Craig Wright / Int.J.Cryp.Curr.Res. 6(1) (2026) 102-133
https://doi.org/10.67191/IJCCR.6.1.2026.102-133

ISSN: 2790-1386

International Journal of Cryptocurrency Research

Publisher's Home Page: https://www.eandtpress.com/

Research Paper **Open Access**

Legal Deterrence in 'Permissionless' Consensus

Craig Wright[1*]

[1]University of Exeter Business School, Exeter, United Kingdom. E-mail: cw881@exeter.ac.uk

Article Info

Volume 6, Issue 1, June 2026
Received : 13 April 2026
Accepted : 06 June 2026
Published : 25 June 2026
doi: 10.67191/IJCCR.6.1.2026.102-133

Abstract

Budish 2025 derives a deterrence condition for proof-of-work consensus under a no-rule-of-law scope condition. This paper extends his framework along two dimensions: it replaces the monolithic anonymous attacker with a pool-centred coalition whose effective capacity depends on member retention, and it admits legal sanctions into the attacker's payoff. Setting both extensions to zero recovers Budish's baseline exactly. The model introduces a value-dependent legal term subject to an enforcer participation constraint: below the enforcement threshold, protocol costs alone determine deterrence. As transaction value rises, the legal term becomes positive and the model distinguishes payout-distortion exit from liability-driven exit. Pool-level data for March 2026 show that the coordination layer is concentrated among publicly attributable operators in a small number of jurisdictions. Pure flow-cost security applies where the enforcer participation constraint binds; for large-value attacks, deterrence is jointly determined by protocol costs, legal sanctions, reputational loss, capital destruction, and longest-chain exclusion.

Keywords: *Rule of law, Legal enforcement, Blockchain security, Mining pools, Reputational sanctions, Double-spending*

1. Introduction

Blockchain security is analyzed in much of the existing literature as if its central economic problem were entirely internal to the protocol. In that view, the relevant question is whether a permissionless consensus system can deter double spending through cryptography, protocol design, and endogenous resource expenditure alone. Budish (2025) gives the clearest statement of that benchmark. His critique of Nakamoto consensus (Nakamoto, 2008) is derived under a no-rule-of-law scope condition and shows that, in a pure permissionless setting, trust at scale requires continuing flow expenditure large enough to deter attack. That result is important and, within its stated domain, precise. The present paper does not reject it. It argues instead that the benchmark becomes progressively less satisfactory as a positive description of economically significant proof-of-work transactions.

* *Corresponding author: Craig Wright, University of Exeter Business School, Exeter, United Kingdom. E-mail: cw881@exeter.ac.uk*

The reason is simple. The relevant attack in Bitcoin-style proof of work is double spending, not arbitrary fabrication of transactions. For trivial transactions, the no-rule-of-law benchmark applies: a one-dollar loss does not justify the cost of legal process, exchange intervention, or tracing and recovery (Polinsky and Shavell, 2007). At that low-value edge, the enforcer participation constraint in equation (6) binds, the legal term is zero, and the system relies on protocol-side deterrence alone. But the same approximation becomes progressively less accurate as transaction value rises. A large double spend does not occur in an institutional vacuum. It implicates victims, exchanges, pools, counterparties, custodians, and legally identifiable organizations with substantial fixed capital, future business value, and operational dependencies. At that point, the attacker's payoff is no longer determined by protocol-resource costs alone.

This paper makes a deliberately narrow move. It extends Budish's framework along two dimensions. First, it replaces the monolithic anonymous attacker with a pool-centred coalition whose effective capacity depends on member retention—an organizational extension motivated by the structure of actual mining. Second, it admits rule-of-law effects into the attacker's payoff—the central institutional extension. Setting both extensions to zero recovers Budish's baseline exactly (Section 3.11). The paper's claim is therefore not that protocol economics cease to matter, nor that legal institutions make proof-of-work security trivial. The claim is that, for economically significant double-spend opportunities, the relevant deterrence condition becomes a hybrid one. Attack profitability depends not only on the cost of assembling and sustaining attacking hash, but also on expected legal sanctions net of enforcement cost, exchange and recovery frictions, reputational and franchise-value loss, member exit from pooled mining arrangements, impairment of specialised ASIC capital, and the possibility of network-based exclusion and longest-chain orphaning. In that environment, Budish's pure flow-cost benchmark is not wrong; it is incomplete.

The paper develops this argument in the context where it matters most: pooled mining. Pooled mining is not an alternative protocol. It is the dominant industrial organization of production under the Nakamoto protocol. The mining literature no longer supports an image of Bitcoin security as the product of a multitude of isolated solo miners. Mining is organised through pools, managers, payout rules, and mobile contributed hash (Cong *et al.*, 2021; Chatzigiannis *et al.*, 2022). That distinction matters analytically. A hash provider may allocate capacity across pools, but that does not make the provider the first-order decision-maker in a deliberate double spend. In pooled mining, the relevant coordinating actor is the pool operator or equivalent block-building entity. Contributors affect the pool's effective capacity and the stability of that capacity, because they can switch when observed payouts deteriorate or when continued association becomes legally or economically dangerous. The pool, not the ordinary contributing machine owner, is therefore the first-order locus of deliberate double-spend attribution.

Once mining is viewed through that pool-centred lens, several consequences follow. First, pooled attacking capacity is endogenous. A covert deviation from honest mining can reduce or delay payouts, increase variance, or otherwise make the pool appear inferior even before contributors understand the precise source of the problem. Those effects can induce exit on ordinary economic grounds. Second, once misconduct is suspected or detected, a separate exit channel opens: continued participation may expose contributors to liability, contractual termination, exchange scrutiny, or the impairment of highly specialised mining capital. Third, the pool operator itself may bear substantial off-chain losses. These include not only civil and criminal exposure, but also franchise-value destruction, counterparty loss, exclusion from off-ramp institutions, and the stranding of ASIC-intensive capital with limited alternative use outside Bitcoin or closely related SHA-256 environments. These channels all weaken the profitability of a large attack relative to the pure no-law benchmark.

The article also argues that the enforcement margin is broader than court judgment alone. A large mining pool depends on naming, routing, and interconnection layers that it does not unilaterally control. Once a major dishonest pool is identified, the relevant response set can extend beyond indictment or civil suit to exchange intervention, counterparty exclusion, route filtering, degraded connectivity, and other operational frictions that raise stale and orphan rates and reduce realised attack value. The point is not that any one state can simply disable global mining at will. It is that large pooled attacks are operationally entangled with externally governed infrastructure and commercially legible intermediaries. That dependence matters for

deterrence because it increases both direct attack cost and the likelihood that ordinary contributors observe worsening pool performance and leave.

The contribution of the paper is threefold. First, it introduces a value-dependent rule-of-law term into Budish's attack condition while preserving the rest of his framework. In the extended model, the expected legal effect is net of enforcement cost, so Budish's result holds at the low-value edge where the enforcer participation constraint binds and invoking law is not worthwhile. Second, it recasts proof-of-work deterrence around pools, members, and mobile hash rather than anonymous standalone miners. This allows the model to distinguish explicitly between payout-distortion exit from hidden attack activity and legal or capital-risk exit from continued association with a dishonest pool. Third, it incorporates two forms of off-chain loss largely absent from the pure benchmark: the destruction of specialised ASIC capital and the fact that once the attacker's effective hash falls below the honest remainder—through member exit, connectivity degradation, or both—the unchanged Nakamoto protocol orphans the attacker's chain. The result is a deterrence condition in which large-value attacks are less profitable than a pure flow-cost model implies, not because the protocol ceases to matter, but because the attack is embedded in a legal, organizational, and operational environment that the no-rule-of-law benchmark intentionally abstracts away from.

This framing also clarifies what the paper does not claim. It does not argue that legal institutions perfectly secure Bitcoin, that all mining is easily attributable, or that courts can reliably restore victims in every case. It does not claim that Nakamoto consensus intrinsically requires law for all uses. A high-throughput system processing large volumes of small payments can operate within Budish's protocol-only regime: each individual transaction is below the enforcement threshold, and the system functions without legal recourse. What the paper claims is narrower: once economically significant transfers occur through legally legible pooled organizations, the no-law benchmark is no longer the correct positive model of attack profitability. The paper accepts Budish's theorem as a theorem about a pure no-law, no-pool benchmark and asks what follows when that benchmark is extended to pooled organization and rule-of-law exposure. The answer is that the relevant security margin becomes transaction-value dependent and organizationally specific.

The remainder of the paper proceeds as follows. Section 2 reviews the protocol-only benchmark literature, the mining-pool and mining-concentration literatures, the enforcement and forensic literature, and the law-and-economics tools needed to extend Budish's payoff structure. Section 3 presents Budish's baseline model, introduces the paper's pool-centred and value-dependent extensions, and develops the formal attack condition including legal sanctions net of enforcement cost, payout-distortion and liability-driven member exit, ASIC-capital impairment, and honest-longest-chain exclusion. Section 4 summarises the enforcement implications and calibrates three components of the model using publicly available data: the enforcer participation threshold, the capital-at-risk ratio, and member-exit dynamics. Section 5 concludes.

2. Literature Review

The modern economics of blockchain security is built around a distinctive benchmark: the security of a permissionless ledger is treated primarily as an internal incentive problem. In that benchmark, the central question is whether a system without trusted intermediaries can deter double spending through cryptography, protocol rules, and endogenous economic cost alone. Budish (2025) states this benchmark most explicitly. His analysis asks whether Nakamoto-style trust can operate economically without government or the rule of law, and his deterrence condition is derived under precisely that scope restriction. That formulation is analytically useful because it isolates the cost of securing a ledger in pure form. But it also narrows the institutional environment in a way that becomes increasingly difficult to sustain once one moves from trivial retail transactions to economically significant transfers mediated through mining pools, exchanges, custodians, and other legally legible organizations.

Important predecessors and adjacent models largely work within the same protocol-centred frame, even when they do not state the no-rule-of-law premise as starkly as Budish. Auer (2019) treats proof-of-work finality as an economic profitability problem and asks how much ongoing expenditure is required to deter double-spend attacks. His comparison is fundamentally between blockchain settlement and conventional forms of finality, not between alternative legal environments inside the crypto sector. Abadi and Brunnermeier

(2018) analyze consensus, finality, and double spending through mechanism design and ledger structure, allowing punishments internal to the system but not building external legal sanction into the attacker's payoff. Sockin and Xiong (2020) broaden the analysis by linking user participation, miner incentives, and platform fragility, yet their treatment of attack remains endogenous to token economics and mining incentives rather than civil liability, criminal exposure, exchange intervention, or recovery risk. Taken together, these papers define the benchmark to which the present article responds: they model deterrence primarily as a protocol-internal problem, with legal institutions either absent, external, or left implicit.

That benchmark remains useful, but it rests on a stylised picture of mining that later work has substantially complicated. The mining literature shows that Bitcoin production is not well described as a mass of atomistic solo miners independently producing blocks. Rather, it is organised through pools, managers, payout rules, fee structures, and mobile contributed hash. Cong *et al.* (2021) show that mining pools became the dominant organizational form early in Bitcoin's history and model them as risk-sharing institutions in which a pool manager coordinates contributors and distributes rewards according to contractual payout rules. Chatzigiannis *et al.* (2022) sharpen this by modelling miners as allocating hash power across multiple pools and currencies in response to expected returns, variance, and portfolio considerations. Albrecher *et al.* (2025) add a further layer by treating miners as firms exposed to ruin risk, operating costs, and stochastic reward streams rather than as frictionless nodes with a single-dimensional incentive.

This literature matters for the present paper for two reasons. First, it shifts attention from isolated machines to the organizational structure of mining. Pooled mining is built around managers who receive and distribute rewards, while contributing miners can move in response to payout expectations and risk. Second, it introduces a distinction that is central here: hash providers are not identical to pools. A provider may allocate capacity across pools, but this affects the pool's effective capacity, not the location of first-order block-building control. For a deliberate double-spend executed through pooled mining, the first-order actor is therefore the coordinating pool entity rather than the ordinary provider merely because that provider contributed hash. Cong *et al.* (2021), together with Chatzigiannis *et al.* (2022), support this organizational distinction by showing that pooled mining is structured around managers, payout design, and mobile contributed hash rather than independent autonomous miners.

Stylianou and Carter (2020) are helpful here, but only with an important qualification. Their contribution is to show that mining pools and exchanges are measurable economic actors and that simple counts of coins, pools, or miners can misstate the underlying structure of the crypto economy. That caution is correct as a matter of industrial organization and measurement. Yet for the present model it should not be pushed too far. The fact that some providers may multihome across pools, or that commonly owned entities may sit behind distinct labels, does not eliminate the pool as the relevant coordination point for a deliberate double-spend. Multihoming changes measured capacity and complicates concentration analysis; it does not displace the coordinating pool operator as the first-order organizational locus of attack. The right implication is therefore not that attribution becomes impossible, but that attribution must run through the economic entity coordinating pooled block production rather than through crude reliance on surface labels alone.

Once mining is understood in this pool-centred way, another literature becomes directly relevant: work on mining concentration and its legal-economic implications. Arnosti and Weinberg (2022) show that Bitcoin mining is a natural oligopoly: cost asymmetries and economies of scale produce highly concentrated equilibria even without collusion. Their model is purely game-theoretic and does not invoke legal or antitrust concepts, but the implication is direct—mining concentration makes the identity and conduct of a small number of large actors economically central. Deuflhard and Heller (2023) build on this by explicitly applying antitrust economics to cryptocurrency mining, arguing that competition-law concepts illuminate the risks of concentration, exclusion, and majority attacks. Together, these papers establish that mining is an oligopolistic market with identifiable, economically significant actors, not a frictionless field of anonymous nodes. Wright (2025) pushes this one step further by listing "no rule of law" as an identifiable Budish assumption and explicitly noting that miners are legal entities and that prosecutions are routine. These papers do not yet integrate a full rule-of-law term into the attacker's payoff. But they show that the idea of mining as a legally insulated, institutionless activity is already under pressure in adjacent work.

The enforcement and forensic literature pushes further still. Foley *et al.* (2019) and Makarov and Schoar (2021) show that even illicit Bitcoin use is not well described by a strong lawlessness premise. Their empirical design depends on tracing, seizures, exchange linkage, and law-enforcement intervention, thereby demonstrating that blockchain activity can be monitored, classified, and connected to identifiable actors through off-chain interfaces. Cong *et al.* (2025) reinforce this point in the context of crypto-enabled cybercrime. They show that major criminal organizations in the sector are not amorphous anonymous swarms but organised, reputation-conscious enterprises, and they emphasise that blockchain transparency and digital footprints facilitate forensics, monitoring, seizure, and shutdown. Atlam *et al.* (2024), in a systematic review of blockchain forensics, synthesise these developments and stress the linkage between on-chain evidence and off-chain data such as IP logs, email records, and financial records, while also noting the evidentiary and jurisdictional limits of such enforcement.

