes is as semi-automated agents pursuing an objective chosen by their human and organizational principals, constrained by the rules of the protocol. Notably, the ample majority of the largest staked blockchain networks by market capitalization embed penalties into the rules that coordinate joint production of tamper-proof blocks of data. Soon there will be network-coordinated contexts where numerous AI agents act on behalf of numerous principals, potentially including individuals, private firms, and government entities alike. These interactions will necessarily involve individual agents making decisions under conditions of considerable uncertainty, both in terms of their individual objectives provided by their human principals, but also as a function of the predicted behavior of other agents in any context of sufficiently complex networked coordination.[63] Agents entering a particular context thus should have their own individual objective functions within a shared set of semi to fully automated penalties that all agents interacting within that context are notified of ex-ante and are subject to as their decisions on behalf of their principal create broader systemic consequences. It is thus quite likely that the institutional variation and ultimately, innovation occurring within staked blockchain networks has much broader implications for individual and social welfare in a world of numerous AI agents interacting and transacting in digitally networked contexts on behalf of a wide range of human principals.[64] Even those uninterested in cryptocurrencies and distributed network governance should take notice and understand that more effective incentive alignment tends to involve rewards and penalties, even as the reliable automation of the latter is a complex task in its own right.

6. Conclusion

Blockchain networks have come to be understood as an organizational economic innovation. While this has led to outsized claims that they will transform all forms of digitally

[63] Large populations of autonomous LLM agents display emergent coordination and norm formation, reinforcing the role for governance hooks and sanctions (Park et al. 2023; Li et al. 2023; Wu et al. 2023).

[64] Agent-based finance shows interacting automated agents can generate market-level externalities and stylized facts; credible, automatic penalties can mitigate system-level risk (Farmer & Foley 2009; LeBaron 2001).

networked organization, they have proven effective at coordinating independent network validators and users in furtherance of several well understood joint productive purposes. While their largest scale applications currently surround the reliable issuance and transfer of abstract network-native unitized instruments (like currency and derivative claims over it), the application of smart contracts to facilitate more complex commitments continues to develop incrementally. This makes the development of these networks of more general organizational and institutional economic interest.

Herein we develop an organizational economic theory of penalty institutions on staked blockchain networks, arguing that the presence of semi to fully automated penalties alongside rewards creates better incentive alignment for validators through more closely approximating the balance of discretionary rewards and penalties that characterize most private organizational contexts where employees' incentives are shaped by both discretionary rewards and termination. Absent penalties, validators have less incentive to ensure their node performs its functions reliably, and those with malicious intent face lower costs to attempt to subvert the network, increasing the likelihood that a bad block could be forced on the network. While the likelihood of a successful attack is low on networks with a sufficiently large validator set, no penalties for bad block proposals or attestations could render the risk of successful attacks more likely. Penalty policies that hold value staked by validators proposing and attesting to the validity of block updates helps promote a joint productive equilibrium that attracts good faith validators and deters bad faith ones. Absent sufficient certainty that malicious attacks will not prevail, good faith validators will be deterred from joining the network, due to the detrimental effect that a successful attack would have on the value of their pledged stake. It is in this sense that the balance of rewards and penalties on staked blockchain networks is akin to the incentive alignment that the employment contracts of private sector firms provide to their workers through discretionary rewards and termination. All firm workers are better incentivized if they know their average coworker is less likely to shirk on the margin (Shapiro & Stiglitz 1984; Kandel & Lazear 1992), just as all validator operators are more incentivized to pledge scarce tokens and ensure their node's performance if they know their counterparts have similarly aligned incentives.[65]

But firm contracts vary considerably in terms of discretionary bonuses and likelihood of termination due to the need to be tailored to the specific nature of the firm's joint productive purpose (Williamson 2005). Whereas much of that variability in corporate governance can be introduced via managerial discretion applied to standard employment contracts with rich reward (bonus) policies, blockchain networks are more tightly constrained by the lack of discretion inherent to their design. Nevertheless, we expect that there ought to be variation in automated penalty institutions to account for Williamsonian customization to fit the context-specific purposes of different staked blockchain networks. We also expect that in the nascent and rapidly evolving cryptocurrency industry and in the face of diverse national and market-specific regulations and business practices, that there would be significant heterogeneity in the rules, maturity, and structure of blockchain networks.

