Collusion-Proof Decentralized Autonomous Organizations

Alexander Braun and Niklas Haeusle*

September 5, 2023

Abstract

We uncover severe collusion vulnerabilities in decentralized autonomous organizations on the blockchain. When voters act strategically and the system is poorly decentralized, payoff matching bribes compromise the punishment of malfeasant actors under conventional decentralized governance. We show that this issue can be mitigated through alternative voting mechanisms. Stochastic voting effectively decouples a tokenholder's influence from the voting behavior of others. Thus, collusion-proofness can be restored in the presence of sufficiently centralized governance tokenholders. In addition, masked voting increases collusion resilience through the obfuscation of individual voting behavior. Our findings are applicable to the blockchain oracle problem and, more generally, to the design of incentive-compatible and collusion-proof decentralized systems.

*Alexander Braun is Associate Professor of Insurance and Capital Markets as well as Director of the Institute of Insurance Economics at the University of St. Gallen, Tannenstrasse 19, CH-9000 St. Gallen, Switzerland. Niklas Haeusle is a Ph.D. student in Finance at the University of St. Gallen. The authors can be contacted via E-mail: alexander.braun@unisg.ch and niklas.haeusle@unisg.ch. We thank Stefan Buehler, Will Cong, Samuel Haefner, Hanna Halaburda, Winfried Koeniger, Fahad Saleh, Larry Samuelson, Gerry Tsoukalas, Rakesh Vohra as well as the participants of the Crypto and Blockchain Economics Research Forum (CBER), European Economic Association (EEA) Annual Meeting (2020), the American Risk and Insurance Association (ARIA) Annual Meeting (2023), the European Group of Risk and Insurance Economists (EGRIE 2021) annual meeting and the household finance research seminar at the University of St. Gallen (2021) for their helpful comments and suggestions. A special thanks to Stephan Karpischek.

1 Introduction

Without a central authority, decentralized autonomous organizations (DAOs) give rise to a unique challenge: How can they be designed in an incentive-compatible and collusion-proof way? In his Ethereum whitepaper, Buterin (2013) discusses the concept of an idealized DAO that operates under a fully transparent and tamper-proof set of business rules recorded in specific pieces of computer code. The latter are called smart contracts and maintained by a blockchain network. There is no hierarchy or human management, meaning that decisions are based on consensus of the network participants, who identify themselves with cryptographic tokens native to the project. All tasks involved in the production of goods or services are completed on an entirely transactional basis.

DAOs have seen unprecedented growth in recent years. In 2022, blockchain sources listed over 1'000 DAOs, 65 of which exhibited an asset base in excess of USD 100 million and more than 1'000 active members.[1] Together, these projects cover a wide range of applications: developer collectives such as BadgerDAO build infrastructure for the Web 3.0, decentralized finance (DeFi) organizations offer lending and borrowing services, and decentralized insurance platforms such as Nexus Mutual enable risk sharing without the need for an insurance company.

The operational procedures of DAOs, such as the mechanism employed for block validation and the rewards granted to miners, are usually meticulously designed by the founders. Yet, no matter how sophisticated the design of a system or an algorithm, there will always be unanticipated behaviors of individuals as well as changes in market conditions that require case-by-case solutions or protocol updates (Buterin, 2021). In traditional firms, governance typically relies on oversight. This is not possible in a decentralized setting, since executives and reporting lines do not exist. Moreover, the service providers of a DAO may participate on a one-off basis so that their liability is strictly limited. Finally, the legal enforcement of property rights causes prohibitively high costs, even if the network is relatively small. The reason is that DAO participants typically remain anonymous and may be distributed across various jurisdictions around the globe.[2]

[1] For more information see https://deepdao.io/organizations.

[2] The incompleteness of contracts governing international transactions is known to be a limiting factor for traditional firms (see, e.g., Antràs, 2005a,b). A decentralized setting, however, exponentiates the problem.

Hence, decentralization creates the need for new governance mechanisms that align different network participants along the common goals of the DAO and allow them to take suitable collective action. Specifically, DAOs require a set of rules that enable strategic and operational decisions, such as human intervention in the context of events that automated smart contracts cannot capture. From a high-level perspective, this paper examines the avoidance of misconduct when decisions rely on decentralized governance. We do not assume that voters are invariably honest or adversary, but instead act strategically and respond to incentives.

This is especially important to capture the aspect of collusion. While participants of a DAO are typically pseudonymous, they may still influence their peers to generate a favorable outcome for themselves. Communication in a blockchain context, relies on wallet to wallet messaging (blockscan chat) or off-chain channels and messengers, such as Telegram. Collusion for votes can be achieved through bribing strategies with different effects on the probability of a malicious actor being convicted.[3] Coordinated collusion tactics already pose a severe problem in decentralized systems. A well-known example are pump and dump schemes, in which malicious members of the crypto community orchestrate their actions to manipulate the price of coins in their favor (Xu and Livshits, 2019). Moreover, bribing in DAOs occurs openly and explicitly and even dedicated bribing protocols exist.[4] Bribers may also use a different name, obfuscate bribes through cryptocurrency exchanges or rent governance tokens.[5] The implications of collusion in blockchain networks extend beyond individual transactions or decisions. They can undermine the trust and security of the entire ecosystem, fundamentally impacting the utility and adoption of blockchain technology.

[3]To implement a bribing strategy, the briber may write a smart contract that specifies the bribe amount, which is conditional on an outcome. The outcome can refer to the aggregated voting result or to the voting decision of an individual wallet address (or a combination of both). The briber then communicates his offer through the usual crypto channels mentioned above. Voters now face a straightforward and fully transparent trade-off. They can either accept the bribe and vote for the preferred outcome of the briber or ignore the offer and vote truthfully. Trust is not required, since smart contracts make the bribe offers credible.

[4]See, e.g., the Bribe protocol and the Paladin protocol.

[5]See, e.g., Buterin (2021).

Our contributions are threefold. First, we show that conventional voting in decentralized systems is prone to collusion and ineffective in discouraging misconduct. A simple bribing strategy, which matches a voter's payoffs, establishes a cost-effective route to evade punishment. Second, we prove that adequate voting schemes for decentralized governance can effectively mitigate the risk of collusion. Stochastic voting restores collusion-proofness in the presence of sufficiently concentrated governance tokenholders. With stochastic voting, each tokenholder's influence becomes independent of the actions of their peers. Centralized tokenholders then have non-linear incentives to resist bribery because their chances of being pivotal are not only fixed but also high. Moreover, due to their large token holdings, they stand to lose more for a given drop in the token value after a malicious attack. Another way to make DAOs more collusion resilient is masked voting, which obscures voting outcomes just enough to make bribery strategies unprofitable for a wide range of parameters. Third, we put forward a sufficient condition under which the inherent frictions of a large decentralized network can contribute to making the system immune to collusion. Our findings may help to design DAOs that are both incentive-compatible *and* collusion proof. We illustrate this with a concrete application to blockchain oracles.