These papers matter here not because they are mining-pool models, but because they weaken the descriptive force of a strong no-law assumption. They support positive values for the probability of detection, attribution, freeze, sanction, and recovery. They also show that realised gains in crypto systems are mediated through exchanges, counterparties, custodians, and other legally legible institutions (Makarov and Schoar, 2021; Cong *et al.*, 2025). Nabilou (2019) makes this institutional point explicitly. He argues that while direct protocol-level regulation is difficult where there is no obvious central command point, effective regulation can still operate through exchanges, banks, payment institutions, and large miners or node operators. For present purposes, this is critical. It means that even if one accepts Budish's pure benchmark as a theorem about protocol-only deterrence, the economically relevant environment for larger transfers is one in which law re-enters through identifiable chokepoints.

That point is strengthened by the current jurisdictional structure of the pool layer. Recent pool-share figures (Appendix A) indicate that pool coordination is concentrated in a small number of jurisdictions. US-linked pools account for 42.01% of reported pool share (with Foundry USA alone at 31.48%), China-linked pools account for 41.51%, Singapore-linked pools (F2Pool) account for 11.33%, and the remaining attributable share is split among Japan (1.88%), the Czech Republic (1.47%), and Slovenia (0.02%). Only 1.77% of hashrate is attributable to pools without verifiable public operator identification. These figures do not establish the physical location of all underlying ASICs and should not be conflated with provider-location data. But they do establish that the coordination layer of mining is highly concentrated and jurisdictionally legible. That fact is more relevant to a deliberate pooled double-spend model than raw hardware geography, because the relevant question is not simply where every machine sits, but which organizations coordinate block production and are exposed to legal and reputational constraints.

The pool literature also shows why pooled attack capacity cannot be treated as fixed once legal and economic consequences enter. Contributors join pools to reduce variance and stabilise returns, not to subsidise a dishonest coordinator indefinitely. Reward systems are built around shares, payout delays, and expected earnings, and switching behaviour responds to maturity structure and expected profitability (Cong *et al.*, 2021; Chatzigiannis *et al.*, 2022). This has a direct implication for any model of a hidden or overt attack. If a pool deviates from honest mining, ordinary contributors may first observe lower realised payouts, abnormal delays, unusual variance, or other performance distortions. Even without proving misconduct, such signals create incentives to move elsewhere. If the pool is then publicly associated with an attack, a second channel opens: continued association may expose members to legal scrutiny, contractual termination, exchange-related frictions, or the impairment of specialised mining capital. The present paper therefore treats pooled attack capacity as endogenous not only to protocol economics, but also to member exit driven by payout distortion and by liability and capital-preservation concerns.

This last point becomes stronger once the specificity of mining capital is brought into view. Bitcoin ASICs are not general-purpose assets. They are highly specialised, system-specific machines with limited alternative use outside Bitcoin or closely related SHA-256 environments. Chatzigiannis *et al.* (2022) emphasise the race toward application-specific hardware, while Deuflhard and Heller (2023) describe contemporary mining as requiring industrial-scale ASIC investment and fixed operational commitments. Wright (2026) formalises the implication: sunk capital in Nakamoto consensus creates a stock-cost component that Budish's flow-cost

framework omits. That matters because exclusion from profitable honest mining, or prolonged impairment of off-ramp access and counterparties, can destroy not merely current revenue but a substantial share of the asset value embedded in the mining operation. Budish's benchmark recognises the importance of stock-cost deterrence in principle, but the present paper argues that for organised pooled mining, legal and operational exclusion can convert a nominally flow-cost attack environment into one that includes very large expected capital losses.

A further complication arises at the network layer. The security and profitability of mining depend not only on hash and law, but also on naming, routing, propagation, and interconnection. Official Internet-governance materials make clear that root-zone management, number-resource coordination, and related identifier functions remain externally governed (Internet Assigned Numbers Authority, 2025), while NIST's routing-security guidance emphasises that BGP route hijacks, leaks, and related instability can deny access, detour traffic, and degrade inter-domain connectivity (Haag *et al.*, 2019). The present paper does not rely on any broad claim that a particular state can simply "turn off" another country's miners. The narrower and more defensible point is that pools depend on naming, routing, and interconnection layers that they do not unilaterally control. Once a major dishonest pool is identified, intervention at those layers can raise propagation delays, stale shares, and orphan rates, thereby worsening ordinary pool performance and reducing effective attack capacity. In other words, network dependence creates an additional enforcement margin that sits between pure criminal liability and pure protocol economics.

At the conceptual level, legal scholarship on blockchain has increasingly moved away from the strong substitution thesis—the idea that blockchain replaces law altogether. De Filippi *et al.* (2022) examine the claim that blockchain is "alegal," outside the legal/illegal binary, and treat that claim critically rather than descriptively. Limata (2019), in a working paper surveying the blockchain-and-institutions literature, applies Werbach's (2018) taxonomy to frame blockchain's relation to law as supplement, complement, or substitute, and concludes that the evidence favours supplementation or complementarity over substitution. These conceptual contributions are not the empirical core of the present article, but they are useful in clarifying that "no rule of law" is not a neutral descriptive statement. It is a scope condition and, in some contexts, an ideological posture. That matters because this paper's contribution is not to deny Budish's theorem, but to argue that the theorem's domain of application narrows as one moves into the world of high-value pooled mining.

The final ingredients come from the law-and-economics literature on organizational misconduct, enforcement selectivity, and deterrence through salience. The Beckerian deterrence framework (Becker, 1968) underlies the present model's legal term: the attacker is a rational agent who weighs expected gain against expected sanction, and deterrence depends on the probability and magnitude of punishment net of enforcement cost. Five recent papers in the *Journal of Law and Economics* provide the specific tools needed to operationalise that framework in pooled mining. Thompson (2024) analyzes the internal organization of La Cosa Nostra and shows that hierarchical firms emerged among criminal actors precisely to limit costly disputes that would raise the group's public profile. The coordination structure—a boss who monopolises the right to authorise extreme action, a court system for peaceful dispute resolution, and residual claimancy to align incentives—maps closely onto the pool-operator/member relationship in mining. The pool operator, like the LCN boss, is the first-order coordinating actor whose decisions determine the organization's risk exposure, and contributing members, like LCN soldiers, retain the option to exit when the coordinator's behaviour threatens their economic or legal position. Bloch and Kranton (2024) study executives' optimal cover-up strategies and show that concealment is sustained when outside investigators face endogenous enforcement costs, but that beliefs evolve: after a successful cover-up, future prosecutors are less likely to investigate, strengthening cover-up incentives. Their model formalises the dynamic that a hidden double-spend attack exploits—the pool conceals its deviation for as long as detection probability remains low—and their result that escalating fines are needed to offset this dynamic supports the value-dependent structure of the legal term in equation (6).

Gulen and Myers (2024) provide direct evidence that enforcement is not uniform. Studying Clean Water Act compliance across US states, they find that violation rates are 39.7% lower in electorally important battleground states, driven by reduced regulatory enforcement intensity rather than improved compliance. Their finding that enforcement is selective and politically contingent supports the enforcer participation

constraint in the present model: the relevant question is not whether law nominally covers double spending, but whether enforcers will bear the cost of invoking it given the political, institutional, and economic environment. Matsuzawa (2025) estimates the causal effect of targeted, salient police enforcement on dangerous behaviour and finds that banning DUI checkpoints increases alcohol-related traffic fatalities by 12.4%. The mechanism is general deterrence through salience: visible, targeted enforcement deters a broader population than the individuals directly affected. This result is directly relevant to the present model because it implies that the deterrence value of identifying and sanctioning a major pool operator extends beyond the specific case—it changes the perceived risk for other pools considering similar conduct.

Finally, Belnap *et al.* (2024) examine automatic information-sharing under FATCA and find that 96.6% of foreign financial institutions publicly commit to sharing account information with the IRS, but that the quality of shared information varies: 20% of total offshore account balances lack taxpayer identification numbers, and compliance costs reduce the profitability of participating institutions. Their study demonstrates that transparency regimes impose real costs on intermediaries and that the effectiveness of detection depends not only on the existence of a reporting obligation but on the quality and completeness of the information actually shared. For blockchain-based transactions, the analogous mechanism operates through exchange-mediated tracing, on-chain forensics, and the coinbase-tag identification of pool operators documented in Appendix A. The critical difference is that blockchain's coordination layer is, by construction, more transparent than the offshore banking system: coinbase tags, public block data, and corporate filings yield attribution rates exceeding 98% at the pool level, substantially above the 80% balance-weighted compliance rate Belnap *et al.* (2024) document for FATCA.

Taken together, these papers supply the modelling vocabulary needed to extend a protocol-only deterrence condition into a rule-of-law framework: organizational hierarchy and coordinator attribution from Thompson (2024), dynamic concealment and escalating penalties from Bloch and Kranton (2024), selective and contingent enforcement from Gulen and Myers (2024), deterrence through salience and identification from Matsuzawa (2025), and the costs and limits of transparency-based detection from Belnap *et al.* (2024). What none of these papers does is apply that vocabulary to proof-of-work double spending in pooled mining. That is the gap this paper addresses. Existing crypto-economics papers model deterrence mainly through protocol costs. Mining-pool papers explain how pooled production and mobile hash work, but do not incorporate rule-of-law sanctions into the attack payoff. Enforcement and legal-interface papers show that crypto activity is not beyond tracing, seizure, or regulation, but do not integrate those channels into the core mining-deterrence condition. The present paper fills that gap by relaxing only Budish's no-rule-of-law assumption and recasting pooled double-spend deterrence as a value-dependent function of protocol costs, legal sanction net of enforcement cost, reputational loss, member exit, capital destruction, and network-based exclusion and longest-chain orphaning.

3. Model

This section retains Budish's basic framework and extends it along two dimensions. The first is organizational: it replaces the monolithic anonymous attacker with a pool-centred coalition whose effective capacity depends on member retention. The second is institutional: it admits rule-of-law effects into the attacker's payoff, including legal sanctions, reputational loss, capital impairment, and longest-chain exclusion. The organizational extension is the necessary premise; the institutional extension is the central contribution. Section 3.11 shows that setting both extensions to zero recovers Budish's baseline exactly.

3.1. Baseline

Let N^* denote equilibrium trust support, c the unit cost of trust support, and p_{block} the block reward. Budish's zero-profit condition is

$$N^*c = p_{block} \quad ...(1)$$

Let A^* denote the attacking share required for the attack and $t(A^*)$ the expected duration for which that attacking share must be maintained; t is decreasing in A because a larger attacking share produces blocks faster on the hidden chain. If the nominal value of a successful double-spend opportunity is V_{tx}, Budish's deterrence condition is:

$A^*N^*c \cdot t(A^*) > V_{tx}$ ...(2)

Equations (1) and (2) are preserved as the pure benchmark. The present paper asks what follows when the attacker is not modelled as operating outside law, outside organization, and outside operational dependencies.

3.2. Actors and Pool Structure

Consider a mining pool p that coordinates block production. Let

- h_P^{own} denote proprietary hash directly controlled by the pool operator;
- h_i denote the hash capacity supplied by outside provider i;
- $\alpha_i \in [0,1]$ denote the fraction of h_i allocated to pool P.

Nominal pool hash is

$$H_P = h_P^{\text{own}} + \sum_{i=1}^{n} \alpha_i h_i \quad \text{...(3)}$$

Equation (3) is a capacity identity only. It does not imply that all contributing providers are equivalent to the pool as attack coordinators. In pooled mining, the pool manager or equivalent coordinating entity receives and distributes rewards, organises block production, and is therefore the first-order actor for a deliberate pooled double spend. Contributing providers matter because they enlarge or shrink the pool's effective attacking capacity and because they may leave once payouts deteriorate or risk rises.

This distinction is central. A provider may allocate capacity across pools, but that affects the pool's feasible power, not the first-order locus of coordinated block-building control. The relevant attacker in the present model is thus not "the miner" in a generic sense, but the pool-centred organization that coordinates the attack and internalises its gains and losses. The structural parallel with Thompson 2024's analysis of criminal organization is direct: in both cases, a hierarchical coordinator monopolises key decisions while autonomous members retain exit options that constrain the coordinator's ability to sustain risky conduct.

3.3. Realized Attack Value

A successful on-chain double spend is not the same as successful realisation of value. Let realised attack value be:

$$V_R(V_{tx}) = \left(1 - \phi_f(V_{tx}) - \phi_r(V_{tx}) - \phi_c(V_{tx})\right) V_{tx} \quad \text{...(4)}$$

where $\phi_f(V_{tx}) \in [0,1]$ is the expected freeze share, $\phi_r(V_{tx}) \in [0,1]$ is the expected recovery or clawback share, and $\phi_c(V_{tx}) \in [0,1]$ is the compliance, liquidation, or monetisation haircut. Each is a function of V_{tx}: $V_{tx} = 0$ at there is no transaction to freeze, recover, or monetise, so $\phi_f(0) = \phi_r(0) = \phi_c(0) = 0$. The frictions become positive and increasing as V_{tx} rises, because larger gains require institutional monetisation that triggers exchange intervention, compliance scrutiny, and counterparty diligence (Makarov and Schoar, 2021; Cong *et al.*, 2025).

Equation (4) captures a point that the pure benchmark omits. Nominal attack value is not identical to value securely captured by the attacker. For economically significant transactions, gains must pass through exchanges, custodians, OTC channels, lenders, or other counterparties before they become durable off-chain wealth (Makarov and Schoar, 2021; Cong *et al.*, 2025). Legal intervention and institutional frictions therefore reduce the relevant prize.

3.4. Pool Legal Profile

Let the legal and organizational profile of pool P be

$$\theta_P = (j_P, o_P, k_P, x_P, m_P, n_P, s_P), \quad \text{...(5)}$$

where j_p is jurisdictional reach, O_p is organizational legibility, k_p is capital specificity and fixed-asset exposure, x_p is exchange, lender, custodian, and counterparty exposure, m_p is dependence on outside mobile hash, n_p is exposure to naming, routing, and interconnection dependencies, and s_p is state-action exposure.

Each element does distinct work. j_p captures whether the pool sits within, or is materially exposed to, a legally reachable jurisdiction. O_p captures whether the pool is linked to a visible firm, public brand, listed company, disclosed operator, or other identifiable entity. k_p captures the degree to which the mining organization depends on highly specialised Bitcoin-specific capital. x_p captures dependence on exchanges, counterparties, custodians, lenders, and fiat-facing off-ramp institutions. m_p captures the extent to which the pool depends on contributed rather than proprietary hash. captures dependence on externally governed routing, naming, and interconnection layers. n_p captures susceptibility to formal or informal intervention by states or regulated infrastructure providers.