[65] The lack of contractual bond between validators and a central firm authority makes such a relationship akin to a contest where performance is rewarded, with lower performers receiving less in expectation, creating incentive alignment among more independent contributors to a joint productive goal. For example, rank-order tournaments provide strong incentives via relative rewards and implicit penalties for under-performance in noisy environments (Lazear & Rosen 1981).

We explore these differences by undertaking a comparative empirical analysis of the top staked blockchain networks by market capitalization to better understand the form and range of penalties that these networks use to align validator incentives. We find considerable variance in these digital penalty institutions and develop a typology of behaviors that commonly trigger penalties, the form these penalties take (including forfeited tokens and removal from eligibility to validate) as well as the means by which these penalties are assessed, ranging from fully automated to discretionary. We further consider outlier penalty institutions on these networks in terms of their relative novelty.

Yet considering how closely related the joint productive purposes of many these networks are, this poses an additional question as to why we uncover the extent of variance in these penalty regimes. We consider four explanations for the institutional variation we identify. First, we believe sufficiently different choices in penalty regimes by protocol designers indicates deliberate choice on the part of the institutional (protocol) designers, likely in furtherance of distinct joint productive purposes toward which a given network is coordinating. Yet institutional design on new frontiers is itself a discovery process, such that we also characterize some of the observed variation as due to random choice, as well as mimicry following codebase similarity from developers forking existing open source blockchain codebases. Finally, we draw from the study of default rules to argue that specific variation (up to a certain maximal range) is less important for aligning incentives than is the credible enforcement of a rule which enables validators to productively rely on the aligned incentives of other independent validator nodes. These explanations are not mutually exclusive, such that several of these forces of institutional variation are likely to be present with respect to any one staked blockchain network's protocol design choices.

Finally, in a world where networked contexts are drastically increasing in terms of the complexity and economic magnitude of the processes they are coordinating, we argue that the institutional development on blockchain networks that we document and analyze is likely to have broader purchase than facilitating reliable block production on these networks. This is especially the case when it comes to AI agents being deployed in contexts where their decisions are governed by scarcity and uncertainty. The very agency they will have in making risky investments and other decisions means that aligning these agents' incentives with their human and organizational principals will likely require embedding the equivalent of penalties in their objective functions to create an appropriate level of risk aversion as they pursue their encoded goals. As agents act alongside one another and are governed by other agents, this will require a sophisticated coordinative understanding on the part of such agents of one another's incentives in a way that is analogous to the alignment achieved by the balance of rewards and penalties on blockchain networks, making our comparative institutional analysis of penalties on staked blockchain networks of interest to scholars of protocol design and digital governance writ large.

7. References

Aghion, P. and Tirole, J. (1997). Formal and real authority in organizations. Journal of Political Economy, 105(1), pp.1-29.

Agur, M.I., Deodoro, J., Lavayssi√®re, X., Peria, S.M., Sandri, M.D., Tourpe, H. and Bauer, M.G.V. (2022). Digital currencies and energy consumption. International Monetary Fund.

Alchian, A.A. and Demsetz, H. (1973). The property right paradigm. The journal of economic history, 33(1), pp.16-27.

Alchian, A.A. and Demsetz, H. (1972). Production, Information Costs, and Economic Organization. American Economic Review, 62(5), pp.777-795.

Alston, E. (2019). Constitutions and blockchains: Competitive governance of fundamental rule sets. Case W. Res. JL Tech. & Internet, 11, p.131.

Alston, E., L. Alston, and B. Mueller. 2025. Handbook on Institutions and Complexity Theory. Edward Elgar.

Alston, E., Law, W., Murtazashvili, I. and Weiss, M. (2022). Blockchain networks as constitutional and competitive polycentric orders. Journal of Institutional Economics, 18(5), pp.707-723.

Amodei, D., et al. (2016). Concrete Problems in AI Safety. arXiv:1606.06565.

Anderson, T.L. and Hill, P.J. (1979). An American Experiment in Anarcho-Capitalism: The Not So Wild, Wild West. Journal of Libertarian Studies, 3(1), pp.9-29.

Ante, L. (2024). Autonomous AI Agents in Decentralized Finance. SSRN Working Paper.

Aoki, M. (2001). Toward a Comparative Institutional Analysis. Cambridge, MA: MIT Press.