Current economic analysis of this problem usually assumes that there are a fixed number of adversaries, but the majority of voters are honest (Adler et al., 2018; Cai et al., 2020).[6] This includes the Chainlink whitepaper (Breidenbach et al., 2021), which suggests a two-stage system consisting of a first-tier network and a second-tier backstop. Upon closer inspection, this *shifts* the incentive and collusion problem from the first tier to the second tier, for which the paper assumes that voters behave honestly. We show that this assumption is problematic because rational voters may behave dishonestly.

Tsoukalas and Falk (2020) and Cong and Xiao (2023) also consider voting schemes in a decentralized context, with a focus on the information side. Our paper, in contrast, considers the problem of dealing with adversaries and strategic tokenholders. Belavina et al. (2020) examines misconduct in a crowdfunding setting and Zhang (2023) develops optimal compensation mechanisms. As oppposed to these papers, our focus is on bribery

[6]Exceptions in the wider blockchain literature that comprise strategic actors are Amoussou-Guenou et al. (2020), Auer et al. (2021), and Halaburda et al. (2021).

and we do not assume that adversaries are inherently dishonest. Instead, they become adversaries in the presence of incentives which make them act malfeasantly.

We also contribute to the recent literature on decentralized consensus (Amoussou-Guenou et al., 2020; Benhaim et al., 2022; Cong and He, 2019; Gans and Holden, 2022; John et al., 2020). The most related piece of research in this area is Cong and He (2019), who examine collusion in a blockchain context. Our work differs in two fundamental ways: First, Cong and He (2019) look at collusive anti-competitive behavior, while we examine collusion subsequent to misconduct. Second, they analyze how collusion arises, given the tension between decentralized consensus and information distribution, whereas we focus on specific collusion strategies and how to avoid collusion from a mechanism design perspective.

Lastly, we add to the literatures on tokenomics (Cong et al., 2021b), decentralized finance (Cong et al., 2022; John et al., 2022) and governance (Grossman and Hart, 1988; Harris and Raviv, 1988; Han et al., 2023). In particular, our results highlight the governance challenges arising from decentralization and the limits of classical remedies for new organizational forms. Our incentive-compatible and collusion-proof solution complements the overall governance toolbox available in financial economics.[7]

The remainder of this paper is structured as follows. In Section 2, we present a generic economic setting for the DAO, discussing the incentives for a service provider in the context of decentralized governance. Our analyses and findings are outlined in Section 3. In Section 4, we discuss our general findings in the context of a specific, but essential service provider, the blockchain oracle. Oracles operate as a vital gateway, mediating the flow of information from the outside environment (off-chain) to the blockchain (on-chain). Section 5 contains additional considerations with regard to the robustness of our main findings, including the impact of reputation and uncertainties. Finally, in Section 6, we conclude the paper. Proofs are mostly confined to Appendix A for the ease of exposition.

[7] In the wider sense, the paper relates to a specific facet of social theory, which addresses the problem of individuals intentionally misreporting their preferences to influence the process of aggregating individual preferences (Gibbard, 1973, 1977; Satterthwaite, 1975). In addition, we connect to the politicial economy literature, especially voting systems (Dal Bó, 2007; Dekel et al., 2008).

2 The Economic Problem

In our basic model, we consider a single service provider and a number of governance tokenholders who are entitled to vote. Service providers are crucial participants in DAOs as they create and maintain the systems and perform all tasks necessary to deliver the value proposition. Without loss of generality we explicitly separate the two roles here. Figure 1 illustrates our Nash game.[8]

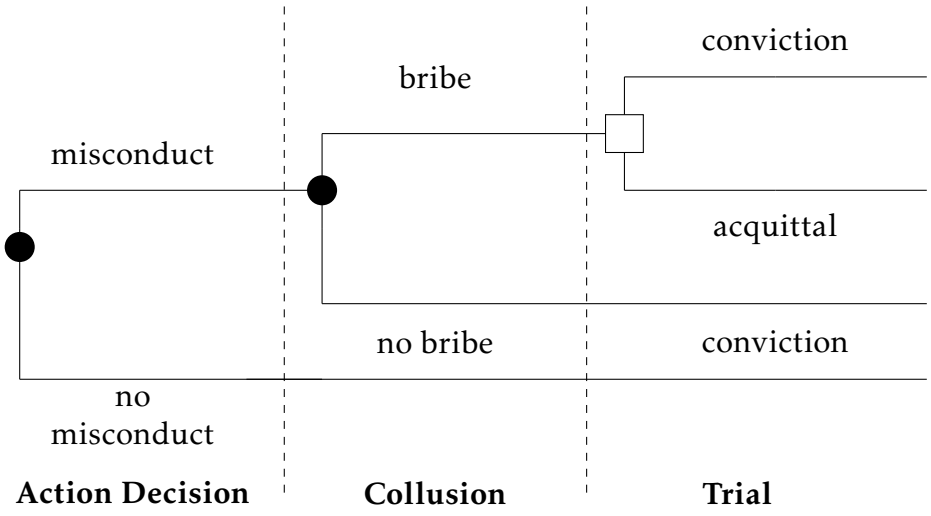

Figure 1: The Nash Game

The service provider posts a stake S as collateral, which consists of native network tokens. He then decides how to behave. If he refrains from misconduct, which can be shirking or manipulation, the outcome is "success" and no further actions occur. Misconduct, in contrast, produces a "failure" (or faulty outcome) about which the governance tokenholders are alerted and a trial process is initiated.[9] Subsequently, the network tokenholders need to decide on the consequences.[10]

There are n governance tokenholders, holding heterogenous amounts of tokens T. The tokens make them eligible to vote on the decision of whether the service provider is guilty of misconduct and should be punished. Since tokens have a positive price, votes cannot

[8]Note that in the highly theoretical case where detection and punishment of misconduct can be completely and reliably automated, this problem does not pose itself.

[9]This step can, e.g., be implemented through a watchdog system as employed by Chainlink (Breidenbach et al., 2021).