Recent pool-share data illustrate the practical content of θ_p. Pool coordination is concentrated in a small number of jurisdictions: US-linked pools account for 42.01% of reported pool share, China-linked pools account for 41.51%, and Singapore-linked pools (F2Pool) account for 11.33%, with Foundry USA alone at 31.48% and MARA Pool (NASDAQ: MARA) at 5.71%. These figures do not establish the physical location of all underlying ASICs, but they establish that the coordination layer is jurisdictionally concentrated and legally legible. In this model, that affects j_p and O_p, not raw hardware geography.

A public, branded, exchange-facing pool with large ASIC fleets and heavy dependence on contributed hash therefore has a larger θ_p in the relevant dimensions than an opaque and weakly intermediated actor.

3.5. Legal Sanction Net of Enforcement Cost

Define the expected rule-of-law term as:

$$\Pi_{\text{law}}(V_{tx}, \theta_P) = \max\{0,\ p_d \cdot p_a \cdot p_s \cdot S - C_E\}, \quad \ldots(6)$$

where all arguments are evaluated at (V_{tx}, θ_p): p_d is the probability of detection, p_a is the probability of attribution, p_s is the probability of effective sanction, freeze, injunction, settlement, recovery, or prosecution, S is the magnitude of sanction and recovery conditional on legal success, and C_E is the expected cost of enforcement invocation and pursuit borne by the enforcing party.

The $\max\{0,\cdot\}$ operator is not a modelling convenience. It captures the participation constraint on enforcers: if the expected cost of pursuing legal action exceeds the expected recovery, no rational enforcer initiates proceedings, and the attacker faces no legal risk. Equation (6) therefore distinguishes the existence of law from the economic value of using law. For very small transaction values, C_E dominates the gross expected sanction, so $\Pi_{\text{law}} = 0$ and the attacker faces protocol costs only. For larger transaction values, expected sanction rises while enforcement costs become relatively less important, and Π_{law} becomes positive.

This distinction is central to the paper. The question is not whether a legal system nominally covers the conduct. The question is whether a victim, prosecutor, exchange, or regulator will rationally bear the cost of invoking it. For a $5 double spend, the answer is no: the cost of a lawyer exceeds the loss. For a $5 million double spend against an identified pool operator with seized assets and exchange-held balances, the calculus changes. The enforcer participation constraint is therefore the mechanism through which Budish's no-law benchmark remains valid at the low-value edge and ceases to be valid at the high-value edge. It is not an exogenous policy parameter. It is the economic condition under which law becomes operative. Gulen and Myers (2024) provide empirical evidence that enforcement intensity varies substantially with the political and institutional incentives of enforcers, reinforcing the point that enforcement is contingent rather than automatic.

Lemma 1 (Value Dependence of Legal Relevance): *There exists an economically relevant range of transaction values over which*

$$\frac{\partial \Pi_{\text{law}}(V_{tx}, \theta_P)}{\partial V_{tx}} > 0. \quad \ldots(7)$$

Proof: By Equation (6), $\Pi_{\text{law}} = \max\{0,\ p_d p_a p_s S - C_E\}$. In the interior where $p_d p_a p_s S > C_E$, $\Pi_{\text{law}} = p_d p_a p_s S - C_E$.

Sanction S is a function of V_{tx}. Disgorgement alone implies $S \geq V_{tx}$; additional penalties (fines, forfeiture, imprisonment-equivalent) are non-negative. Hence $S(V_{tx}) \geq V_{tx}$ and $dS/dV_{tx} \geq 1$: total expected sanctions rise at least dollar-for-dollar with transaction value. Enforcement cost C_E may also increase in V_{tx} (larger cases involve more discovery, expert testimony, and judicial time), but the bulk of enforcement cost is fixed infrastructure: filing, investigation setup, prosecutorial staffing, and institutional capacity. Marginal enforcement cost therefore grows sub-linearly in V_{tx}, so $dC_E/dV_{tx} < dS/dV_{tx}$. Then $\partial \Pi_{\text{law}} / \partial V_{tx} = p_d p_a p_s \cdot dS/dV_{tx} - dC_E/dV_{tx} > 0$, since $dS/dV_{tx} \geq 1 > dC_E/dV_{tx} \geq 0$ and $p_d, p_a, p_s > 0$. Outside the interior, $\Pi_{\text{law}} = 0$ with $\partial \Pi_{\text{law}} / \partial V_{tx} = 0$. At the boundary $V_{tx} = V^{\dagger}$ where $p_d p_a p_s S = C_E$, transitions from zero to positive. For all $V_{tx} > V^{\dagger}$, the derivative is strictly positive.

This does not require a hard threshold. It requires only that legal sanction net of enforcement cost becomes more material as the stake rises. The paper therefore does not claim that law "switches on" at a single exact value. It claims that the no-law approximation weakens as transaction value and actor legibility rise.

3.6. Member Participation and Exit

For each contributing provider *i*, define the expected payoff from staying with pool *P* as

$$U_i^{\text{stay}} = \pi_i - c_i - \ell_i - \lambda_i - z_i \quad \text{...(8)}$$

where π_i is expected payout from staying, c_i is operating cost, ℓ_i is payout-distortion loss from hidden attack activity, λ_i is expected liability cost from continued association, and z_i is expected capital-risk cost, including hardware impairment, freezing, taint, or impaired monetisation.

Let the outside option be

$$U_i^{\text{switch}} = \pi_i^{\text{alt}} - c_i^{\text{alt}} \quad \text{...(9)}$$

Provider remains only if $U_i^{\text{stay}} \geq U_i^{\text{switch}}$, and exits if $U_i^{\text{stay}} < U_i^{\text{swit ch}}$.

This participation structure captures two distinct channels through which pooled attack capacity becomes unstable.

3.6.1. Channel A: Payout-Distortion Exit

Let

$$\ell_i = \ell_i(\delta_i, \nu_P) \quad \text{...(10)}$$

where δ_i is observed payout distortion and ν_p is observed network impairment affecting the pool.

The function ℓ_i is defined as payout-distortion loss: the shortfall in provider *i*'s realised payout attributable to distortion δ_i and network impairment ν_p. By construction,

$$\frac{\partial \ell_i}{\partial \delta_i} > 0, \qquad \frac{\partial \ell_i}{\partial \nu_P} > 0 \quad \text{...(11)}$$

These signs are not assumptions. They follow from the definition of ℓ_i as a loss function: greater observed payout distortion (δ_i) and greater network impairment (ν_p) each increase the shortfall between expected and realised payout, which is ℓ_i.

Contributors need not know that an attack is occurring. The exit condition $U_i^{\text{stay}} < U_i^{\text{switch}}$ is satisfied whenever pool performance worsens relative to alternatives, regardless of whether the contributor identifies the cause. This is the mining analogue of the dynamic formalised by Bloch and Kranton (2024): a hidden deviation is sustained as long as outside observers' beliefs remain favourable, but the concealment itself generates observable distortions that eventually erode those beliefs.

3.6.2. Channel B: Legal and Capital-Risk Exit

Let

$$\lambda_i = \lambda_i(p_d, p_a, p_s, \kappa_i; \theta_P) \quad ...(12)$$

where κ_i is provider i's legal-exposure profile, and let

$$z_i = z_i(\omega_i, \eta_i; \theta_P) \quad ...(13)$$

where ω_i captures hardware and monetisation risk and η_i captures the risk of exclusion from continued profitable mining.

Lemma 2 (Continued-Association Risk)

$$\frac{\partial \lambda_i}{\partial p_d} > 0, \quad \frac{\partial \lambda_i}{\partial p_a} > 0, \quad \frac{\partial \lambda_i}{\partial p_s} > 0, \quad \frac{\partial z_i}{\partial \omega_i} > 0, \quad \frac{\partial z_i}{\partial \eta_i} > 0 \quad ...(14)$$

Proof: By Equation (12), $\lambda_i = \lambda_i(p_d, p_a, p_s, \kappa_i; \theta_P)$ is defined as expected liability cost from continued association with the pool. Expected liability cost is the product of the probability of incurring liability and the magnitude of liability conditional on incurrence. The probability of incurring liability is increasing in the probability of detection (p_d), the probability of attribution (p_a), and the probability of sanction (p_s), since each raises the likelihood that continued association results in a legal consequence for provider i. Formally, if $\lambda_i = p_d \cdot p_a \cdot p_s \cdot L_i(\kappa_i, \theta_P)$ where $L_i > 0$ is the conditional liability magnitude, then $\partial \lambda_i / \partial p_d = p_a p_s L_i > 0$, and similarly for p_a and p_s. By Equation (13), $z_i = z_i(\omega_i, \eta_i; \theta_P)$ is expected capital-risk cost. ω_i captures hardware and monetisation risk; η_i captures exclusion risk. Each is a component of expected loss: higher ω_i raises the expected loss from hardware impairment, and higher η_i raises the expected loss from exclusion. Since z_i is a sum or product of non-negative loss components, it is increasing in each.

As detection and legal exposure rise, and as hardware becomes more vulnerable to stranding or taint, the exit condition $U_i^{\text{stay}} < U_i^{\text{switch}}$ is satisfied for a larger share of contributors.

Let exit probability be

$$q_i = \Pr\left(U_i^{\text{stay}} < U_i^{\text{switch}}\right) \quad ...(15)$$

Then

$$\frac{\partial q_i}{\partial \ell_i} > 0, \qquad \frac{\partial q_i}{\partial \lambda_i} > 0, \qquad \frac{\partial q_i}{\partial z_i} > 0 \quad ...(16)$$

Effective retained pool hash is therefore

$$H_P^{\text{eff}} = h_P^{\text{own}} + \sum_{i=1}^{n} (1 - q_i)\, \alpha_i h_i \quad ...(17)$$

Equation (17) is what distinguishes pooled attack from a static outsider model. The pool's feasible attacking share is not mechanically equal to nominal committed capacity once payout quality, legal exposure, and capital risk begin to deteriorate.

3.7. Reputational, Capital, Longest-Chain Exclusion, and Network Terms

In addition to the legal term, the pool faces four off-chain loss components.

3.7.1. Reputational and Franchise Loss

$$L_{\text{rep}}(V_{tx}, \theta_P) = \rho(V_{tx}, \theta_P) \cdot K_P^F \quad ...(18)$$

where K_P^F is the pool's franchise and future-business capital.

Maintained Restriction 1 (Reputational Exposure): *Reputational loss is increasing in attack size, organizational visibility, and counterparty dependence:*

$$\frac{\partial \rho}{\partial V_{tx}} > 0, \qquad \frac{\partial \rho}{\partial o_P} > 0, \qquad \frac{\partial \rho}{\partial x_P} > 0 \qquad \text{...(19)}$$

This restriction states that reputational loss from a detected attack is increasing in attack size (V_{tx}), organizational visibility (o_p), and counterparty dependence (x_p). The signs cannot be derived from the model alone; they are an empirical claim about how reputational penalties scale. Karpoff and Lott (1993) provide the foundational evidence: studying 132 firms subject to corporate fraud enforcement, they find that reputational losses—measured as the decline in market value net of legal penalties—constitute over 90% of the total penalty, and are increasing in the severity and visibility of the misconduct. The channel is counterparty-facing: fraud victims and trading partners withdraw future business, destroying the firm's quasi-rents.

The scope of this channel is not universal. Karpoff *et al.* (2005), studying 478 firms subject to environmental enforcement actions, find that reputational penalties for environmental violations are *negligible*: market value losses approximately equal the legal penalties imposed, with no additional reputational component. The authors conclude that legal penalties, not reputational penalties, are the primary deterrent to environmental violations. The divergence is informative. Reputational penalties are large where the misconduct involves direct counterparty fraud—the defrauded party can withdraw future business—and negligible where harm is diffuse and external, so that no individual counterparty bears a concentrated loss. A double-spend attack is structurally analogous to counterparty fraud: the attacker deceives specific transaction recipients who accept invalidated payments. The 1993 finding is therefore the relevant empirical analog for the sign restrictions in Equation (19). The 2005 finding supports the broader argument of this paper: where reputational mechanisms alone are insufficient, legal penalties carry the deterrence burden.

3.7.2. ASIC-Capital Loss

Let K_P^{ASIC} be the stock of specialised mining capital tied to the pool's economic operation. Define

$$L_{\text{cap}}(V_{tx}, \theta_P) = \chi(V_{tx}, \theta_P) \cdot K_P^{\text{ASIC}} \qquad \text{...(20)}$$

Lemma 3 (Capital Specificity)

$$\frac{\partial \chi}{\partial V_{tx}} > 0, \qquad \frac{\partial \chi}{\partial k_P} > 0 \qquad \text{...(21)}$$

Proof: The function $\chi(V_{tx}, \theta_P)$ is defined as the expected fraction of ASIC capital destroyed conditional on the attack being detected and attributed. The sign $\partial\chi/\partial k_P > 0$ follows from the definition of capital specificity: k_p measures the degree to which the pool's capital is specialised to Bitcoin mining and has no alternative productive use. A higher k_p means a larger fraction of the pool's capital stock is at risk from exclusion, because specialised assets cannot be redeployed. For the sign $\partial\chi/\partial V_{tx} > 0$: by Lemma 1, expected legal sanction Π_{law} is increasing in V_{tx} for $V_{tx} > V^{\dagger}$. Higher expected sanction raises the probability of enforcement actions that result in seizure, freezing, or exclusion. Each of these outcomes impairs the economic value of the pool's ASIC fleet. Hence the expected fraction of capital destroyed is increasing in V_{tx}.

The term is not a symbolic placeholder. Publicly listed miners disclose rig counts in the hundreds of thousands and energised hashrate measured in tens of exahashes per second.[1] Section 4.2.2 calibrates the capital-at-risk ratio using MARA's disclosed figures and shows that capital exposure exceeds illustrative attack targets by two to three orders of magnitude.

3.7.3. Longest-Chain Exclusion

Let a be the attacker's nominal share. Let $\sigma \in [0,1]$ denote suppression of the attacker's effective connectivity or propagation. Then the attacker's effective share is

[1] MARA Holdings (NASDAQ: MARA) reported approximately 400,000 mining rigs and 53.2 EH/s of energised hashrate as of December 31, 2024, with 44,893 bitcoin held at year-end 2024. Source: MARA Holdings, Inc., Annual Report on Form 10-K for the fiscal year ended December 31, 2024, filed March 3, 2025.23at August 15, 2021.

$$a^{\text{eff}} = a(1-\sigma) \quad \text{...(22)}$$

The honest remainder of the network is (1 – a). Honest miners follow the unchanged Nakamoto protocol: they mine on the longest valid chain they observe. No coordination, protocol modification, or strategic decision is required. The longest-chain rule excludes the attacker whenever honest hash exceeds the attacker's effective hash:

$$(1-a) > a^{\text{eff}} \quad \text{...(23)}$$

Substituting Equation (22), this reduces to

$$(1-a) > a(1-\sigma) \quad \text{...(24)}$$

Equation (24) is the Nakamoto protocol operating as designed. The attacker builds a hidden chain; honest miners continue mining on the public chain. When the attacker reveals its chain, the longest valid chain prevails. If honest hash exceeds the attacker's effective hash, the attacker's chain is orphaned. The σ term captures infrastructure-layer degradation—route filtering, propagation delays, connectivity impairment—that reduces the attacker's effective share without any change to the protocol rules. Without degradation (σ= 0), Equation (24) reduces to (1 – a) > a, which is the standard Nakamoto majority condition a < 0.5.