Appendix A: Penalty Characteristics of Staked Blockchain Networks

Aronoff, D.J. (2022). Essays on Incentive Designs to Improve Market Performance (Doctoral dissertation, Massachusetts Institute of Technology).

Arthur, W.B. (1989). Competing technologies, increasing returns, and lock-in by historical events. Economic Journal, 99(394), pp.116-131.

beaconcha.in (n.d.). Slashings (Ethereum Beacon Chain).

BeaconScan (n.d.). Validators that were slashed (Ethereum Beacon Chain).

Becker, G.S. (1968). Crime and punishment: An economic approach. Journal of political economy, 76(2), pp.169 -217.

Bentov, I., Lee, C., Mizrahi, A. and Rosenfeld, M. (2016). Proof of Activity: Extending Bitcoin,'s Proof of Work via Proof of Stake. ACM SIGMETRICS Performance Evaluation Review, 42(3), pp.34-37.

Bradley, M.E. (2007). Efficiency wages and classical wage theory. Journal of the History of Economic Thought, 29(2), pp.167-188.

Bronjes, L., Koutsoupias, E., Kyropoulou, M., and Stouka, A. (2020). Reward Sharing Schemes for Stake Pools. In: AFT 2020 (Proceedings of the 2nd ACM Conference on Advances in Financial Technologies).

Buterin, V. (2014). A Next-Generation Smart Contract & Decentralized Application Platform.

Buterin, V. and Griffith, V. (2017). Casper the Friendly Finality Gadget. arXiv:1710.09437.

Calabresi, G., & Melamed, A. D. (1972), "Property rules, liability rules, and inalienability: one view of the cathedral," Harvard Law Review, 1089-1128

Cambridge Centre for Alternative Finance (n.d.). Cambridge Bitcoin Electricity Consumption Index (CBECI).

Cardano (n.d.). Delegation and Stake Pools (official documentation).

Catalini, C. and Gans, J.S. (2020). Some Simple Economics of the Blockchain. Communications of the ACM, 63(7), pp.80-87.

Coase, R.H. (1977). The wealth of nations. Economic Inquiry, 15(3), pp.309-325.

Conitzer, V. and Sandholm, T. (2002). Complexity of mechanism design. UAI 2002.

Cosmos SDK (2024,2025). x/slashing module documentation.

Cosmos SDK (n.d.). Slashing module documentation.

Cowen, N. (2020). Markets for rules: The promise and peril of blockchain distributed governance. Journal of Entrepreneurship and Public Policy, 9(2), pp.213-226.

Dafoe, A., et al. (2020). Open Problems in Cooperative AI. arXiv:2012.08630.

Daian, P., Goldfeder, S., Kell, T., Li, Y., Zhao, X., Bentov, I., Breidenbach, L. and Juels, A. (2019). Flash Boys 2.0: Frontrunning, Transaction Reordering, and Consensus Instability in Decentralized Exchanges. arXiv:1904.05234.

David, B., Gai, P., Kiayias, A. and Russell, A. (2018). Ouroboros Praos: An Adaptively-Secure, Semi-synchronous Proof-of-Stake Blockchain. In: EUROCRYPT 2018.

David, P. (1986), Narrow Windows, Blind Giants and Angry Orphans: The Dynamics of Systems Rivalries and Dilemmas of Technology Policy, Technological Innovation Project Working Paper, No.10, Center of Economic Policy Research, Stanford University, Paper presented to the Conference on Innovation Diffusion, held in Venice, Italy, March 17 -22, 1986.

David, P.A. (1985). Clio and the Economics of QWERTY. American Economic Review, 75(2), pp.332-337.

de Vries, A. (2018). Bitcoin’s Growing Energy Problem. Joule, 2(5), pp.801-805.

DiMaggio, P.J. and Powell, W.W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality. American Sociological Review, 48(2), pp.147-160.

Edgington, B. (2025). Inactivity leak (Eth2Book).

Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. Quarterly Journal of Economics, 75(4), pp.643-669.

Ethereum.org (2025). Proof-of-Stake rewards and penalties.

Ethereum.org (2025). PoS attack and defense.

Fama, E.F. (1980). Agency problems and the theory of the firm. Journal of Political Economy, 88(2), pp.288-307.