[10]In our standard setting, tokenholders can discard false alarms with certainty. Later on, we discuss a setting in which the tokenholders have imperfect information on regarding the validity of the alarm but receive a noisy signal (see Section 5).

be acquired costlessly. If at least 50 percent of the votes are cast for conviction, the DAO confiscates the stake of the malfeasant service provider. Before this decision, the service provider has the option to bribe the tokenholders. If he does not bribe, he gets rightly convicted and loses his stake. If he bribes the governance tokenholders using an optimal strategy, he might get acquitted.

Let R (reward) denote the service provider's payoff when he refrains from misconduct and G (gain) the payoff for misconduct. G could, e.g., be the benefit of not exerting effort. R and G can also be interpreted in the context of exploits, in which case G might be significantly larger. We will discuss this notion in more detail in Section 4.

In the following, we introduce our notation for the bribing strategies. Let $\mathcal{K}$ denote the outcome space of the voting process and $k \in \mathcal{K}$ a specific realization therein. $\mathcal{C}$, $\mathcal{A}$, and $\mathcal{Q}$ represent degenerate versions of $\mathcal{K}$. $\mathcal{C}$ subsumes all outcomes which lead to conviction, $\mathcal{A}$ subsumes all outcomes which lead to acquittal and $\mathcal{Q}$ subsumes all outcomes where an individual voting decision is pivotal, i.e., a switch of one voter changes the outcome from $\mathcal{C}$ to $\mathcal{A}$ or vice versa. Note that $\mathcal{C}$, $\mathcal{A}$, and $\mathcal{Q}$ are not disjoint sets, but ex-post outcome sets which help us to describe the optimal bribing strategies later on. Furthermore, $B_{j,k}$ stands for the bribe paid to person j in state k, and m denotes the number of individuals to which a positive bribe was offered. As the structure of a DAO is loose and distributed, voters do not have close contact to each other. The briber thus incurs search costs $c(m)$, that grow in the number of contacted network participants m. Finally, D is the relative drop in token value caused by the acquittal of a guilty provider, implying that the tokens of every tokenholder are then worth $(1 - D)T < T$ (in real value terms).[11]

The planner of the DAO aims to minimize the staking (collateral) costs rS by choosing an optimal stake (see Problem 1). Here, r is the service provider's (exogenous) cost of capital or opportunity cost. The optimization has to adhere to several constraints. The incentive compatibililty constraint (IC) and the participation constraint (PC) are also found in economic analyses of traditional organizations. (IC) states that it needs to be preferable for a service providers to not misconduct and lose the stake. (PC) implies that service

[11] This assumes that the secondary market for the network's native token prices financial and/or reputation loss caused by the unpunished misconduct of the service provider, a premise also indirectly supported by empirical studies suggesting a negative token response to bad news (Hashemi Joo et al., 2020; Anamika and Subramaniam, 2022).

providers will only be prepared to join the DAO, if they do not expect to suffer a loss.[12] (BC) is specific to our DAO problem. It says that a solution without bribing can only exist if it is more expensive to bribe than to forfeit the stake.

Problem 1

$$\arg\min_{S} \quad -rS$$

$$s.t.$$

$$R \geq G - S \tag{IC}$$

$$R - rS \;\geq\; 0 \tag{PC}$$

$$S \;\leq\; \sum_{k \in \mathcal{K}} P(k|\boldsymbol{B}) \left(\sum_{j=1}^{m} B_{j,k} \right) + P(\mathcal{C}|\boldsymbol{B})S + P(\mathcal{A}|\boldsymbol{B})SD + c(m) \tag{BC}$$

The first term on the right hand side of (BC) is the expected value of the bribes across the whole probability space $\mathcal{K}$.[13] The second term is the expected stake loss in case of conviction, i.e., if the bribing fails. The third term is the expected loss in case of acquittal. In this case, the provider does not lose his stake, but since he stakes in native tokens, he participates in the drop in token value. The probability of a specific voting outcome $P(k|\mathbf{B})$, the probability of conviction $P(\mathcal{C}|\mathbf{B})$, and the probability of acquittal $P(\mathcal{A}|\mathbf{B})$ are contingent on the allocation of the bribes $\mathbf{B}$ to the voters.[14] Finally, the fourth term represents the search costs $c(m)$ that arise from the decentralized setting. These costs make bribing more expensive as they increase with the number of voters m that have to be identified and contacted.

[12] In the context of shirking, the participation constraint is slightly modified by incorporating the additional cost of effort into the left-hand side of the equation.

[13] Note that the bribes can also be conditioned on states where the malefactor gets acquitted ($k \in \mathcal{A}$).

[14] In Lemma 1, we describe the most efficient bribing strategies and explain how the bribe amounts affect these probabilities.

3 Analysis

We are interested in the nash equilibria where it is economically unattractive to misbehave and then collude with others to avoid conviction. We define these as collusion-proof equilibria. If **Problem 1** has a solution, the system is feasible, incentive compatible, and collusion proof.

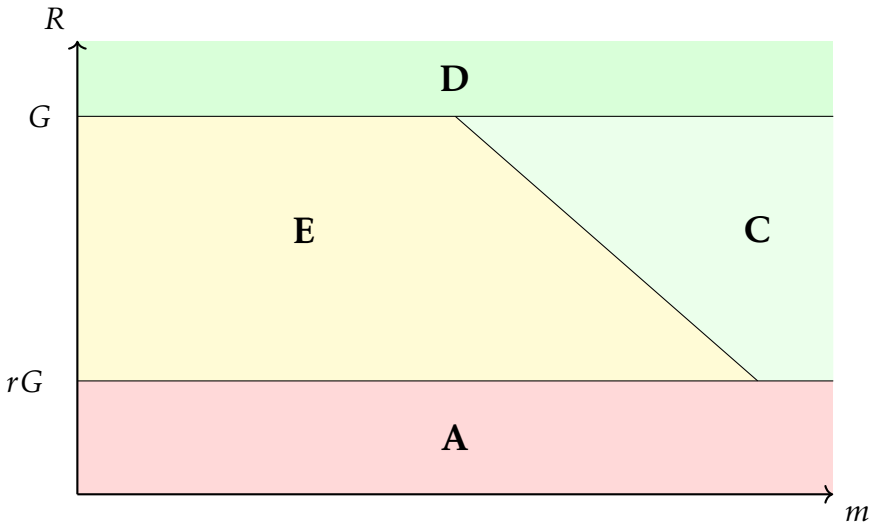

Figure 2: Economic Feasibility of a DAO

This figure shows the feasibility of a DAO from the perspective of a service provider, depending on the reward R and the number of participants m that need to be contacted for bribing. In Area A, the DAO is not feasible. In Area D, it is both feasible and inherently collusion proof. Areas C and E imply that the DAO is feasible, but not necessarily collusion proof.