Let the expected longest-chain-exclusion loss be

$$L_{\text{excl}}(a,\sigma) = \psi(a,\sigma)\cdot B_P \quad \text{...(25)}$$

where B_p > 0 is the continuation value of mining rewards (Biais *et al.*, 2019), $\psi \geq 0$ by construction (it is the expected fraction of continuation value lost when the attacker's chain is orphaned), and ψ rises sharply as condition (24) is approached or satisfied.

3.7.4. Network-Friction Loss

Define

$$L_{\text{net}}(V_{tx},\theta_P) = \tau(V_{tx},n_P,s_P)\cdot B_P \quad \text{...(26)}$$

where $\tau \geq 0$ is the expected fraction of continuation value lost through degraded naming, routing, propagation, stale shares, orphaning, or interconnection suppression.

Maintained Restriction 2 (Network Dependence): *Operational losses are increasing in dependence on external naming, routing, and interconnection layers (n_p) and in state-action exposure (s_p):*

$$\frac{\partial\tau}{\partial n_P} > 0, \qquad \frac{\partial\tau}{\partial s_P} > 0 \quad \text{...(27)}$$

This restriction states that pools more dependent on external naming, routing, and interconnection layers (n_p) and more exposed to state-action (s_p) face larger operational losses if those layers are impaired. The signs cannot be derived from the model alone. They are supported by the empirical evidence on BGP manipulation and routing attacks: Apostolaki *et al.* (2017) demonstrate that routing-level intervention can isolate mining capacity, delay block propagation, and degrade effective hashrate, and Haag *et al.* (2019) confirm that BGP instability can deny access and detour traffic across inter-domain boundaries. A pool with greater dependence on these externally governed layers is, by the mechanism documented in these papers, more vulnerable to degradation.

3.8. Full Net-Gain Function

Let H_{total} denote total network hash. The pool's effective attacking share is:

$$A_P = \frac{H_P^{\text{eff}}}{H_{\text{total}}} \quad \text{...(28a)}$$

and the protocol-resource cost of sustaining the attack for duration $t(A_p)$ is $A_P \cdot N^* c \cdot t(A_P)$. The function is decreasing in A_p: a larger attacking share discovers blocks faster on the hidden chain, reducing the time to

overtake the honest chain. This follows from the Poisson structure of Nakamoto block production (Nakamoto, 2008; Budish, 2025). H_{total} is held fixed at the point of comparison; the model analyzes the attacker's profitability conditional on the current network state rather than endogenising aggregate hashrate response. This is a maintained simplification. During a hidden double-spend attack, honest miners continue to mine on the public chain and earn block rewards at the prevailing rate; they have no information that an attack is underway until the attacker reveals the hidden chain. Honest miners therefore have no reason to exit during the covert phase. Once the attacker reveals, the attack is either successful (and complete) or fails; in either case, the attack duration is determined before honest miners can respond. The simplification therefore does not bias the analysis during the attack phase. If anything, endogenising H_{total} would strengthen deterrence: after a revealed attack, honest miners attracted by temporarily elevated expected rewards (from the attacker's impaired capacity) would enter, raising H_{total} and reducing the attacker's share in any subsequent attempt. In Budish's benchmark, where the attacker is a monolithic outsider that can acquire arbitrary hash, A_p reduces to the cost-minimising attacking share A^*. In pooled mining, A_p is constrained by effective retained hash, which is endogenous to member exit.

The pool's expected net gain from attack is:

$$\begin{aligned} G(V_{tx}) = \quad & V_R(V_{tx}) - A_P \cdot N^* c \cdot t(A_P) - \Pi_{\text{law}}(V_{tx}, \theta_P) \\ & - L_{\text{rep}}(V_{tx}, \theta_P) - L_{\text{cap}}(V_{tx}, \theta_P) \\ & - L_{\text{excl}}(a, \sigma) - L_{\text{net}}(V_{tx}, \theta_P). \end{aligned} \quad \text{...(28)}$$

Attack is profitable if and only if $G(V_{tx}) > 0$, and deterred if $G(V_{tx}) \leq 0$.

Equation (28) is the paper's central extension. It preserves Budish's resource-cost logic but embeds it in a legal, organizational, and operational environment that becomes increasingly relevant as transaction value rises.

3.9. Comparative Statics

The model yields the following comparative statics. As transaction value rises, legal and reputational losses become more material over the relevant range:

$$\frac{\partial \Pi_{\text{law}}}{\partial V_{tx}} > 0, \qquad \frac{\partial L_{\text{rep}}}{\partial V_{tx}} > 0, \qquad \frac{\partial L_{\text{cap}}}{\partial V_{tx}} > 0 \quad \text{...(29)}$$

Through the direct legal and loss channels, holding effective pool hash H_P^{eff} fixed, stronger detection, attribution, and sanction reduce attack profitability:

$$\left.\frac{\partial G}{\partial p_d}\right|_{H_P^{\text{eff}}} < 0, \qquad \left.\frac{\partial G}{\partial p_a}\right|_{H_P^{\text{eff}}} < 0, \qquad \left.\frac{\partial G}{\partial p_s}\right|_{H_P^{\text{eff}}} < 0 \quad \text{...(30)}$$

The conditioning on H_P^{eff} is needed because these parameters also enter the member-exit channel: higher p_d, for instance, raises λ_i, raises q_i, and lowers H_P^{eff}, which reduces the protocol-cost term $A_P \cdot N^* c \cdot t(A_P)$ in Equation (28). That indirect effect is not a threat to deterrence. If effective pool hash falls far enough that A_p drops below the threshold for a feasible majority attack, the attack is prevented entirely—a stronger form of deterrence than reduced profitability. The direct and indirect channels therefore reinforce each other: the direct channel makes the attack less profitable, while the indirect channel can make it infeasible.

Similarly, through the direct channels, greater organizational visibility, jurisdictional reach, and counterparty dependence increase deterrence:

$$\left.\frac{\partial G}{\partial j_P}\right|_{H_P^{\text{eff}}} < 0, \qquad \left.\frac{\partial G}{\partial o_P}\right|_{H_P^{\text{eff}}} < 0, \qquad \left.\frac{\partial G}{\partial x_P}\right|_{H_P^{\text{eff}}} < 0 \quad \text{...(31)}$$

Greater capital specificity increases expected attack loss:

$$\frac{\partial G}{\partial k_P} < 0 \qquad \text{...(32)}$$

Greater network and state-action exposure reduce profitability both directly and through member exit:

$$\frac{\partial G}{\partial n_P} < 0, \qquad \frac{\partial G}{\partial s_P} < 0 \qquad \text{...(33)}$$

These comparative statics generate the paper's central implication: Budish's no-law benchmark is strongest at the low-value edge and weakens as transaction value, actor legibility, and off-chain dependence rise.

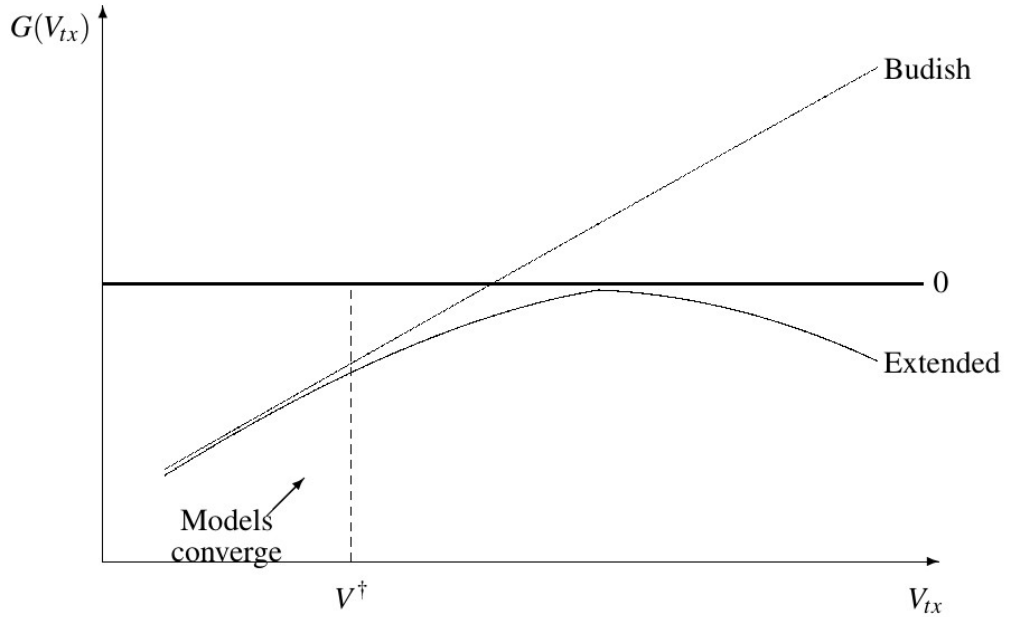

Figure 1: Schematic Attack Profitability Under the Budish Baseline and the Extended Model

Note: Horizontal axis: Nominal transaction value V_{tx}. Vertical axis: pool expected net gain $G(V_{tx})$. Under the Budish baseline, G rises with V_{tx}. Under the extended model, G is compressed at higher values by legal sanctions, reputational loss, capital impairment, member exit, and longest-chain exclusion. Below $V^\dagger$, where enforcement cost dominates expected recovery, the models coincide. The figure is illustrative and does not represent calibrated values.

3.10. Propositions

Proposition 1 (Budish is Locally Correct for Trivial Transaction Values): *For sufficiently small V_{tx}, enforcement cost dominates expected sanction, so $\Pi_{\text{law}}(V_{tx}, \theta_P) = 0$ by the enforcer participation constraint in Equation (6). The pure protocol-cost benchmark then applies.*

Proof: By Equation (6), $\Pi_{\text{law}} = \max\{0, p_d p_a p_s S - C_E\}$. Since $S \leq \bar{S}(V_{tx})$ for some upper bound that is finite for finite V_{tx}, and $C_E > 0$, there exists $V^\dagger > 0$ defined by $p_d p_a p_s S(V^\dagger) = C_E$ such that for all $V_{tx} < V^\dagger$, $p_d p_a p_s S < C_E$ and hence $\Pi_{\text{law}} = 0$. With $\Pi_{\text{law}} = 0$: by Restriction 1, $L_{\text{rep}} = \rho \cdot K_P^F$ where ρ is increasing in V_{tx}. At $V_{tx} = 0$ there is no attack, hence no reputational loss; therefore $\rho(0, \theta_p) = 0$ by definition, and $L_{\text{rep}} \to 0$ as $V_{tx} \to 0$. By Lemma 3, $L_{\text{rep}} = \chi \cdot K_P^{\text{ASIC}}$ where χ is increasing in V_{tx}. At $V_{tx} = 0$ there is no attack, hence no capital impairment; therefore $\chi(0, \theta_p) = 0$ by definition, and $L_{\text{cap}} \to 0$ as $V_{tx} \to 0$. L_{excl} and L_{net} depend on V_{tx} through enforcement response intensity: when $V_{tx} < V^\dagger$, no rational enforcer initiates proceedings, so there is no public identification trigger for longest-chain orphaning or network-level suppression, and both terms are zero. In this regime, $V_R = \left(1 - \phi_f(V_{tx}) - \phi_r(V_{tx}) - \phi_c(V_{tx})\right) V_{tx}$. The freeze, recovery, and compliance frictions are themselves functions of V_{tx}: at $V_{tx} = 0$ there is no transaction to freeze, recover, or monetise, so $\phi_f(0) = \phi_r(0) = \phi_c(0) = 0$ and $V_R = V_{tx}$. For small positive V_{tx}, the frictions remain small because the institutional response (exchange

alerts, counterparty scrutiny) is not triggered for trivial amounts. Equation (28) then reduces to $G(V_{tx}) = V_{tx} - A_P \cdot N^* c \cdot t(A_P)$, which reproduces Budish's deterrence condition up to the pool-structure term in A_p.

Throughput and the Low-Value Regime: Proposition 1 is not merely a concession to Budish. It has a positive economic implication. Budish's deterrence condition compares attack cost to V_{tx} for a single double-spend opportunity. It does not say the system is uneconomic in aggregate. It says each transaction must be secured by protocol cost proportional to its own value. For small transactions, that cost is small.

If transaction volume scales, the per-transaction share of p_{block} falls. A system processing N transactions per block has a per-transaction security cost of p_{block}/N. If N is sufficiently large and individual transaction values are small—each below the threshold at which enforcement becomes economic—then the per-transaction security cost is low, and no individual transaction justifies legal process. Each of those transactions falls within the regime where $\Pi_{\text{law}} = 0$ and Budish's protocol-only model applies. The model therefore permits large aggregate turnover through many low-value payments, all of which remain in the protocol-only security regime.

The paper's extension therefore covers the complementary regime. For the subset of transactions whose individual value is large enough to make enforcement economic, the extended model applies and the deterrence condition widens. The system thus has two security regimes operating simultaneously: protocol-only for small payments, protocol-plus-law for large payments. Budish's benchmark describes the first regime. The present model describes the second. Neither displaces the other.

Proposition 2 (Attribution Runs First to the Pool Operator): *In pooled mining, the first-order coordinating actor for a deliberate double spend is the pool operator or equivalent block-building entity, not the ordinary contributing hash provider.*

Proof: By equation (3), nominal pool hash $H_P = h_P^{\text{own}} + \sum_i \alpha_i h_i$. In pooled mining under the pool-operator architecture described by Cong *et al.* (2021), the pool operator constructs the block template, selects transactions, determines which chain tip to extend, and propagates completed blocks to the network. Contributing providers submit proof-of-work shares against the template provided by the pool; they do not independently select transactions, construct block headers, or choose which chain to build on. A provider who multihomes across pools $\{P_1, \ldots, P_k\}$ submits shares to each pool's template but does not coordinate block production for any of them: the decision to build an alternative (attacking) chain is the pool operator's decision, implemented through the template distributed to providers. First-order attribution for a deliberate double spend therefore runs to the entity that constructs and propagates the attacking chain—the pool operator—not to the provider whose hash is an input to the pool's aggregate capacity. The provider's contribution is necessary for the pool's capacity but not sufficient for the attack decision.