Fama, E.F. and Jensen, M.C. (1983). Separation of ownership and control. Journal of Law & Economics, 26(2), pp.301-325.

Farmer, J.D. and Foley, D. (2009). The economy needs agent-based modelling. Nature, 460, pp.685-686.

Farrell, J. and Saloner, G. (1985). Standardization, compatibility, and innovation. RAND Journal of Economics, 16(1), pp.70-83.

Foucault, M. (1977), Discipline and Punish: the Birth of the Prison, trans. Alan Sheridan, Random House, 1977.

Gates, P.W. (1968). History of Public Land Law Development. Washington, DC: Public Land Law Review Commission.

Gibbons, R. (1998). Incentives in organizations. Journal of economic perspectives, 12(4), pp.115-132.

Gibbons, R. and Roberts, J. (2013). Economic theories of incentives in organizations. Handbook of organizational economics, pp.56-99.

Gillan, S.L. and Nguyen, N.Q. (2016). Incentives, termination payments, and CEO contracting. Journal of Corporate Finance, 41, pp.445-465.

Goetz, C.J. and Scott, R.E. (1977). Liquidated damages, penalties and the just compensation principle. Columbia Law Review, 77(4), pp.554-594.

Greif, A. (2006). Institutions and the Path to the Modern Economy. Cambridge: Cambridge University Press.

Hendershott, T., Jones, C.M. and Menkveld, A.J. (2011). Does algorithmic trading improve liquidity? Journal of Finance, 66(1), pp.1-33.

Hendrycks, D., Mazeika, M. and Woodside, T. (2023). An Overview of Catastrophic AI Risks. arXiv:2306.12001.

Holland, J.H. (1992). Complex adaptive systems. Daedalus, 121(1), pp.17-30.

Holmstrom, B. (1982). Moral hazard in teams. Bell Journal of Economics, 13(2), pp.324-340.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal,agent analyses: Incentive contracts, asset ownership, and job design. Journal of Law, Economics, & Organization, 7, pp.24-52.

Hylton, K.N., Punitive Damages and the Economic Theory of Penalties. Georgetown Law Journal, Vol. 87, No. 2, November 1998.

Jensen, M.C. and Meckling, W.H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics, 3(4), pp.305-360.

Kandel, E. and Lazear, E.P. (1992). Peer Pressure and Partnerships. Journal of Political Economy, 100(4), pp.801-817.

Karim, M.M., et al. (2025). AI Agents Meet Blockchain: A Survey on Secure and Scalable Collaboration. Future Internet, 17(2), 57.

Katz, M.L. and Shapiro, C. (1985). Network externalities, competition, and compatibility. American Economic Review, 75(3), pp.424-440.

Khalili, J. (2025). Tornado Cash Developer Roman Storm Guilty on One Count in Federal Crypto Case. Wired.

Kiayias, A., Russell, A., David, B. and Oliynykov, R. (2017). Ouroboros: A Provably Secure Proof-of-Stake Blockchain Protocol. In: CRYPTO 2017.

Knight, F.H. (1921). Risk, Uncertainty and Profit. Boston: Houghton Mifflin.

Kubanek, J., Snyder, L. H., & Abrams, R. A. (2015). Reward and punishment act as distinct factors in guiding behavior. Cognition, 139, 154-167.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, R.W. (1998). Law and Finance. Journal of Political Economy, 106(6), pp.1113-1155.

Laffont, J.-J. and Martimort, D. (2002). The Theory of Incentives: The Principal-Agent Model. Princeton: Princeton University Press.

Lamoreaux, N.R. (1998). Partnerships, Corporations, and the Theory of the Firm. The American Economic Review, 88(2), pp.66-71.

Langlois, R.N. (2023). The corporation and the twentieth century: the history of American business enterprise. Princeton University Press.

Lazear, E.P. and Rosen, S. (1981). Rank-order tournaments as optimum labor contracts. Journal of Political Economy, 89(5), pp.841-864.

LeBaron, B. (2001). Empirical regularities from agent-based computational finance. Journal of Economic Dynamics & Control, 25(9,10), pp.1447-1475.

Lehr, W. (2022), "Smart Contracts: Myths and Implications for Economics and Financial Regulation," TPRC50: Policy Research Conference on Communications, Information and the Internet, September 2022.