Figure 2 illustrates how the feasibility of the DAO depends on the two critical parameters R and m. In Area A ($R < rG$), the existence of the DAO is not socially optimal or not feasible due to collateral costs.

Furthermore, in Area D, where R is very high relative to the gain from misconduct, the DAO is feasible and inherently collusion proof, even without staking. Problems arise in Areas C and E. In these cases, the DAO is feasible, but the reward R is not high enough to incentivize the service provider automatically. Thus, staking is required and it may be beneficial to engage in malfeasance. To prevent this and the associated negative externality on all tokenholders, incentivization via staking needs to be complemented by a governance that avoids collusion. With regard to the optimal stake S^*, we postulate the following:

Remark 1 *The optimal incentive-compatible stake is* $S^* = G - R$.

Proof. This can be simply solved by setting $S = G - R$ in **Problem 1**. Ignoring (PC) and (BC), $S^* = G - R$ is the optimal stake, because it minimizes the staking costs, while still being large enough to ensure incentive compatibility. Lowering it would violate (IC). ■

The designer must know G to calculate the optimal incentive-compatible stake. In most applications this can be well estimated.[15] (BC) places an upper bound on the stake. For stakes that exceed this limit, the malefactor has an incentive to bribe other members of the DAO in order to avoid conviction. Any increase in the stake tightens (BC) rather than relaxing it. The same is true for (PC). Hence, the optimal stake S^* is only collusion proof (and therefore a solution to Problem 1), if the expected costs of bribing, i.e. the right hand side of (BC) is high enough.

A natural situation for this condition to be fulfilled is when the DAO is large and decentralized. In this case (Area C), very high search costs $c(m)$ relax (BC). We delay the analysis of Area C for now and set $c(m) = 0$.

As we will see, the remaining terms that are part of the expected bribing cost crucially depend on the voting system used by the DAO for the tokenholder decisions (Area E). This will be our focus in the next sections. In particular, we explicitly compare three potential voting schemes: i) majority voting, ii) masked voting, and iii) stochastic voting. Further alternatives will be discussed in Section 5.

3.1 Majority Voting

Under a majority voting scheme, individual votes are weighted with the voting power of the tokenholders. If a 50 percent majority is reached, the accused service provider will be convicted. After the voting, all details become fully transparent. To examine the collusion-proofness of this voting scheme, we first derive the optimal bribing strategy by

[15] E.g. as the monetary gain to be expected by a data oracle from the manipulation of a smart contract.

minimizing the bribing costs:[16]

$$\arg\min_{\mathbf{B}} \quad \sum_{k\in\mathcal{K}} P(k|\mathbf{B})\left(\sum_{j=1}^{m} B_{j,k}\right) + P(\mathcal{C}|\mathbf{B})S + P(\mathcal{A}|\mathbf{B})SD. \tag{1}$$

Lemma 1 describes the optimal bribing strategy in the spirit of Dal Bó (2007). With this strategy, the payoffs of the tokenholders in each state are marginally outbid.

Lemma 1 *The most efficient bribing strategy takes the following form:*

$$B_{j,k} = \begin{cases} \epsilon & \text{if } k \in \mathcal{C} \setminus \mathcal{Q} \\ DT_j + \epsilon & \text{if } k \in \mathcal{Q} \\ \epsilon & \text{if } k \in \mathcal{A} \setminus \mathcal{Q}, \end{cases} \tag{2}$$

where ϵ is a small but positive number.

Proof. See Appendix. ■

The intuition behind this dominant bribing strategy is as follows. If the majority of the other governance tokenholders vote for conviction or acquittal, voter j, who cannot influence the outcome, at least obtains a small inflow ϵ, given he votes for acquittal. In case his acquittal vote is decisive, however, he obtains a payoff that overcompensates the loss in the value of his token holdings by ϵ. Hence, with majority voting the following result can be derived:

Proposition 1 *Majority voting does not prevent collusion in a DAO.*

Proof. The bribing strategy outlined in Lemma 1 implies that it is strictly dominant for any individual to accept the bribe and vote for acquittal. If this bribe is offered to everbody, everybody will accept the bribe and $P(\mathcal{A} \setminus \mathcal{Q}) = 1$. Because the voter is then never pivotal, the payout per voter is only ϵ. The bribing costs for a malfeasant provider are then only $n\epsilon$ and therefore arbitrarily close to zero. ■

[16] Recall that $c(m) = 0$ for this analysis.

This proposition remains consistent across varying thresholds for conviction or acquittal, no matter whether these thresholds are higher than or lower than 50%. However, it's important to highlight a particular exception: When a single tokenholder possesses more than 50% of the entire voting power, majority voting is resilient against collusion, at least given reasonable sets of parameter values.

3.2 Masked Voting

Standard majority voting in a blockchain system implies that the decision and the voters (or their public keys) that carried it are transparent after the vote has taken place. Under masked voting, in contrast, conviction still requires at least 50 percent of the votes, but the individual voting behavior of the network participants is not disclosed. The following then holds true:

Proposition 2 *Masked voting increases the collusion resilience of a DAO.*

Proof. See Appendix. ■

Now the malfeasant actor does not know who voted for acquittal and who for conviction. Accordingly, he can only condition bribes on the total outcome, but not on individual votes. This substantially reduces the set of available bribing strategies. The proof shows that bribing under a large range of reasonable parameter values, bribing is now very costly, because the few available bribing strategies require the malfeasant service provider to compensate 50 percent of the voting power in the DAO. It should be highlighted, however, that masked voting makes a DAO more *collusion resilient*, but not necessarily *collusion proof*. If the drop in token value after a false acquittal is very small, the overall network value (sum of all tokens) is small, or the voting power in the network is concentrated, collusion cannot be prevented with certainty.[17]

Despite its potential, masked voting is rarely observed in reality.[18] This could be

[17] This finding has similarities to Cong and He (2019). In their paper, sellers can use the blockchain to gather information about market demand. This information can be used to sustain a collusive equilibrium where prices are set in a non-competitive fashion by more accurately punishing deviating behavior. In our model, restricting the space of vote outcomes prevents a briber to accurately "punish" deviating voting behavior.