Proposition 3 (Pooled Attack Capacity is Unstable through Two Exit Channels): *As payout distortion rises, or as liability and capital-risk terms rise, contributor exit probabilities increase:*

$$\delta_i \uparrow \Rightarrow \ \ell_i \uparrow \Rightarrow \ q_i \uparrow$$

$$(p_d, p_a, p_s, \omega_i, \eta_i) \uparrow \Rightarrow \ (\lambda_i, z_i) \uparrow \Rightarrow \ q_i \uparrow$$

Hence effective retained pool hash falls: $q_i \uparrow \Rightarrow \ H_P^{\text{eff}} \downarrow$.

Proof

Channel A: By the definition of ℓ_i (Equation 11), $\partial \ell_i / \partial \delta_i > 0$. By Equation (8), $U_i^{\text{stay}} = \pi_i - c_i - \ell_i - \lambda_i - z_i$. An increase in ℓ_i reduces U_i^{stay} by $\partial U_i^{\text{stay}} / \partial \ell_i = -1 < 0$. The outside option $U_i^{\text{switch}} = \pi_i^{\text{alt}} - c_i^{\text{alt}}$ (Equation 9) does not depend on ℓ_i. By Equation (15), $q_i = \Pr\left(U_i^{\text{stay}} < U_i^{\text{switch}}\right)$. Since U_i^{stay} is decreasing in ℓ_i and U_i^{switch} is invariant, the probability that $U_i^{\text{stay}} < U_i^{\text{switch}}$ is increasing in ℓ_i, hence increasing in δ_i. Therefore $\delta_i \uparrow \Rightarrow \ell_i \uparrow \Rightarrow q_i \uparrow$.

Channel B: By Lemma 2 (Equation 14), $\partial\lambda_i/\partial p_d > 0$, $\partial\lambda_i/\partial p_a > 0$, $\partial\lambda_i/\partial p_s > 0$, $\partial z_i/\partial\omega_i > 0$, $\partial z_i/\partial\eta_i > 0$. By Equation (8), $\partial U_i^{\text{stay}}/\partial\lambda_i = -1 < 0$ and $\partial U_i^{\text{stay}}/\partial z_i = -1 < 0$. The outside option is again invariant. By the same argument as Channel A, q_i is increasing in λ_i and z_i, hence increasing in (p_d, p_a, p_s, ω_i, η_i).

Hash Erosion: By Equation (17), $H_P^{\text{eff}} = h_P^{\text{own}} + \sum_i (1-q_i)\,\alpha_i h_i$. Since $\partial(1-q_i)/\partial q_i = -1 < 0$ and $\alpha_i h_i \geq 0$, H_P^{eff} is strictly decreasing in each q_i for all i with $\alpha_i h_i > 0$.

Proposition 4 (Large-Value Attacks are Less Profitable than the Pure Flow-Cost Model Implies): *As rises, realised attack value remains discounted by monetisation frictions (equation 4), legal sanctions net of enforcement cost become positive and increasing (Lemma 1), reputational loss rises (Restriction 1), ASIC-capital risk rises (Lemma 3), and longest-chain and network exclusion become more effective. Hence the relevant deterrence condition is not (2) alone but* $G(V_{tx}) \leq 0$.

Proof: By Equation (28), $G(V_{tx}) = V_R(V_{tx}) - A_P N^* c\, t(A_P) - \Pi_{\text{law}} - L_{\text{rep}} - L_{\text{cap}} - L_{\text{excl}} - L_{\text{net}}$.

Step 1-Gain Term: By Equation (4), $V_R = \left(1 - \phi_f - \phi_r - \phi_c\right)V_{tx}$. For $\phi_f, \phi_r, \phi_c > 0$, $V_R < V_{tx}$. The gain is strictly smaller than under Budish's benchmark.

Step 2-Legal Term: By Lemma 1, $\partial\Pi_{\text{law}}/\partial V_{tx} > 0$ for $V_{tx} > V^\dagger$. Hence Π_{law} is zero for $V_{tx} \leq V^\dagger$ and strictly positive and increasing for $V_{tx} > V^\dagger$.

Step 3-Reputational Term: By Restriction 1, $\partial\rho/\partial V_{tx} > 0$, so $L_{\text{rep}} = \rho \cdot K_P^F$ is increasing in V_{tx} for $K_P^F > 0$.

Step 4-Capital Term: By Lemma 3, $\partial\chi/\partial V_{tx} > 0$, so $L_{\text{cap}} = \chi \cdot K_P^{\text{ASIC}}$ is increasing in V_{tx} for $K_P^{\text{ASIC}} > 0$.

Step 5-Exclusion and Network Terms: L_{excl} and L_{net} are non-negative by construction (Equations 25 and 26).

Step 6-Indirect Channel: By Proposition 3, the legal and capital-risk parameters that enter Π_{law} and L_{cap} also raise exit probabilities q_i, which reduces H_P^{eff} and hence $A_P = H_P^{\text{eff}}/H_{\text{total}}$. This reduces the protocol-cost term $A_P N^* c\, t(A_P)$ but also reduces the attacker's ability to sustain a majority chain. If A_P falls below the threshold for a feasible majority attack, the attack is prevented entirely.

Conclusion: In Budish's benchmark, $G^B(V_{tx}) = V_{tx} - A^* N^* c\, t(A^*)$. In the extended model, $(V_{tx}) < G^B(V_{tx})$ for all $V_{tx} > V^\dagger$, because: (i) the gain is reduced ($V_R < V_{tx}$), (ii) five additional non-negative loss terms are subtracted, at least two of which (Π_{law}, L_{rep}) are strictly positive and increasing for $V_{tx} > V^\dagger$, and (iii) the indirect member-exit channel can render the attack infeasible. The deterrence condition is therefore $G(V_{tx}) \leq 0$, which is strictly weaker (easier to satisfy) than Budish's $G^B(V_{tx}) \leq 0$ for all $V_{tx} > V^\dagger$.

3.11. Reversion to Budish

The model extends Budish along two dimensions: organizational structure (pooled mining with endogenous member exit) and rule of law (legal sanctions, reputational loss, capital impairment, longest-chain exclusion, and network friction). The rule-of-law extension is the paper's central contribution; the pool-structure extension is its necessary organizational premise, because the legal term cannot be assigned to an actor until the actor is specified. The two extensions nest cleanly:

Proposition 5 (Exact Reversion to Budish): *Setting* $\Pi_{\text{law}} = 0$, $L_{\text{rep}} = 0$, $L_{\text{cap}} = 0$, $L_{\text{excl}} = 0$, $L_{\text{net}} = 0$, $\phi_f = \phi_r = \phi_c = 0$, $\lambda_i = 0$, $z_i = 0$, *and* $\ell_i = 0$ *for all i, Equation (28) reduces to Budish's Equation (2).*

Proof: The reduction proceeds in two steps.

Step 1-Eliminate Rule of Law: Set $\Pi_{\text{law}} = 0$, $L_{\text{rep}} = 0$, $L_{\text{cap}} = 0$, $L_{\text{excl}} = 0$, $L_{\text{net}} = 0$. Set all monetisation frictions to zero: $\phi_f = \phi_r = \phi_c = 0$, so $V_R = V_{tx}$. Set all liability and capital-risk exit terms to zero: $\lambda_i = 0$, $z_i = 0$ for all i. Equation (28) becomes

$$G(V_{tx}) = V_{tx} - A_P \cdot N^* c \cdot t(A_P)$$

with effective pool hash still given by $H_P^{\text{eff}} = h_P^{\text{own}} + \sum_i (1 - q_i)\,\alpha_i h_i$, where q_i now depends only on payout distortion ℓ_i. This is an intermediate case: pooled mining without law. The attacker faces protocol cost and endogenous member exit from worsening payouts, but no legal, reputational, or capital-destruction penalties. This intermediate model is itself a contribution: it shows that pooled attack capacity can be fragile even before law enters.

Step 2-Eliminate Pool Structure: Further set $\ell_i = 0$ for all i (no payout distortion, so contributors never observe a reason to exit via Channel A). With $\lambda_i = z_i = 0$ from Step 1 and $\ell_i = 0$, the stay payoff $U_i^{\text{stay}} = \pi_i - c_i$ is unaffected by the attack. Exit probability $q_i = 0$ for all i, so $H_P^{\text{eff}} = H_P$. The pool's capacity is static. Setting $H_p = A^* H_{\text{total}}$ (the pool controls exactly the cost-minimising attacking share), $A_p = A^*$ and the deterrence condition reduces to

$$A^* N^* c \cdot t(A^*) > V_{tx}$$

which is Budish's Equation (2). The same result obtains if one starts from a monolithic solo miner with no outside contributors ($n = 0$): then $H_P = h_P^{\text{own}}$, there are no contributors to exit, and $A_P = h_P^{\text{own}} / H_{\text{total}}$, which reduces to A^* in Budish's setup. Both routes yield Equation (2).

The model therefore contains Budish as the no-law, no-pool special case. It generalises in two nested layers: first to pooled organization with endogenous exit (an economic extension), then to rule-of-law exposure (the legal extension). Each layer adds terms that are zero at the relevant boundary and become positive as the institutional environment strengthens.

3.12. Interpretation of the Maintained Restrictions

The model contains two maintained restrictions (Restrictions 1 and 2) whose sign conditions are supported by cited empirical evidence but cannot be derived from the model's definitions alone. All other sign conditions (Lemmas 1, 2, and 3, and the properties of ℓ_i) are proved from the model's definitions and equations. The maintained restrictions are narrow in scope. The model does not require perfect enforcement, certain attribution, or universal legal reach. Restriction 1 requires only that reputational damage is increasing in attack severity and organizational visibility. Karpoff and Lott (1993) establish this for counterparty-facing fraud across 132 enforcement cases. Karpoff *et al.* (2005) find that the channel does *not* operate for environmental violations, where harm is diffuse and no individual counterparty bears a concentrated loss; in that setting, legal penalties alone carry the deterrence burden. Because a double-spend attack defrauds specific transaction recipients, it falls within the counterparty-fraud channel where the sign restrictions are empirically supported. Restriction 2 requires only that pools dependent on externally governed infrastructure face larger losses when that infrastructure is impaired—a finding established by Apostolaki *et al.* (2017) for Bitcoin specifically. Heterogeneous exposure is built directly into , , and the capital-specificity term.

As Proposition 5 establishes, Budish's theorem is recovered as the no-law, no-pool special case. The present paper's extension is therefore nested: for high-value double spends embedded in pooled organization, legally reachable institutions, specialised capital, mobile contributors, and operational dependencies, the economically relevant deterrence condition is wider than the pure no-rule-of-law benchmark.

3.13. Formal Attack Condition

This subsection derives the attacker's profitability condition once Budish's benchmark is embedded in pooled mining and extended to admit rule-of-law effects. The objective is not to replace the pure protocol-cost logic, but to show how it is altered when one admits four additional features of actual attack environments: legal sanctions net of enforcement cost, member exit induced by payout distortion and liability risk, impairment of specialised ASIC capital, and exclusion by an honest remainder.

3.13.1. From Budish's Inequality to a Pool-Centred Attack Condition

Budish's pure benchmark requires that the resource cost of attack exceed the nominal gain from double spending. In the present framework, the relevant prize is not the nominal transaction value but the value the attacker can actually realise after freezes, recovery, and monetisation frictions. The relevant cost of attack is not simply the cost of assembling nominal hash, because a pooled attacker does not control a fixed block of passive capacity. Its feasible attacking power depends on whether members continue to supply hash during and after the attack attempt.

The pool's expected net gain from attack is therefore given by Equation (28). Attack is profitable if and only if $G(V_{tx}) > 0$, and deterred if $G(V_{tx}) \leq 0$.

3.13.2. Legal Sanctions Net of Enforcement Cost

The first extension is the legal term, Equation (6). The max{0, .} operator ensures that if the expected cost of enforcement exceeds the expected recovery, no rational enforcer initiates proceedings and the attacker faces no legal risk (Polinsky and Shavell, 2007). At the low-value edge, $\Pi_{\text{law}} = 0$ and the pure no-law benchmark applies. Once V_{tx} rises, the gross magnitude of sanction and recovery rises while enforcement costs become relatively less important, and Π_{law} becomes positive.

Large attacks are increasingly exposed to legal responses that are uneconomic for trivial losses. This localises Budish's theorem: the pure benchmark is strongest where $\Pi_{\text{law}} = 0$ and weakens where Π_{law} becomes materially positive.

3.13.3. Payout-Distortion Exit

The second extension concerns the instability of pooled attacking capacity. In a solo-miner model, attack capacity is fixed once the attacker assembles sufficient hash. In pooled mining, that assumption is too strong because contributors can leave.

The first exit channel operates before misconduct is necessarily identified. Hidden attack activity may distort ordinary pool performance. Contributors may observe delayed payments, reduced payouts, abnormal variance, stale-share increases, or unexplained underperformance (Gervais *et al.*, 2016), even if they do not infer the precise cause. Contributor i's payoff from staying with the pool is Equation (8), while the outside option is Equation (9). Exit occurs when $U_i^{\text{stay}} < U_i^{\text{switch}}$.

This yields a direct implication: covert attacks can unravel even before the legal system responds, because ordinary contributors react to worsening economics. The pool does not need to be publicly exposed for its effective attacking capacity to begin shrinking. The exit condition (Equation 8 vs. Equation 9) binds once members observe that their payout falls below the outside option, regardless of attribution.

3.13.4. Liability-Driven and Capital-Risk Exit

The second exit channel opens once the pool is detected, suspected, or publicly associated with dishonest conduct. Continued participation is no longer a neutral commercial decision. It may expose contributors to legal scrutiny, exchange frictions, contractual termination, or the impairment of their specialised mining capital.

A member does not need to be the primary attacking actor to be affected by attack exposure. Once the pool becomes risky, contributors may rationally leave to preserve access to exchanges, counterparties, hosting, and the continued economic usefulness of their mining equipment. Together, payout-distortion and liability-driven exit imply that pooled attack capacity is not simply purchased once and for all. It must be held together under deteriorating economic and legal conditions.

3.13.5. ASIC-Capital Impairment

The third extension concerns the capital structure of mining. Bitcoin mining relies heavily on application-specific hardware. Exclusion from honest mining, or material impairment of monetisation and counterparties, can destroy more than current revenue. It can destroy a large share of the value of the mining operation's capital stock.