Lehr, W. (2021) Smart Contracts, Real -Virtual World Convergence and Economic Implications, TPRC49: Policy Research Conference on Communications, Information and the Internet, September 2021.

Lehr, W. and V. Stocker (2024a), "Competition Policy over the Generative AI Waterfall," forthcoming Concurrences, Fall 2024.

Lehr, W. and V. Stocker (2024b), "AI Regulation and 6G: The Measurement Ecosystem Challenge," in Dynamics of Generative AI (ed. Thibault Schrepel & Volker Stocker), Network Law Review, Winter 2023.

Lehr, W. (1992). Standardization: Understanding the Process. Journal of the American Society for Information Science (1986 -1998), 43(8), 550. Mackenzie, J. S. The Relation Between Ethics and Economics., International Journal of Ethics 3, no. 3 (1893): 281-308.

Leibo, J.Z., Zambaldi, V., Lanctot, M., Marecki, J. and Graepel, T. (2017). Multi-agent reinforcement learning in sequential social dilemmas. AAMAS 2017.

Leshy, J.D. (1987). The Mining Law: A Study in Perpetual Motion. Washington, DC: Resources for the Future.

Levinthal, D.A. (1997). Adaptation on rugged landscapes. Management Science, 43(7), pp.934-950.

Li, G., et al. (2023). CAMEL: Communicative Agents for Mind, Exploration of LLMs. arXiv:2303.17760.

Mahoney, J. and Thelen, K. (2010). Explaining Institutional Change: Ambiguity, Agency, and Power. Cambridge: Cambridge University Press.

Maskin, E.S. (2007). Mechanism Design: How to Implement Social Goals. Nobel Prize Lecture, 8 December 2007.

Mattei, U. (1995). The Comparative Law and Economics of Penalty Clauses in Contracts. The American Journal of Comparative Law, 43(3), 427-444.

Mintscan (n.d.). Interchain Block Explorer (Cosmos ecosystem).

Motepalli, S., & Jacobsen, H. -A. (2023). Analyzing Geospatial Distribution in Blockchains. arXiv (Cornell University).

NEAR (2019-2025). White Paper and Validator documentation (Nightshade; progressive slashing).

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System.

Narayanan, A., Bonneau, J., Felten, E., Miller, A. and Goldfeder, J. (2016). Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction. Princeton: Princeton University Press.

Nguyen, Trung-Viet, et al. PenChain: A Blockchain -Based Platform for Penalty-Aware Service Provisioning., IEEE Access, vol. 12, 2024, pp. 1005-30.

North, D.C. (1990). Institutions, Institutional Change and Economic Performance. Cambridge: Cambridge University Press.

Ostrom, E. (1990). Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge: Cambridge University Press.

Page, S.E. (2011). Diversity and Complexity. Princeton: Princeton University Press.

Park, J., et al. (2023). Generative Agents: Interactive Simulacra of Human Behavior. arXiv:2304.03442.

Polinsky, A.M. and Shavell, S. (2000). The economic theory of public enforcement of law. Journal of Economic Literature, 38(1), pp.45-76.

Polkadot (2025). Slashing (documentation).

Posner, Richard A., An Economic Theory of the Criminal Law, Columbia Law Review, vol. 85, pp. 1193-1231 (1985).

Prendergast, C. (1999). The provision of incentives in firms. Journal of Economic Literature, 37(1), pp.7,63.

Perolat, J., et al. (2017). A multi-agent reinforcement learning model of common-pool resource appropriation. AAMAS 2017.

Rajagopalan, S. (2018). Blockchain and Buchanan: Code as constitution. James M. Buchanan: A theorist of political economy and social philosophy, pp.359-381.

Rothchild, Jonathan. Ethics, Law, and Economics: Legal Regulation of Corporate Responsibility. Journal of the Society of Christian Ethics 25, no. 1 (2005): 123-46.

Roughgarden, T. (2024), Provable Slashing Guarantees. In ACM Symposium on Principles of Distributed Computing (PODC'24), June 17-21, 2024, Nantes, France. ACM, New York, NY, USA.

Saleh, F. (2021). Blockchain Without Waste: Proof-of-Stake. Review of Financial Studies, 34(3), pp.1156-1190.