[18] A notable exception is Aragon Court.

attributable to the goal of transparency that is characteristic for most blockchain networks, which contradicts the idea of masked voting. Yet, with the advancements in zero-knowledge proofs, implementing such a solution without compromising transparency may increasingly become a viable option.

3.3 Stochastic Voting

Under stochastic voting a single decisive vote is drawn from the overall pool of votes with a probability that is proportional to the voter's token holdings.[19] A 50 percent majority is thus not required for the conviction of the malfeasant service provider. Such a voting scheme leads to the following outcome:

Proposition 3 *A DAO with stochastic voting and a concentration of voting power among governance tokenholders is collusion proof.*

Proof. See Appendix. ■

This proposition comprises two crucial elements: stochastic voting and concentrated tokenholders. Both are needed to ensure collusion proofness. Cheap bribing strategies such as the one described in Lemma 1 rely on the fact that, under majority voting, a tokenholder's influence is dependent on the voting behavior of all other tokenholders. Stochastic voting, in contrast, disentangles a tokenholders' influence from the voting behavior of other voters and thus *fixes* the pivotal voting probability.

Furthermore, the proposition highlights the importance of a concentration in the voting power. When voting is stochastic but every voter has only small token holdings, reaching an acquittal decision will still be almost costless for the briber. Concentrated tokenholders, in contrast, are non-linearly incentivized to refuse bribes. There are two drivers behind this:

i Under the stochastic voting mechanism, there is a high fixed probability that a large tokenholder is pivotal.

[19] This is similar to the determination of the next block writer under Proof-of-Stake consensus.

ii Given a large tokenholder is pivotal, he has much more to lose, since any value decrease of the network tokens has a more substantial impact on his wealth position.

The concentration of tokenholders was already suggested as a governance instrument by Vitalik Buterin in a 2017 blog post:

> "It's worth noting that this [...] is not a prediction that all tightly coupled voting systems will quickly succumb to bribe attacks. It's entirely possible that many will survive for one simple reason: all of these projects have founders or foundations with large premines, and these act as large centralized actors that are interested in their platforms' success that are not vulnerable to bribes, and hold enough coins to outweigh most bribe attacks. However, this kind of centralized trust model, while arguably useful in some contexts in a project's early stages, is clearly one that is not sustainable in the long term."

This view relies on the implicit assumption that concentrated tokenholders are inherently honest. We show that with strategic behavior, in contrast, concentrated tokenholders are only sufficient to ensure collusion proofness if the DAO additionally relies on a stochastic voting scheme.[20]

There are three (small) drawbacks to consider. First, stochastic voting will always be associated with a positive probability of acquittal if the malfeasant service provider himself has a right to vote. The reason is that he could be randomly chosen as the pivotal voter. A trivial solution to this issue is to ensure that accused DAO participants are not allowed to vote. Second, the off-equilibrium path might still include partial bribes, even if voting power is sufficiently concentrated. One possible solution could be the establishment of a voting threshold which excludes those DAO members from voting that are most easily targeted by the briber. Those are the ones with small token holdings.[21] Even without such a threshold restriction, however, the formerly dominant bribing strategies will no longer work in expectation. This is sufficient to discourage misconduct. Third, given voting power is concentrated, it is not trivial to ensure a proper decentralization when

[20] Network participants that own a large fraction of the overall token supply, so-called "whales" can be empirically observed in the vast majority of DAOs (and other blockchain-based projects) that exist today, e.g., Aragon, BadgerDAO, Compound, Etherisc etc.

[21] We formalize this through Equation A.14 in the Appendix.

the DAO grows larger and the token concentration is no longer needed for governance reasons.

3.4 Size/Decentralization

Recall that in the previous sections, we had set $c(m)$ to zero. We now introduce search costs for the briber and take a closer look at area C of Figure 2. Here, the reward R is not sufficient to make the DAO incentive compatible and collusion proof, but the number of tokenholders m that the briber needs to contact is large. This implies that the search costs $c(m)$ for the malfeasant service provider are high enough to ensure the feasibility of the optimal incentive-compatible stake. First, we need to develop a more detailed understanding of the drivers of m. We draw on the Herfindahl index[22] H_V as a measure for the concentration of the voting power:

$$H_V = \sum_{j=1}^{n} \left(\frac{T_j}{\sum_{j=1}^{n} T_j} \right)^2 . \tag{3}$$

H_V is a function of i) the total number of tokenholders n (i.e., the network size) and ii) the distribution of tokens among the tokenholders. It ranges from $\frac{1}{n}$ (perfect decentralization) to one (perfect concentration). When all participants in the DAO have equal token holdings, $H_V = n\left(\frac{T}{nT}\right)^2 = \frac{1}{n}$.

Given H_V, we can model how m is related to the degree of decentralization of the DAO network. Intuitively, the less concentrated (more decentralized) the voting power (i.e., the closer H_V is to $\frac{1}{n}$), the higher the number of tokenholders (for a given network size) that the malfeasant service provider needs to contact with his bribing offer. The DAO will therefore be collusion proof through a high $c(m)$, if the network is both sufficiently decentralized *and* large. To see this, reconsider the bribing constraint (BC):

$$S \leq \sum_{k \in \mathcal{K}} P(k|\mathbf{B}) \left(\sum_{j=1}^{m} B_{j,k} \right) + P(\mathcal{C}|\mathbf{B}) S + P(\mathcal{A}|\mathbf{B}) SD + c(m). \tag{4}$$

Evidently, a high $c(m)$ relaxes (BC). Deriving an explicit expression for m, given a

[22] The Herfindahl index is widely used to measure concentration, for example by the Federal Trade Commission (FTC) to measure market concentration.

particular network size n as well as a specific degree of concentration of the voting power among the participants H_V, and inserting it into (BC) allows us to derive the following result:

Proposition 4 *A sufficient condition for a collusion-proof DAO under majority voting is*

$$S^* \leq c\left(\frac{n}{2} - \frac{1}{2}\sqrt{n^2 - \frac{1}{H_V}}\right). \tag{5}$$

Intuitively, for an equal distribution of tokens $\left(H_V = \frac{1}{n}\right)$, the Herfindahl index is inversely correlated with the number of participants in the DAO and a large n will lead to a low concentration of the voting power. This implies a large m and, in turn, prohibitively high costs of collusion. However, if the network tokens are very unequally distributed, voting power can be highly concentrated, even if n is large ($H_V \lesssim 1$). The malfeasant service provider then only has to contact and bribe a small number of influential DAO members that can provide him the decisive edge in the voting process.