The pool-level capital loss, Equation (20), performs two functions. First, it captures the loss of value from sanctions, freezing, taint, or inability to continue profitable honest mining. Second, it converts part of the attacker's exposure from a flow-cost problem into a stock-loss problem. Budish's own benchmark recognises how effective such stock-based deterrence would be if it were available. The present model shows one route by which it can arise in practice: an organised pooled attacker that becomes excluded, impaired, or commercially toxic may lose a large fraction of the value embedded in its ASIC-intensive operation. In the current mining environment, the capital at stake in a large pool's coalition is not measured in days of forgone revenue. It is measured in the replacement cost of specialised hardware fleets, associated infrastructure, and the financing and contractual relationships that support them. The gap between temporary flow expenditure and permanent stock destruction is the gap between Budish's benchmark and the present model.

The same logic also reinforces contributor exit. If a member's hardware is economically tied to continued participation in honest mining, then continued support for a dishonest pool can threaten not just current payout but the future value of the hardware itself. For a contributor operating a fleet of Bitcoin ASICs with no alternative productive use, the relevant loss is not a few days of pool earnings; it is the economic value of the fleet as a going concern. This increases z_i, which increases q_i, which reduces H_P^{eff}.

3.13.6. Honest-Longest-Chain Exclusion

The fourth extension concerns the response of the honest remainder of the network. Budish's outsider-attacker framework is static in the sense that the honest side is treated as a background quantity of resistance. In the present model, the honest side does not need to become strategic. Honest miners continue following the unchanged protocol. What changes is the attacker's effective hash: member exit and connectivity degradation reduce it, and once it falls below the honest remainder, the protocol does the rest.

Equation (24) is the standard Nakamoto majority condition adjusted for connectivity degradation. Honest miners do not need to coordinate or change behavior: they continue mining on the longest valid chain under the unchanged protocol. If the attacker's effective hash falls below the honest remainder's hash—whether through member exit, connectivity degradation, or both—the attacker's chain is orphaned. The greater the infrastructure-layer suppression of the attacker, the lower the attacker's effective share and the more certain the protocol-level exclusion.

3.13.7. Main Implications

The formal attack condition yields four immediate implications.

First, Budish's benchmark remains strongest at the low-value edge. Where V_{tx} is trivial, $\Pi_{\text{law}} = 0$ by the enforcer participation constraint, reputational loss is small, and infrastructure-layer responses are not triggered. The system then behaves as the pure protocol-cost case.

Second, the profitability of a pooled attack is more fragile than a static outsider model suggests. Even before formal legal action, covert attack behaviour can lower member payouts and induce exit. Once suspicion or attribution emerges, liability and capital-risk exit reinforce this effect. Effective attacking hash is endogenous and can contract during the attack itself.

Third, capital specificity changes the deterrence margin materially. A pool or contributor with large stocks of Bitcoin-specific hardware is not merely risking current flow income. It risks the productive value of specialised capital. This turns part of the problem from one of temporary expenditure into one of lasting wealth destruction.

Fourth, once the attacker's effective hash is degraded—through member exit, connectivity suppression, or both—the unchanged Nakamoto protocol does the rest. Honest miners continue mining on the longest valid chain. If honest hash exceeds attacker effective hash, the attacker's chain is orphaned. The continuation value of attack can collapse before nominal on-chain gains are fully realized.

3.13.8. Attribution and Liability

A clarification is needed on the scope of the attribution claim. The model assigns first-order double-spend attribution to the pool operator because the pool is the entity that coordinates block production, receives and distributes rewards, and controls the decision to build an attacking chain. That does not exhaust the liability

surface. Secondary liability may extend to contributing hash providers depending on knowledge, collusion, facilitation, or continued participation after misconduct becomes apparent. The model captures that extension through λ_i and z_i: a contributor who continues to supply hash to a pool credibly suspected of dishonest conduct bears expected costs from legal exposure and capital-risk impairment. The distinction is between first-order attribution, which runs to the pool, and secondary exposure, which depends on each contributor's information, conduct, and exit decision. The model does not make every provider the primary attacker. It makes every provider a participant whose continued support is conditional on the costs of staying.

4. Enforcement and Calibration

4.1. Enforcement Channels

The model's terms correspond to enforcement channels documented in the empirical literature. Detailed discussion with full citations is in Appendix B; this subsection summarises the five main channels.

First, the gap between nominal and realised attack value ($V_R < V_{tx}$) reflects exchange freezes, recovery actions, and monetisation frictions documented by Makarov and Schoar (2021) and Cong *et al.* (2025). Second, public identification of a dishonest pool triggers reputational and franchise-value losses (L_{rep}) that extend beyond formal adjudication; Matsuzawa (2025) shows that visible enforcement deters a broader population than the individuals directly affected. Third, the pool's own coalition becomes unstable before any court judgment: members exit when payout deteriorates or legal risk rises, following the organizational logic of Thompson (2024). Fourth, honest-longest-chain exclusion (L_{excl}) operates through the longest-chain rule once the honest fraction's effective hash exceeds the attacker's degraded share. Fifth, operational dependence on naming, routing, and interconnection layers (L_{net}) creates an auxiliary enforcement margin: Apostolaki *et al.* (2017) show that routing-level intervention can isolate mining capacity and degrade effective hashrate.

Gulen and Myers (2024) provide direct evidence that enforcement is selective and politically contingent: Clean Water Act violation rates differ by 39.7% points between battleground and non-battleground states. The model requires only that expected enforcement is positive and increasing in transaction value, actor legibility, and capital exposure—conditions formalised in Equations (29)-(33) and documented by Gulen and Myers (2024) and Belnap *et al.* (2024). Each channel enters Equation (28) as a separate cost term. Any combination that makes $G(V_{tx}) \leq 0$ is sufficient for deterrence.

4.2. Calibration

The preceding sections derive the extended deterrence condition in general terms. This subsection asks whether the model's terms are quantitatively meaningful by calibrating three components using publicly available data: the enforcer participation threshold (Section 4.2.1), the capital-at-risk ratio (Section 4.2.2), and the dynamics of member exit (Section 4.2.3). The exercises are illustrative, not econometric. They use disclosed figures from SEC filings, DOJ enforcement settlements, and pool-level hashrate data to show that the model's extensions are not merely theoretical possibilities but correspond to magnitudes that matter for attack profitability.

4.2.1. Enforcer Participation Threshold

The enforcer participation constraint in Equation (6) states that the legal term Π_{law} is zero when enforcement cost C_E exceeds expected sanction recovery $p_d \cdot p_a \cdot S$. The threshold transaction value $V^\dagger$ at which Π_{law} first becomes positive is therefore the value at which a rational enforcer—victim, prosecutor, exchange, or regulator—finds it worthwhile to act.

Three inputs are required: the joint probability of detection and attribution at the pool coordination layer ($p_d \cdot p_a$), the probability and magnitude of sanction conditional on attribution ($p_s \cdot S$), and the cost of enforcement (C_E). Each can be bounded from observable data.

Detection and Attribution: Pool-level data (Appendix A) show that 98.22% of hashrate is attributable to pools with verifiable public operator identification through coinbase tags, corporate filings, and public websites. At the coordination layer, $p_d \cdot p_a$ is therefore close to unity for pool-mediated attacks. The relevant uncertainty is not whether the pool can be identified—it can—but whether the legal system will act.

Sanction Magnitude: Recent US enforcement actions against cryptocurrency intermediaries provide a revealed lower bound on for identified, legally reachable actors. Binance agreed to pay $4.3 billion in November 2023 to settle criminal charges with the DOJ, FinCEN, and OFAC.[2] BitMEX settled for $100 million in 2021. FTX and Alameda reached a $12.7 billion settlement with the CFTC in 2024. These figures reflect sanctions for operational misconduct (AML failures, sanctions violations), not double spending. A deliberate double-spend attack would expose the attacker to criminal liability under wire fraud (18 U.S.C. §1343), computer fraud (18 U.S.C. §1030), and conspiracy statutes, each of which carries penalties including imprisonment and forfeiture. The enforcement magnitudes observed for operational misconduct therefore represent a lower bound on the sanctions available for a deliberate attack.

Enforcement Cost: The Norton Rose Fulbright 2024 Litigation Trends Survey (Norton Rose Fulbright, 2024) reports that the median breach-of-contract case with $250,000 at stake costs $91,000-$145,000 to litigate through trial, implying enforcement costs of 36-58% of the disputed amount at that scale ($91,000/$250,000 = 36.4%; $145,000/$250,000 = 58.0%). At the 36-58% ratio, a $1 million case implies enforcement cost of $360,000-$580,000; a $5 million case implies $1.8-2.9 million. These ratios are upper bounds for the present setting because they reflect full private litigation through trial. For institutional plaintiffs (exchanges, custodians) with in-house legal capacity, and for criminal prosecutors with existing infrastructure, marginal enforcement cost per case is lower. Table 1 uses C_E values of $500,000-$2,000,000, spanning the range implied by the Norton Rose Fulbright ratio applied to the $1-5 million transaction range.

The model treats C_E as exogenous to the attacker's decision. In practice, enforcement cost depends on case complexity, available evidence, and the enforcer's existing capacity. Two observations bound the concern. First, the model's enforcer participation constraint already incorporates the key implication of variable C_E: enforcement is not invoked when C_E exceeds expected recovery, regardless of why C_E is high. Second, the calibration uses a range of C_E values ($500,000–$2,000,000) rather than a point estimate, and the threshold $V^\dagger$ varies monotonically in C_E (Table 1), so the sensitivity of the result to enforcement cost is transparent.

Threshold: Table 1 reports the implied enforcement threshold $V^\dagger$ under several parameter combinations, solving $p_d \cdot p_a \cdot p_s \cdot V^\dagger = C_E$ for $V^\dagger$.

Table 1: Enforcer Participation Threshold Under Alternative Parameter Values

$p_d \cdot p_a$	p_s	S/V_{tx}	C_E ($ million)	$V^\dagger$ ($ million)
0.95	0.50	1.0	0.5	1.05
0.95	0.50	1.0	1.0	2.11
0.95	0.50	1.0	2.0	4.21
0.95	0.30	1.0	1.0	3.51
0.80	0.50	1.0	1.0	2.50
0.95	0.50	2.0	1.0	1.05

Notes: $V^\dagger = C_E/(p_d \cdot p_a \cdot p_s \cdot S/V_{tx})$. Row 1: baseline with low enforcement cost. Row 2: baseline. Row 3: high enforcement cost. Row 4: lower sanction probability. Row 5: reduced attribution (e.g., opaque pool). Row 6: sanction exceeds attack value (as observed in Binance $4.3B and FTX $12.7B settlements).

The threshold ranges from $1.05 million to $4.21 million across the scenarios. Even under the parameter values most favourable to the attacker (row 3: C_E = $2 million, $p_s = 0.5$, $S = V_{tx}$), the threshold is below $5 million. Row 6 shows that if sanctions exceed the attack value—as they do in practice, given that Binance's $4.3 billion penalty exceeded any single transaction value—the threshold drops further. Below $V^\dagger$, Budish's protocol-only benchmark applies. Above it, the legal term becomes positive and attack profitability is reduced.

[2] United States Department of Justice, press release, 21 November 2023; OFAC Enforcement Release, 21 November 2023.

These estimates are upper bounds on $V^{\dagger}$ because they exclude criminal prosecution (which carries additional penalties beyond disgorgement), reputational and counterparty losses, and the threat of imprisonment. They also set S/V_{tx} at 1 or 2, whereas Binance's $4.3 billion penalty and FTX's $12.7 billion settlement both exceed the value of any single underlying transaction by orders of magnitude. Including any of these omitted channels would reduce $V^{\dagger}$ below the figures in Table 1.

4.2.2. Capital-at-Risk Ratio

The capital-loss term L_{cap} in Equation (20) depends on the stock of specialised ASIC capital tied to the pool's economic operation. Publicly disclosed data from a single listed mining firm illustrate the magnitude.

MARA Holdings (NASDAQ: MARA) reported in its 10-K for fiscal year 2024 that it operated approximately 400,000 mining rigs globally with an energised hashrate of 53.2 EH/s, held 44,893 bitcoin (valued at $4.19 billion at the 31 December 2024 closing price of $93,429 per BTC), and generated revenues of $656.4 million with energy and hosting costs of $381.6 million.[3] MARA Pool accounted for 5.71% of network hashrate during the sample period (237 of 4,149 blocks; Appendix A).

Table 2 reports the ratio of MARA's capital exposure to illustrative double-spend target values. The "ASIC fleet" column reports replacement cost at $3,000-$5,000 per rig. The "total capital exposure" column adds the bitcoin treasury. Hosting contracts, infrastructure, and financing commitments are excluded from both columns; their inclusion would increase the ratio.

Table 2: Capital-at-Risk Ratio: MARA Holdings Capital Exposure versus Illustrative Attack Targets

V_{tx} ($ million)	ASIC Fleet ($ billion)	BTC Treasury ($ billion)	Total Exposure ($ billion)	Ratio (Exposure/V_{tx})
1	1.2–2.0	4.19	5.4–6.2	5,400–6,200×
10	1.2–2.0	4.19	5.4–6.2	540–620×
50	1.2–2.0	4.19	5.4–6.2	108–124×
100	1.2–2.0	4.19	5.4–6.2	54–62×

Notes: ASIC fleet: 400,000 rigs at $3,000-$5,000 per unit (Bitmain Antminer S21 series retail pricing: $3,000 to $5,400 depending on model and efficiency tier, Q4 2024). BTC treasury: 44,893 BTC at $93,429 per BTC (CoinGecko closing price, 31 December 2024), yielding $4.19 billion. Total exposure excludes hosting contracts, infrastructure, and financing commitments; their inclusion would increase the ratio.

Source: MARA Holdings 10-K (FY2024)

MARA is a single firm with 5.71% of network hashrate. Table 2 shows that even for this single firm, the capital-at-risk ratio exceeds 54:1 for a $100 million attack and exceeds 540:1 for a $10 million attack. A pool coordinating 31% of hashrate (the scale of Foundry USA) aggregates contributed hash from multiple firms, each with its own capital exposure. The aggregate coalition capital cannot be computed without capital disclosures from all contributing operators, but it is necessarily larger than any single contributor's exposure.

The implication is direct. A pool operator contemplating a double spend faces capital destruction measured in billions against a prize measured in millions. The L_{cap} term in the model is not a marginal adjustment. For large, identified, ASIC-intensive pools, it dominates the attack payoff by orders of magnitude.

4.2.3. Member-Exit Dynamics

The model predicts that pooled attack capacity erodes endogenously through two exit channels: payout distortion (Channel A) and legal/capital-risk exposure (Channel B). This section illustrates the dynamics with a simple simulation.

[3] MARA Holdings, Inc., Annual Report on Form 10-K for the fiscal year ended December 31, 2024, filed March 3, 2025. BTC price: CoinGecko, 31 December 2024.

Setup: A pool with 252 EH/s (31.5% of network hashrate) diverts an increasing fraction of its hash to a hidden chain: from 5% initially to 25% after 24 hours. The network total of 800 EH/s corresponds to the approximate Bitcoin network hashrate in March 2026.[4] The simulation requires six parametric choices for which no direct empirical source exists; each is stated explicitly and tested in the robustness analysis below.