Shapiro, C. and Stiglitz, J.E. (1984). Equilibrium unemployment as a worker discipline device. American Economic Review, 74(3), pp.433-444.

Shavell, S. (1984). A model of the optimal use of liability and safety regulation. RAND Journal of Economics, 15(2), pp.271-280.

Shavell, Steven, Criminal Law and the Optimal Use of Nonmonetary Sanctions as a Deterrent, Columbia Law Review, vol. 85, pp. 1232-1262 (1985).

Shleifer, A. and Vishny, R.W. (1997). A Survey of Corporate Governance. Journal of Finance, 52(2), pp.737-783.

Shoham, Y. and Leyton-Brown, K. (2009). Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. Cambridge: Cambridge University Press.

Sompolinsky, Y. and Zohar, A. (2015). Secure High-Rate Transaction Processing in Bitcoin. In: Financial Cryptography and Data Security.

Telser, L. G. A Theory of Self -Enforcing Agreements. The Journal of Business 53, no. 1 (1980): 27-44.

Tezos (2025). Consensus & Slashing (Tenderbake/Emmy) documentation.

Umbeck, J. (1977). The California Gold Rush: A Study of Emerging Property Rights. Explorations in Economic History, 14(3), pp.197-226.

Wallis, J.J. (2022). An Alternative Institutional Approach to Rules, Organizations, and Development. The Journal of Economic History, 82(2), pp.335-367.

Williamson, O. E. (1979a), Commentary: Williamson on Predatory Pricing II, The Yale Law Journal, 88(6), 1183-1200.

Williamson, O.E. (2005). The economics of governance. American Economic Review, 95(2), pp.1-18.

Williamson, Oliver E. (1979b), Assessing Vertical Market Restrictions: Antitrust Ramifications of the Transaction Cost Approach, University of Pennsylvania Law Review, vol. 127, no. 4, 1979, pp. 953-93, https://doi.org/10.2307/3311789

Wood, K., Mendes, H., & Pulaj, J. (2024). An Optimal Multilevel Quorum System for Probabilistic Consensus. arXiv (Cornell University).

Wright, A. The Rise of Decentralized Autonomous Organizations: Opportunities and Challenges. Stanford Journal of Blockchain Law and Policy. 2021.

Wu, Q., et al. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. arXiv:2308.08155.

Appendix A: Penalty Characteristics of Staked Blockchain Networks

Staked Blockchain Network Penalty Characteristics	Double Sign Penalty	Downtime Penalty	Jailing/Removal from Validator Set	Withdrawal Queue
Ethereum	1 eth (1/32 of validator stake)	.08 eth per 36 days	Yes	Dynamic - 16 withdrawals maximum in a single block
BNB	200 BNB	10 BNB	Yes	7 Days
Toncoin	Variable	101 TON	No	9 hours
Polkadot	Dynamic dependent on the percentage of validators committing the same infraction	Same as double sign penalty although minor offenses such as a downtime penalty may not incur slashing	Yes	28 days
Avalanche	None	None	N/A	14 day lockup
Tron	None	None	N/A	14 days
Cardano	None	None	N/A	0
Polygon	None	None	N/A	80 checkpoints (3-4 days)
OKT Chain	0	0	Ues	N/A
Cosmos	0.05	0.01	Yes (10 minutes for downtime; tombstoned for double signing)	21 days
Cronos	0.05	0.001	Yes (permanently banned for byzantine fault)	28 days
Hedera	None	None	N/A	0 days
ICP	Dynamic	Dynamic	N/A	7 days
Aptos	None	None	N/A	This depends on the number of days into the 30-day cycle (e.g., 25 days into the 30-day cycle means a 5-day withdrawal queue).
Thorchain	0.05	slash points that deduct block rewards	N/A	None
Near	None	None	N/A	4 epochs (~2 days)
Injective	0.005	0.0001	Yes (10 minutes for downtime;	21 days