Proposition 4 will be fulfilled if the DAO exhibits both a sufficient degree of decentralization *and* a large network size. However, this is not a design choice. Size is exogenous and intertwined with a chicken and egg problem: The DAO becomes collusion proof if its sufficiently large and decentralized, but economic agents will not want to participate in a DAO where collusion inhibits the decentralized value creation process. Furthermore, empirical facts point to the possibility that size and decentralization may be difficult to achieve at the same time: in Proof-of-Work as well as in Proof-of-Stake blockchain networks, size is often correlated with a tendency to recentralize (He et al., 2020).[23]

4 Application to the Oracle Problem

4.1 Blockchain Oracles

In this section, we illustrate the relevance of our findings based on the case of the oracle, a key service provider in DAOs. Oracles grant blockchain-based systems access to off-chain data. They thus unlock the true potential of DAOs by allowing them to expand

[23] A well-known example is the highly concentrated hashing power among Bitcoin mining pools (Cong et al., 2021a).

into many interesting real-world applications. Blockchain oracles gather diverse types of data, including sensor data, price data, weather data, exchange rate data, geolocation data, sports data, and social media data. For instance, an educational data oracle might provide the answer to questions like, "What is this student's ID number?". A financial data oracle might report the current ETH/USD exchange rate. Similarly, in the realm of supply chain management, a sensor oracle could transmit temperature data for a cold chain in food transportation. Chainlink (2021) estimates that up to 90% of all smart contracts rely on such external data. While on-chain information can be easily verified through the ledger, this is not the case for external data feeds. Hence, oracles are vulnerable to manipulation and collusion, presenting a significant challenge and the risk of high losses for blockchain-based systems such as DAOs.[24]

4.2 Decentralized Oracles

An alternative to a single (centralized) oracle is a system of distributed oracle nodes. These decentralized data oracles, also known as Decentralized Oracle Networks (DoNs), introduce redundancy and collate data from multiple sources or entities, thereby enhancing system resilience by providing identical information from various nodes. Both types, centralized and decentralized oracles, find application in reality. For example, Etherisc utilizes a centralized oracle to acquire satellite data for their crop insurance application.[25] According to Cong et al. (2023), 17% of DeFi protocols employ decentralized oracles.

There are significant cost concerns with this approach. While adding redundancy may enhance data quality, it can become highly inefficient when data oracles need to exert effort to gather data that is not readily available or inexpensive. Additionally, situations may arise where seemingly unrelated oracles, intended to provide redundancy, all rely on the same upstream data provider, which becomes the sole true source. Besides these issues, a decentralized oracle will not be effective against collusion. After all, multiple oracles in a DoN must also agree to determine and penalize malfeasant actors. Thus, instead of bribing tokenholders, a malicious oracle can bribe other oracles to influence

[24]Case in point, MakerDAO, a prominent DAO, had recently incurred substantial losses due to a stuck oracle. See Coindesk Website for more information.

[25]https://forum.etherisc.com/t/etherisc-oracle-questions/167

the information fed into the blockchain. Therefore, our main findings are also relevant for this institutional setting in the DAO context.

4.3 Technical Solutions

There are certain technical solutions that create an authenticated data feed.[26] Intel's SGX, e.g., is a set of processor instructions, providing hardware-based protection for user-level code. It supports remote attestation, allowing external parties to verify the integrity and authenticity of code within an enclave. In addition, Town Crier (TC) is a centralized oracle solution that uses a trusted execution environment (TEE) within Intel's Software Guard Extensions enclave to serve source-authenticated data to smart contracts. DECO is a decentralized oracle that allows to securely prove statements about the data while keeping the information itself secret. We argue that such technical solutions and a collusion-proof mechanism design are not an either/or choice. Even the most sophisticated systems may fail and in such a case, incentives play a key role with regard to the decisions of DAO participants that act strategically. Technical solutions and a proper governance should thus be used in a complementary way.

4.4 Exploits and Pricing Oracles

Our analysis is applicable to most, but not all oracle problems. In particular, collusion-proofness cannot be achieved in the presence of exploit opportunities that imply a very large G. Extreme gains substantially inflate the incentive-compatible stake and thus affect the system in two ways. On the one hand, the high stake will likely violate (PC) and thus compromise the feasibility of the entire DAO, independent of the voting scheme and governance system. On the other hand, large stakes make it ceteribus paribus more profitable to bribe leading to a violation of (BC).

Decentralized finance and decentralized insurance are areas subject to this problem. An unscrupulous pricing oracle can manipulate data to trigger the payout of smart contracts, for example those governing highly leveraged derivative or parametric insurance transactions. In this case, the exploit opportunities can exceed the total value of the DAO,

[26] An overview can be found in Pasdar et al. (2021).

i.e., G is larger than the sum of all tokens combined. Even if the negative effects of such an exploit to all tokenholders are internalized, it would still be profitable for a malfeasant actor to alter the data.

There are mitigation strategies that limit the gains of pricing oracles from such exploits and therefore relax (PC) and (BC). One option is to disallow highly leveraged derivative contracts. Another one is to define a strict cap for the payoffs or for the number of contracts that a single oracle can serve.[27] Such constraints do not necessarily limit growth opportunities. As the DAO expands, its value will grow alongside it. This increases the capacity that the DAO can handle without violating either one of the aformentioned constraints.

4.5 Further Applications

Our model particularly applies to service providers which develop and maintain the systems of a DAO. Apart from blockchain oracles, these include, e.g., infrastructure contributors such as Infura, who offer reliable access to blockchain networks. Also included are software developers who work on the code, storage providers who manage data securely, and Know Your Customer (KYC) and Anti-Money Laundering (AML) service providers who ensure compliance with critical regulations.

Moreover, our findings carry over to DAO proposals, which are common for strategic decisions and upgrades to the source code. Proposals are submitted to the governance tokenholders for a collective decision, who then vote following the rules laid out by the DAO. Upon approval, the proposed modifications are implemented through blockchain-based smart contracts. Proposals can be broadly classified into strategic proposals and operational proposals. An example of a strategic proposal would be a lending-DAO adding a new coin to its portfolio. Operational proposals, on the other hand, might suggest the addition of new functionalities such as the ability to monitor user positions or the establishment of an address provider registry. Our research, which emphasizes the necessity of a collusion-proof governance voting scheme, could hold particular significance for operational proposals, especially in its capacity to stave off and discourage poorly crafted or malicious proposals.