1. **Proprietary Hash Fraction:** Of the pool's 252 EH/s, 25 EH/s (9.9% of nominal, i.e., 25/252) is proprietary and the remainder is contributed. No pool publicly discloses its proprietary/contributed hash split. The 10% figure is a parametric choice. The qualitative result does not depend on it: the proprietary floor determines where the curves converge, not whether erosion occurs.
2. **Number of Contributing Providers:** Set at 200 with heterogeneous hash allocations drawn from a log-normal distribution. Axtell (2001) shows that US firm sizes follow a Zipf (power-law) distribution. The Zipf and log-normal are distinct distributional families, but both generate heavy right tails; the log-normal is used here as a tractable parametric choice that produces the skewed size heterogeneity documented by Axtell without imposing the infinite-variance property of a pure power law. The number 200 is a parametric choice; robustness to provider count is addressed through the sensitivity of exit dynamics to the distribution's variance, not the count.
3. **Exit Function:** Exit probability is a quadratic function of the observed payout shortfall, scaled by provider-specific sensitivity. The quadratic form ensures that small shortfalls produce low exit rates and large shortfalls produce high exit rates. This is a functional-form choice, not an empirical estimate.
4. **Detection Lag:** The legal-risk exit channel (Channel B) activates after 48 blocks (8 expected hours at 10 minutes per block). It is a parametric choice.
5. **Rolling Detection Window:** Providers observe a 24-block (4 expected hours) rolling average of payout performance. This is a parametric choice.
6. **Diversion Schedule:** Hash diversion ramps linearly from 5% to 25% over 24 hours. This is a parametric choice representing a gradual rather than instantaneous deviation.

5. Results

Figure 2 reports mean effective pool hash over a 48-hour attack window from 1,000 Monte Carlo replications, with shaded 90% intervals (5^{th}-95^{th} percentile). Under the Budish baseline, effective hash remains constant at 252 EH/s. Under Channel A alone, payout distortion drives mean contributed hash to 132.7 EH/s (52.6% of nominal) after 24 hours (90% interval: [109, 156] EH/s). Adding Channel B (legal and capital-risk exit) reduces mean effective hash to 84.4 EH/s (33.5%) at 24 hours (90% interval: [69, 102] EH/s). The full model, including network friction, brings mean effective hash to 67.5 EH/s (26.8%) at 24 hours (90% interval: [55, 82] EH/s). All three extended scenarios converge toward the proprietary-hash floor of 25 EH/s as contributed hash exits.

The simulation uses parametric functional forms and parameter values (enumerated above) rather than estimated coefficients. Its purpose is to show that the model's member-exit channels produce quantitatively meaningful capacity erosion over economically relevant time horizons, and that this result is robust to the choice of exit functional form (Table 3). Even a large initial pool share is reduced to a fraction of its nominal level within hours, not days, once payout distortion and legal exposure begin to bind.

Robustness: The qualitative result—substantial hash erosion within 24 hours—holds across functional-form alternatives and parameter variations. Table 3 reports results from 1,000 Monte Carlo replications under three exit functional forms (linear, quadratic, exponential), all under the full model. The linear form produces the most aggressive exit (mean 10.6% retained at 24 hours), while the quadratic (26.8%) and exponential (24.4%) produce similar results. Across all three forms, mean retained capacity at 24 hours falls well below the 50% threshold required for a majority attack.

[4] Mempool.space and Hashrate Index both report Bitcoin network hashrate in the range of 750-850 EH/s during February–March 2026.

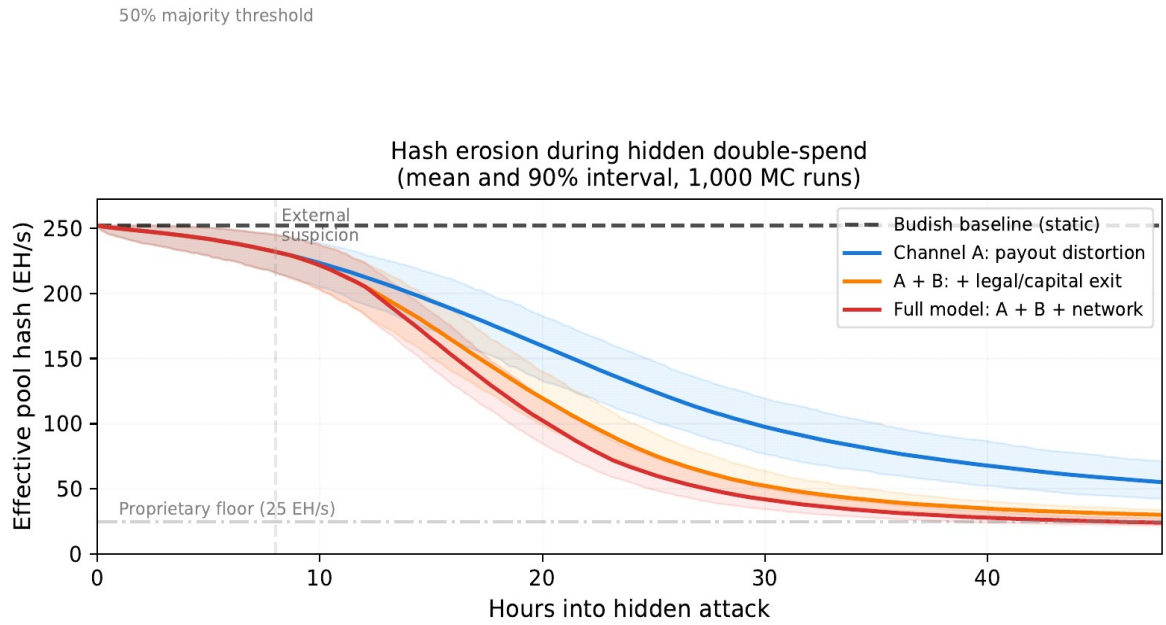

Figure 2: Simulated Effective Pool Hash During a Hidden Double-Spend Attack (Mean and 90% Interval from 1,000 Monte Carlo Replications). The Pool Begins with 252 EH/s (31.5% of Network) and Diverts Hash to a Hidden Chain at a Rate Rising from 5% to 25% Over 24 hours. The Budish Baseline Treats Capacity as Static. The Three Extended Scenarios Show Hash Erosion through Payout-Distortion Exit (Channel A), Legal/Capital-Risk Exit (Channel B), and Network-Friction Penalties. After 24 hours, Mean Effective Hash Under the Full Model is 26.8% of Nominal (90% CI: [21.8%, 32.5%]). Parameters: 200 Heterogeneous Contributors, 24-Block Rolling Detection Window, 48-Block Detection Lag for Channel B, Quadratic Exit Function

Table 3: Functional-Form Robustness: Retained Capacity at 24 hours (Full Model, 1,000 MC Runs)

Exit Function	Mean (EH/s)	Mean (% of Nominal)	90% CI (EH/s)
Linear	26.8	10.6	[22, 33]
Quadratic	67.5	26.8	[55, 82]
Exponential	61.5	24.4	[49, 76]

The functional-form comparison in Table 3 provides the primary robustness test: across all three exit specifications, mean retained capacity at 24 hours falls between 10.6% and 26.8% of nominal, well below majority-attack feasibility in every case.

The policy implication is that Budish's static-capacity assumption understates the fragility of pooled attack capacity once institutional and organizational constraints enter.

6. Conclusion

This paper has developed a narrow response to Budish's (2025) critique of permissionless consensus. It does not reject the theorem Budish proves. It accepts that theorem as a result for a pure benchmark in which deterrence must arise from protocol-side resource expenditure alone. The paper's claim is instead that this benchmark is locally accurate but globally incomplete. For trivial transaction values, the enforcer participation constraint binds (Equation 6), the legal term is zero, and Budish's protocol-only condition applies. For economically significant double-spend opportunities executed through pooled mining, the attacker's payoff is no longer determined by protocol-resource cost alone.

The reason is institutional rather than rhetorical. Mining is organised through pools, managers, payout rules, and mobile contributed hash rather than a mass of isolated autonomous miners. In that environment, the relevant first-order actor for a deliberate pooled attack is the coordinating pool entity. That entity's expected gain from attack is reduced not only by the direct cost of sustaining attacking hash, but also by legal sanctions net of enforcement cost, exchange and recovery frictions, reputational and franchise-value loss, the impairment of specialised ASIC capital, and the endogenous erosion of pooled attacking capacity through contributor exit. Once the attacker's effective hash falls below the honest remainder, the unchanged Nakamoto protocol orphans the attacker's chain; naming, routing, and interconnection dependencies can further degrade the attacker's effective share.

The paper makes four claims. First, for pooled mining, the relevant coordinating actor in a deliberate double spend is the pool operator or equivalent block-building entity, not the ordinary hash provider. Pool operators are identifiable at the coordination layer: pool-level data for March 2026 show that the coordination layer is concentrated among publicly attributable pool operators in a small number of jurisdictions, with less than 2% of hashrate attributable to pools without verifiable public operator attribution (Appendix A). Second, pooled attack capacity is endogenous rather than static. Hidden attack activity distorts payouts and induces member exit on ordinary economic grounds; detected or suspected misconduct induces further exit through liability and capital-risk channels. Third, the attacker's expected off-chain losses include not only legal sanctions net of enforcement cost, but also reputational destruction, the stranding of specialised ASIC capital whose replacement cost for a large pool's coalition can reach nine or ten figures, and the possibility of protocol-level exclusion once attacker effective hash falls below the honest remainder. An attacking pool is not risking a few days of forgone revenue; it is risking the productive value of a capital base that has no profitable alternative deployment. Fourth, the enforcement surface for a large attack extends beyond courts and prosecutors to exchange intervention, counterparty exclusion, reputational sanction, and operational disruption of naming, routing, and interconnection dependencies.

These claims are not independent. They form an interlocking structure. Identification enables attribution. Attribution enables legal sanction. Legal sanction triggers reputational and counterparty consequences. Those consequences induce member exit. Member exit reduces effective attacking hash. Reduced effective hash increases the honest remainder's ability to exclude the attacker. The no-law benchmark breaks all of these links by removing identification at the first step.

The calibration exercises in Section 4.2 show that these channels are quantitatively meaningful. The enforcer participation constraint becomes non-binding at transaction values above $1.05-$4.21 million depending on parameter values (Table 1). For a single listed mining firm (MARA Holdings), capital exposure exceeds a $10 million attack target by 540-620 to 1 (Table 2). A simulation of member exit under a hidden attack (1,000 Monte Carlo replications) shows that a pool beginning with 31.5% of network hashrate retains a mean of 26.8% of its nominal capacity after 24 hours under the quadratic exit specification (90% interval: [21.8%, 32.5%]), with qualitatively similar results under linear and exponential alternatives (Table 3). These magnitudes establish that the model's extensions are not marginal corrections. For large, identified, ASIC-intensive pools, the off-chain terms dominate the attack payoff.

The paper does not claim that legal institutions make Bitcoin safe, that all mining is easily attributable, or that enforcement is perfect. Nor does it claim that permissionless consensus intrinsically requires law. For small-value transactions processed at high volume, the system can operate within Budish's protocol-only regime without legal recourse. The maintained restrictions are narrow: reputational loss is increasing in attack severity and organizational visibility for counterparty-facing misconduct (supported by Karpoff and Lott, 1993); where harm is diffuse and no counterparty bears a concentrated loss, Karpoff *et al.* (2005) find that reputational penalties are negligible and legal penalties alone carry the deterrence burden. That divergence reinforces the present paper's thesis: for the class of misconduct at issue—double-spend attacks that defraud identifiable transaction recipients—reputational loss operates through the counterparty channel, and where it does not, legal sanctions must fill the gap. Operational losses are increasing in network dependence (supported by Apostolaki *et al.*, 2017). All other sign conditions are derived from the model's definitions. Bloch and Kranton (2024) emphasise that the dynamics of concealment and investigation are jointly determined: escalating

penalties are needed to offset the rational incentive to cover up, and enforcement effectiveness depends on the informativeness of the evidence available to investigators. The paper does not reject Budish's theorem. It localises it. Budish's deterrence condition applies where the enforcer participation constraint binds. For economically significant transactions executed through pooled mining, the relevant deterrence condition is jointly determined by protocol costs, legal sanctions, reputational loss, member mobility, capital destruction, and operational exclusion.

The core implication is that proof-of-work security at scale is best understood as a hybrid deterrence environment. Protocol economics remain central, but they do not operate in isolation once transactions are large enough to implicate real victims, off-ramp institutions, visible organizations, and operational dependencies. Budish's no-rule-of-law condition therefore remains a useful lower-bound benchmark for the low-value edge of the system, but it ceases to describe the full deterrence condition for large-value pooled attacks. A system processing large volumes of small transactions operates within Budish's benchmark: the per-transaction security cost is low, no individual payment justifies enforcement, and aggregate turnover can be large under protocol-only deterrence. The present model covers the complementary regime where individual transaction values are large enough to make legal enforcement economic.

Seen in this light, the economic limits of proof-of-work are not exhausted by the cost of assembling hash. They also depend on whether attackers can preserve the legal, commercial, and operational environment necessary to convert nominal on-chain gains into durable off-chain value. For small transactions, the enforcer participation constraint binds, the legal term is zero, and the attacker retains gains. For large transactions, the off-chain response set documented in Section 4 reduces realised value. That is the sense in which rule of law does not displace Budish's model, but materially changes its domain of application.

The broader contribution of the paper is therefore methodological. It suggests that the relevant contrast is not between a lawless blockchain and a law-governed outside world, but between alternative bundles of protocol incentives, organizational structure, legal sanction, and infrastructure dependence. Once the problem is framed in those terms, the economics of double spending are no longer captured by a pure flow-cost inequality alone. They are captured by a value-dependent attack condition in which protocol cost, legal exposure, member mobility, capital specificity, and exclusion dynamics jointly determine whether attack is profitable. That is the deterrence condition that matters for proof-of-work at scale.

References

Andrew Belnap, Jacob Thornock and Braden Williams. (2024). Hidden Wealth and Automatic Information Sharing. *Journal of Law and Economics*, 67(4), 905-949. doi: 10.1086/731738.

Arthur Gervais, Ghassan O. Karame, Karl Wüst, Vasileios Glykantzis, Hubert Ritzdorf and Srdjan Capkun. (2016). On the Security and Performance of Proof of Work Blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 3-16, ACM. doi: 10.1145/2976749.2978341.

Bruno Biais, Christophe Bisière, Matthieu Bouvard and Catherine Casamatta. (2019). The Blockchain Folk Theorem. *Review of Financial Studies*, 32(5), 1662-1715. doi: 10.1093/rfs/hhy095.