			tombstoned for double signing)	
MultiversX	Potentially all stake (based on rating system)	Rating system (no direct slashing; rating decreases)	Yes (jailed if rating drops below 10 points)	10 days
Algorand	None	None	N/A	N/A
Theta Network	None	None	N/A	60 hours (2.5 days).
Flow	Case by case basis	50% of stake rewards removed for liveness issues	N/A	2 epochs (14 days)
Sui	Tally system. Under tally score of 0, rewards slashed	Under tally score of 0, rewards slashed.	N/A	0 days (instant withdrawal)
Fantom	Potentially all stake	None	N/A	7 days
CORE	Entire reward and CORE validator deposit slashed	50 blocks missed = accrued token rewards slashed; 150 blocks missed = 10% of stake slashed	Yes	7 days (liquid staking tokne)
EOS	None	None	N/A	21 days
Sei	5% of stake	1% of stake	Yes (jailed for a period after slashing)	21 days
Chiliz	N/A	Missed blocks lose epoch rewards; jailed for 7 epochs for extended downtime	Yes (jailed for 7 epochs)	7 epochs (~7 days)
Flare	Any rewards during staking period (no stake slashing)	Any rewards during staking period (no stake slashing)	N/A	Not specified
Axie infinity	RON 250,000.00	1000 RON (for missing more than 500 blocks)	Yes (jailed for 2 days for unavailability; permanent for double signing)	Instantly (if staked for 3 days)
Akash	5% of stake	.01% of stake	Yes (jailed for downtime; can submit unjail transaction)	21 days
Gnosis	1/32 of staked GNO	0 (penalties equal to foregone rewards)	Yes	1-2 days
Tezos	5% of frozen deposit; adaptive slashing for double attestation	Not specified	Yes	4 cycles (~10 days)

Conflux	None (node excluded from next committee period)	None (node excluded from next committee period)	N/A	1-14 days (usually 14)
Oasis Network	100 tokens (minimum stake amount)	0	No	14 days
Klaytn	Not specified	Not specified	N/A	7 days
AIOZ Network	5% of stake	.01% of stake	Yes	28 days
Kava	5% of stake	.01% of stake	Not specified (most likely yes; uses Cosmos consensus)	21 days
Terra Luna	5% of stake	.01% of stake	Yes (600 seconds for downtime; tombstoned for double signing)	21 days
XDC Network	Not specified	Not specified	Yes (jailed for 4 epochs)	Not specified
IoTex	None	.None	not specified	Stake duration + 3 days
Radix	None (planned for the future)	None	Not specified	7 days
Zilliqa	None	None	Not specified	14 days
Enjin	10-100% of stake	0-1% of stake	Yes	28 days (can accept offer to bypass)
Aelf	Not specified	100,000 ELF	Yes	Not specified
Qtum	None	None	Not specified	Not specified
Casper	Currently disabled	None	Not specified	7 eras (14-16 hours)
Chromia	Not specified	Not Specified	Not specified	14 days
Moonbeam	None	None	Not specified	7 days
SAGA	5% of stake	.01% of stake	Potentially yes (uses cosmos consensus)	21 days
Aleph Zero	No automatic slashing implemented	No automatic slashing implemented	Not specified	14 days
WAXP	None	None (staking rewards decrease for underperformance)	Not specified	72 hours (3 days)
Hive	None	None	Not specified	Not specified
Oasys	Suspended for a few days (no stake slashing specified)	Suspended for a few days	Yes	10 days
CUDOS	5% of stake	.01% of stake	Yes (minimum 600 seconds for downtime)	21 days

Humans.AI	5% of stake	None (subject to change)	Not specified	14 days
CANTO	5% of stake	.75% of stake	Yes (30 minutes for prolonged downtime; permanent for consensus faults)	21 days
Lukso	Not specified (portion of LYX burned for malicious behavior)	Increasing amount (1 LYX for 1 day; 60% of stake lost after 18 days)	Ejected if below 16 LYX	16 validators per block exit rate
Agoric	5% of stake	None (temporary removal for missing more than 5% of blocks in 8 hours)	Yes (minimum 10 minutes)	21 days
Entangle	5% of stake	None (jailed for missing more than 50 of the last 100 blocks)	Yes	21 days
Meld	Variable, 3 warnings before slashing (full slashing for operators; percentage for delegators based on pool pledge)	Variable (greater than 24 hours of downtime or less than 98% of uptime over 90 days triggers slashing)	Not specified	Dependent on lockup period
Viction	None	No rewards for 5 epochs	Not specified	96 epochs (~48 hours)
Oraichain	5% of stake	.01% of stake	Not specified	14 days
Zetachain	1% of stake	.1% of stake	Yes	Not specified