[27] In practice this becomes more feasible if a large number of independent pricing oracles is available.

5 Robustness

5.1 Uncertainty

There are three cases of uncertainty in which our stylized model may deviate from real-world settings: i) detection of uncertainty, ii) false alarms and iii) the inability to differentiate deliberate misconduct from bad luck. In the first case, malfeasance may go undetected such that no alert will be issued. In the second case, governance tokenholders may not be able to distinguish a genuine issue from a false alarm with certainty.[28] Finally, a faulty outcome may arise even though there was no misconduct.[29]

While these sources of uncertainty are not considered by our base model, they constitute trivial extensions as long as a noisy, but informative signal exists. Noisiness leads to a positive probability of acquittal. To counter this effect, the expected punishment has to be adjusted upwards. This is achieved through an increase in the optimal incentive-compatible stake. The noisier the signal, the higher the required stake. While this generally reduces the feasibility of the DAO through (PC) and the collusion-proofness through (BC), our core results remain unchanged.

5.2 Sybil attacks

Our base model does not account for the possibility that service providers may build up voting power by purchasing governance tokens under pseudonymous wallet addresses, a strategy known as a Sybil attack. The larger the total token holdings of the service provider himself, the smaller the frictions associated with bribery. In an extreme scenario, a service provider might even amass half of the network's voting power, which would allow him to avoid conviction with certainty.

The possibility of such Sybil attacks does not change our results. We illustrate this by comparing the costs of bribery and token acquisition. To achieve a majority, a provider must acquire half of the voting power. Given the token's value decrease upon a wrongful acquittal $(1 - D)$, the associated cost is $\frac{1}{2} D \sum_{j=1}^{n} T_j$. This amount equals the bribing cost

[28] For oracle failures, this is usually not an issue. For instance, a price oracle for Synthetix misreported on June 25, 2019 the price of the Korean Won to be 1000x higher than the true rate. It was obvious to all tokenholders that this oracle failed (Medium, 2021).

[29] This form of uncertainty only matters, if the DAO aims to distinguish willful misconduct from bad luck.

under masked voting (Proposition 2) and exceeds the cost of the optimal bribing strategies under majority and stochastic voting.[30] Intuitively, a service provider that acquires more tokens than necessary for the stake internalizes part of the losses from his own misconduct. This offsets the gains in voting power. Our results thus remain valid.[31]

5.3 Reputation

So far we did not consider the potential impact of reputation on a service provider's inclination to misbehave and bribe. Reputation manifests itself in two primary forms. The first is through the forgone future fee opportunity in the DAO network (Breidenbach et al., 2021). The second form concerns the service provider's reputational equity outside the specific blockchain, e.g., in the off-chain world. However, reputation effects are only a concern for unsuccessful bribing attempts. After all, the essential aspect of bribery is that it alters the "truth" itself. Consequently, when a bribe is successful, the service provider's reputation will likely remain high.

Our model can readily be adapted to include the reputation concerns of rational service providers. From an economic perspective, reputation functions like an implicit stake, which can be combined with the explicit stake S (Breidenbach et al., 2021). A full or partial replacement of the explicit stake with the implicit stake derived from reputation relaxes all three constraints. It is particularly helpful for (PC), as explicit staking entails a fixed cost of capital, which does not arise for implicit stakes associated with reputation concerns. Hence, our main results also hold in the presence of reputational effects.

5.4 Vote Delegation

Over the past two years, the idea of token delegation — where on-chain token owners transfer their voting rights to others — has significantly grown in popularity. Benhaim et al. (2022), for instance, examine voting behavior in a blockchain context when votes are delegated to a decision-making committee. The primary benefit of such a committee is that a small group of experts can make more informed, uniform, and impactful decisions.

[30] Recall that the pivotal bribing strategy of Lemma 1 was associated with negligible costs of $(n-1)\epsilon$ (as ϵ is infinitesimally close to 0).

[31] Particularly Proposition 3.

Without the committee, thousands of stakeholders, some holding minimal stakes and others lacking ample time, would need to sift through numerous proposals.

Our results are robust to such a change in governance design. First, although token delegation to a committee consolidates voting power and thus makes it a bit easier to bribe, the pivotal bribing strategy of Lemma 1 under majority voting is nearly costless anyway. Second, an artifical tokenholder concentration through token delegation does not help, since committee members do not bear the externalities of administered tokens. Therefore, stochastic voting and a genuine concentration of governance tokenholders is still required to overcome the collusion problem.

5.5 Short selling of tokens

In our model framework, the token drop D disincentivizes governance tokenholders to accept bribes. When misconduct happens and is detected but falsely not punished, tokenholders suffer a negative wealth shock. Such a token drop could increase the gains G for the malfeasant service provider, if short selling of the token is possible. As already discussed in Section 4.4, a substantial inflation of G is at odds with both the feasibility and collusion-proofness of the DAO. Theoretically, G could become extensive if D is large and the shortable token amount is high. In reality, however, the amount of tokens that can be shorted via derivatives or contracts for difference is limited. Therefore, a system with stochastic voting and concentrated tokenholders remains collusion-proof.

5.6 Quadratic Voting

Apart from the voting schemes discussed in Section 3, quadratic voting is regularly proposed in a blockchain context (Lalley and Weyl, 2018; Buterin et al., 2019; Benhaim et al., 2023).[32] Under quadratic voting, tokenholders vote proportionally to the strength of their preferences. This explicitly entails the option to pay for additional votes to express one's support for an outcome. A quadratic cost function for vote pricing is put in place and ensures that each additional vote becomes more and more expensive.

[32] According to Benhaim et al. (2023) quadratic voting is used as a funding method for cryptocurrency startups (Gitcoin 2021a), and also in blockchain governance (Panther.io 2021).

There are two main reasons why quadratic voting does not solve our collusion problem. First, quadratic voting assumes that individual users are unable to vote as multiple entities. Yet, in a Sybil attack, any large tokenholder may distribute his tokens across multiple wallets and thus avoid the quadratic voting costs. Second, under quadratic voting, voters act on the assumption of a significant marginal pivotality of their votes (Mueller, 1973; Lalley and Weyl, 2018). However, with strategic behavior and especially under the bribing strategies described in Section 3, this pivotality is close to 0.