Craig S. Wright. (2025). Bitcoin Protocol Analysis: Game Theory and Protocol Integrity. Ph.D. Thesis, University of Exeter.

Craig S. Wright. (2026). Sunk Capital and Repeated Interaction in Nakamoto Consensus. *Economics Letters*, 249, 112940. doi: 10.1016/j.econlet.2026.112940.

Eric Budish. (2025). Trust at Scale: The Economic Limits of Cryptocurrencies and Blockchains. *Quarterly Journal of Economics*, 140(1), 1-62. doi: 10.1093/qje/qjae033.

Florian Deuflhard and Philipp Heller, C. (2023). Antitrust Economics of Cryptocurrency Mining. *Journal of Competition Law & Economics*, 19(2), 333-356. doi: 10.1093/joclec/nhad006.

Francis Bloch and Rachel Kranton. (2024). Cover-Ups. *Journal of Law and Economics*, 67(2), 337-369. doi: 10.1086/727849.

Gary S. Becker. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2), 169-217. doi: 10.1086/259394.

Haag Jr. William A., Douglas C. Montgomery, Allen Y. Tan and William C. Barker. (2019). Protecting the Integrity of Internet Routing: Border Gateway Protocol (BGP) Route Origin Validation. Special Publication 1800-14, National Institute of Standards and Technology.

Hansjörg Albrecher, Dina Finger and Pierre-Olivier Goffard. (2025). Empirical Risk Analysis of Mining a Proof-of-Work Blockchain. *Decisions in Economics and Finance*, 48, 1481-1508. doi: 10.1007/s10203-024-00458-w.

Hany F. Atlam, Ndifon Ekuri, Muhammad Ajmal Azad and Harjinder Singh Lallie. (2024). Blockchain Forensics: A Systematic Literature Review of Techniques, Applications, Challenges, and Future Directions. *Electronics*, 13(17), 3568. doi: 10.3390/electronics13173568.

Henry A. Thompson. (2024). The Industrial Organization of the Mafia. *Journal of Law and Economics*, 67(3), 555-587. doi: 10.1086/727850.

Hossein Nabilou. (2019). How to Regulate Bitcoin? Decentralized Regulation for a Decentralized Cryptocurrency. *International Journal of Law and Information Technology*, 27(3), 266-291. doi: 10.1093/ijlit/eaz008.

Huseyin Gulen and Brett W. Myers. (2024). The Selective Enforcement of Government Regulations: Battleground States, State Regulators, and the Environmental Protection Agency. *Journal of Law and Economics*, 67(1), 225-263. doi: 10.1086/728369.

Igor Makarov and Antoinette Schoar. (2021). Blockchain Analysis of the Bitcoin Market. Working Paper 29396, National Bureau of Economic Research.

Internet Assigned Numbers Authority. (2025). Root Zone Management. Available at https://www.iana.org/domains/root

Jonathan M. Karpoff and Jr. Lott, John R. (1993). The Reputational Penalty Firms Bear from Committing Criminal Fraud. *Journal of Law and Economics*, 36(2), 757-802. doi: 10.1086/467297.

Jonathan M. Karpoff, Jr. Lott John R. and Eric W. Wehrly. (2005). The Reputational Penalties for Environmental Violations: Empirical Evidence. *Journal of Law and Economics*, 48(2), 653-675. doi: 10.1086/430806.

Joseph Abadi and Markus K. Brunnermeier. (2018). Blockchain Economics. Working Paper 25407, National Bureau of Economic Research.

Kevin Werbach (2018). The Blockchain and the New Aechitecture of Trust. MIT Press, Cambridge, MA. ISBN: 9780262038935.

Konstantinos Stylianou and Nic Carter. (2020). The Size of the Crypto Economy: Calculating Market Shares of Cryptoassets, Exchanges, and Mining Pools. *Journal of Competition Law & Economics*, 16(4), 511-551. doi: 10.1093/joclec/nhaa016.

Kyutaro Matsuzawa. (2025). The Deterrent Effect of Targeted and Salient Police Enforcement: Evidence from Bans on Checkpoints for Driving Under the Influence. *Journal of Law and Economics*, 68(2), 311-360. doi: 10.1086/732666.

Lin William Cong, Campbell R. Harvey, Daniel Rabetti and Zong-Yu Wu. (2025). An Anatomy of Crypto-Enabled Cybercrimes. *Management Science*, 71(4), 3622-3633. doi: 10.1287/mnsc.2023.03691.

Lin William Cong, Zhiguo He and Jiasun Li. (2021). Decentralized Mining in Centralized Pools. *Review of Financial Studies*, 34(3), 1191-1235. doi: 10.1093/rfs/hhaa040.

Maria Apostolaki, Aviv Zohar and Laurent Vanbever. (2017). Hijacking Bitcoin: Routing Attacks on Cryptocurrencies. In *2017 IEEE Symposium on Security and Privacy*, 375-392. doi: 10.1109/SP.2017.29.

Michael Sockin and Wei Xiong. (2020). A Model of Cryptocurrencies. Working Paper 26816, National Bureau of Economic Research.

Mitchell Polinsky, A. and Steven Shavell. (2007). The Theory of Public Enforcement of Law. In Mitchell Polinsky, A. and Steven Shavell (Editors), *Handbook of Law and Economics*, 1, 403-454, Elsevier. doi: 10.1016/S1574-0730(07)01006-7.

Nick Arnosti and Matthew Weinberg, S. (2022). Bitcoin: A Natural Oligopoly. *Management Science*, 68(7), 4755-4771. doi: 10.1287/mnsc.2021.4095.

Norton Rose Fulbright. (2024). 2024 Annual Litigation Trends Survey. Survey Report, Norton Rose Fulbright LLP.

Panagiotis Chatzigiannis, Foteini Baldimtsi, Igor Griva and Jiasun Li. (2022). Diversification Across Mining Pools. *Journal of Cybersecurity*, 8(1), tyab027. doi: 10.1093/cybsec/tyab027.

Plinio Limata. (2019). Blockchains' Twilight Zones: A Reasoned Literature Review for a Critical Primer. Working Paper 76, Econometica.

Primavera De Filippi, Morshed Mannan and Wessel Reijers. (2022). The Alegality of Blockchain Technology. *Policy and Society*, 41(3), 358-372. doi: 10.1093/polsoc/puac006.

Raphael Auer. (2019). Beyond the Doomsday Economics of "Proof-of-Work" in Cryptocurrencies. Working Papers 765, Bank for International Settlements.

Robert L. Axtell. (2001). Zipf Distribution of US Firm Sizes. *Science*, 293(5536), 1818-1820. doi: 10.1126/science.1062081.

Satoshi Nakamoto. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Available at https://bitcoin.org/bitcoin.pdf, 2008. Accessed 13 March.

Sean Foley, Jonathan R. Karlsen and Talis J. Putninš. (2019). Sex, Drugs, and Bitcoin: How Much Illegal Activity is Financed through Cryptocurrencies?. *Review of Financial Studies*, 32(5), 1798-1853. doi: 10.1093/rfs/hhz015.

Appendix

A. Pool-Level Data

This appendix documents the pool-level hashrate data used in the paper. The data cover 4,149 blocks mined during the 30-day period ending 13 March 2026, as recorded by mempool.space. Pool identification is based on coinbase transaction tags—strings embedded in each block's generation transaction that identify the mining pool. This identification method is used by Cong *et al.* (2021), Arnosti and Weinberg (2022) and Stylianou and Carter (2020).

Classification Rule

Where the data source reports a country of origin, it is used directly. Where country of origin is listed as "unknown," the pool's jurisdiction is determined from corporate filings, website disclosures, and public reporting. F2Pool is registered in Singapore (co-founded by Chun Wang). MARA Pool is operated by MARA Holdings, Inc., listed on NASDAQ (ticker: MARA), headquartered in Fort Lauderdale, Florida. SBI Crypto is a subsidiary of SBI Holdings, listed on the Tokyo Stock Exchange (TSE: 8473). OCEAN is a US-based pool. NiceHash is registered in Slovenia. Pools for which no verifiable corporate registration, disclosed operator, or public reporting could be identified are classified as "Unknown."

The classification is strictly by operator jurisdiction for the coordination layer: the jurisdiction of the entity that receives block rewards and distributes payouts. It is not a classification by hardware location, nationality of founders, or comprehensive corporate-group mapping. The purpose is to identify the coordination-layer actors to whom first-order double-spend attribution would run, not to map every machine or every beneficial owner.

OCEAN and Ocean appear as separate entries in the data source with different payment methods (unknown and PPLNS respectively). Both are US-linked and are aggregated under the US jurisdiction total.

Pool-Level Hashrate

Table 4: Reports All 22 Pools in the Dataset

Rank	Pool	Share (%)	Blocks	Jurisdiction
1	Foundry USA	31.48	1,306	United States
2	AntPool	15.74	653	China
3	F2Pool	11.33	470	Singapore
4	ViaBTC	9.50	394	China
5	SpiderPool	8.82	366	China
6	MARA Pool	5.71	237	United States
7	SecPool	4.15	172	China
8	Luxor	3.23	134	United States
9	Binance Pool	2.43	101	China
10	SBI Crypto	1.88	78	Japan
11	Braiins Pool	1.47	61	Czech Republic
12	OCEAN	1.23	51	United States
13	Unknown	0.96	40	Unknown
14	WhitePool	0.43	18	Unknown
15	Ultimus Pool	0.41	17	China
16	Ocean	0.36	15	United States
17	MiningCore	0.34	14	Unknown
18	BTC.com	0.29	12	China
19	Poolin	0.17	7	China
20	NiceHash	0.02	1	Slovenia
21	Parasite	0.02	1	Unknown
22	Public Pool	0.02	1	Unknown
	Total	**100.00**	**4,149**	

Appendix (Cont.)

Jurisdictional Aggregation

Table 5 aggregates hashrate by jurisdiction after applying the classification rule above.

Table 5: Bitcoin Hashrate by Pool-Operator Jurisdiction, March 2026

Jurisdiction	Hashrate Share (%)	Pools
United States	42.01	5
China	41.51	8
Singapore	11.33	1
Japan	1.88	1
Czech Republic	1.47	1
Slovenia	0.02	1
Publicly attributable	98.22	17
Unknown	1.77	5
Total	**99.99**	**22**

Interpretation

These figures establish that the *coordination layer* of Bitcoin mining—the set of entities that receive block rewards, distribute payouts, and organise pooled block production—is concentrated among publicly attributable pool operators in a small number of jurisdictions. Less than 2% of hashrate is attributable to pools without verifiable public operator attribution.

Pool-jurisdiction data do not establish: the physical location of all underlying ASIC hardware; the legal ownership of every contributing machine; or total jurisdictionally reachable hashrate. Nor do they establish ultimate beneficial ownership, which may differ from the disclosed operator in some cases. The paper uses these data for the narrower purpose of identifying the coordination-layer actors to whom first-order double-spend attribution would run in the event of a deliberate pooled attack. The evidentiary basis for that identification is public: pool operators self-identify through coinbase tags, public websites, and corporate filings.

Source: Mempool.space, accessed 13 March 2026. Jurisdictional corrections from corporate filings and public reporting as described above.

B. Enforcement Implications: Extended Discussion

This appendix provides the detailed enforcement discussion summarised in Section 4.

Exchange Intervention and Realised Attack Value

The gap between nominal and realised attack value ($V_R < V_{tx}$) reflects exchange freezes, recovery actions, and monetisation frictions. Foley *et al.* (2019) and Makarov and Schoar (2021) show that tracing, exchange linkage, and seizures are important features of the real Bitcoin environment. Cong *et al.* (2025) show that large crypto-enabled criminal enterprises are vulnerable because monetisation requires off-chain institutional interfaces. Nabilou (2019) argues that exchanges, payment institutions, banks, and large service providers remain legal chokepoints. A major double spend generates a sequence of responses: victim complaint, exchange alert, counterparty scrutiny, and pressure on off-ramp institutions. These interventions lower realised value and increase delay.

Pool Visibility, Public Identification, and Franchise Loss

Once a pool is named as the coordinator of a deliberate attack, it may lose counterparties, exchange access, banking relationships, hosting services, insurance, commercial credibility, and future member participation. Stylianou and Carter (2020) treat pools and exchanges as measurable economic actors. Deuflhard and Heller (2023) show that mining concentration and attack incentives are best understood in market and legal terms. Matsuzawa (2025) demonstrates that removing targeted enforcement (DUI checkpoints) produces a 12.4% increase in alcohol-related traffic fatalities, showing visible enforcement deters a broader population. Sanctioning one identified pool operator changes the perceived risk for all others.

Member Exit as an Enforcement Channel

The pool's coalition can become unstable before any court judgment. Thompson (2024) provides the analogue: La Cosa Nostra families became unstable when coordinators attracted public scrutiny, because members retained exit options. In pooled mining, a large identified attack increases both the direct risk to the coordinator and the indirect risk that members withdraw support. Where capital is specialised (Bitcoin ASICs), contributors face longer-term impairment through exclusion, taint, or loss of counterparties.

Appendix (Cont.)

Honest-Longest-Chain Exclusion

Under the longest-chain rule, the chain built by the greater effective hash prevails. Once a dishonest pool is identified and its connectivity degraded ($\sigma > 0$), the honest remainder's hash ($1 - a$) exceeds the attacker's effective hash $a(1 - \sigma)$, and the attacker's chain is orphaned. Honest miners do not need to coordinate: they follow the unchanged Nakamoto protocol. Infrastructure-layer degradation of the attacker's connectivity is the external element; the longest-chain rule is the protocol doing its job.

Naming, Routing, and Interconnection Dependence

A mining pool depends on domains, APIs, hosting, inter-domain routing, and interconnection paths. Root-zone management and identifier coordination remain centralised (Internet Assigned Numbers Authority, 2025). BGP instability can deny access and degrade connectivity (Haag *et al.*, 2019). Apostolaki *et al.* (2017) show that BGP manipulation can isolate mining capacity, slow block propagation, and facilitate double-spend exploits. Once a dishonest pool is identified, degradation of its routing raises stale and orphan rates and worsens payout distortions observed by contributing hash providers, strengthening both the operational-cost and member-exit channels.

Cross-Border Application

A pool may operate in one jurisdiction, target victims in another, and depend on exchanges in a third. The relevant question is whether the attacker remains dependent on enough reachable institutions for a coordinated response to become materially costly. Belnap *et al.* (2024) document that transparency regimes impose differential costs on intermediaries: capital-intensive, exchange-facing pool operators with large ASIC fleets and disclosed corporate structures are more exposed to enforcement than opaque, low-capital actors.

Cite this article as: Craig Wright (2026). Legal Deterrence in 'Permissionless' Consensus. *International Journal of Cryptocurrency Research,* 6(1), 102-133. doi: 10.67191/IJCCR.6.1.2026.102-133.