5.7 Voter Punishment

Another proposed solution to reduce collusion is to slash the tokens of voters who did not vote with the majority. If, e.g., 60% voted for conviction and 40% for acquittal, the 40% would lose a predefined amount of tokens.[33] There are two major flaws with this approach. Firstly, this links the voting behavior to the belief about the decision of other voters instead of the truth. Fearing that they might not be voting with the majority, rational actors will want to vote in line with prevailing sentiment. Secondly, the bribing strategy from Lemma 1 can be modified to account for this policy change and therefore remains effective.[34]

6 Conclusion

We take a normative approach to uncover severe collusion vulnerabilities in DAOs with majority (and masked) voting schemes when voters act strategically and the network is poorly decentralized. We then prove that stochastic voting in combination with a sufficient concentration of votes among governance tokenholders renders bribing ineffective. Once collusion-proofness is established, DAO governance can rely on the implementation of the optimal incentive-compatible stake to deter misconduct by service providers. We highlight blockchain oracles as an interesting real-world application of our findings. Finally, we illustrate the robustness of our results against various additional considerations such as uncertainy, Sybil attacks, reputation, and vote delegation. The insights of

[33] The decentralized prediction market platform Augur, e.g., uses such a mechanism.

[34] Specifically, the bribe in the case $k \in \mathcal{C} \setminus \mathcal{Q}$ would need to be extended by the predefined amount of tokens that voters stand to lose if they turn out to be among the minority.

our analysis stretch beyond the oracle problem and have important implications for the fast-growing realm of DAOs.

Our work leaves several interesting directions for future research, of we want to highlight two. First, while we deliver an economic rationale for vote concentration among tokenholders, such a setting may create unwanted side effects. Specifically, large tokenholders could be inclined to make decisions that serve their own interest to the detriment of smaller token holders. It will thus be interesting to analyze the trade-off between collusion-proofness and other economic goals, such as the avoidance of exploitative practices. Second, DAO governance should be further examined from an empirical perspective. In particular, insights on the dynamics of misconduct and collusion are currently still scarce. Using the transparent nature of blockchain systems, researchers could analyze suspicious transactions and voting patterns in on-chain data and network structures.

We wrap up the paper by emphasizing that the problem of collusion is not simply an academic concern. Instead, it is a major issue that needs to be tackled before DAOs can unfold their potential in many more real-world applications. Collusion undermines the principle of one token, one vote, thus distorting the collective decision-making process that this organizational form is designed to uphold. The consequence is an erosion of trust that ultimately threatens the very feasibility of DAOs. If collusion becomes rampant, honest members will be disincentivized from participating, which could lead to the complete downfall of the DAO as an institutional arrangement.

References

Adler, J., Berryhill, R., Veneris, A., Poulos, Z., Veira, N., and Kastania, A. (2018). Astraea: A decentralized blockchain oracle. In *2018 IEEE international conference on internet of things (IThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData)*, pages 1145–1152. IEEE.

Amoussou-Guenou, Y., Biais, B., Potop-Butucaru, M., and Tucci-Piergiovanni, S. (2020). Committee-based blockchains as games between opportunistic players and adversaries. *Working Paper*.

Anamika, A. and Subramaniam, S. (2022). Do news headlines matter in the cryptocurrency market? *Applied Economics*, 54(54):6322–6338.

Antràs, P. (2005a). Incomplete contracts and the product cycle. *American Economic Review*, 95(4):1054–1073.

Antràs, P. (2005b). Property rights and the international organization of production. *American Economic Review*, 95(2):25–32.

Auer, R., Monnet, C., and Shin, H. S. (2021). Permissioned distributed ledgers and the governance of money. *SSRN Electronic Journal*.

Belavina, E., Marinesi, S., and Tsoukalas, G. (2020). Rethinking crowdfunding platform design: mechanisms to deter misconduct and improve efficiency. *Management Science*, 66(11):4980–4997.

Benhaim, A., Hemenway Falk, B., and Tsoukalas, G. (2022). Scaling blockchains: Can committee-based consensus help? *SSRN Electronic Journal*.

Benhaim, A., Hemenway Falk, B., and Tsoukalas, G. (2023). Balancing power in decentralized governance: Quadratic voting under imperfect information. *SSRN Electronic Journal*.

Breidenbach, L., Cachin, C., Chan, B., Coventry, A., Ellis, S., Juels, A., Koushanfar, F., Miller, A., Magauran, B., Moroz, D., et al. (2021). Chainlink 2.0: Next steps

in the evolution of decentralized oracle networks. https://research.chain.link/whitepaper-v2.pdf.

Buterin, V. (2013). A next-generation smart contract and decentralized application platform. https://cryptorating.eu/whitepapers/Ethereum/Ethereum_white_paper.pdf.

Buterin, V. (2021). Moving beyond coin voting governance. https://vitalik.ca/general/2021/08/16/voting3.html.

Buterin, V., Hitzig, Z., and Weyl, E. G. (2019). A flexible design for funding public goods. *Management Science*, 65(11):5171–5187.

Cai, Y., Fragkos, G., Tsiropoulou, E. E., and Veneris, A. (2020). A truth-inducing sybil resistant decentralized blockchain oracle. In *2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS)*, pages 128–135. IEEE.

Chainlink (2021). What is the blockchain oracle problem? https://chain.link/education-hub/oracle-problem.

Cong, L. W. and He, Z. (2019). Blockchain disruption and smart contracts. *The Review of Financial Studies*, 32(5):1754–1797.

Cong, L. W., He, Z., and Li, J. (2021a). Decentralized mining in centralized pools. *The Review of Financial Studies*, 34(3):1191–1235.

Cong, L. W., Li, Y., and Wang, N. (2021b). Tokenomics: Dynamic adoption and valuation. *The Review of Financial Studies*, 34(3):1105–1155.

Cong, L. W., Li, Y., and Wang, N. (2022). Token-based platform finance. *Journal of Financial Economics*, 144(3):972–991.

Cong, L. W. and Xiao, Y. (2023). Information cascades and threshold implementation: Theory and an application to crowdfunding. *Journal of Finance*, forthcoming.

Cong, W., Prasad, E., and Rabetti, D. (2023). Financial and informational integration through oracles. *SSRN Working paper*.

Dal Bó, E. (2007). Bribing voters. *American Journal of Political Science*, 51(4):789–803.

Dekel, E., Jackson, M. O., and Wolinsky, A. (2008). Vote buying: general elections. *Journal of Political Economy*, 116(2):351–380.

Gans, J. and Holden, R. (2022). Mechanism Design Approaches to Blockchain Consensus. *SSRN Electronic Journal*.

Gibbard, A. (1973). Manipulation of voting schemes: a general result. *Econometrica: Journal of the Econometric Society*, pages 587–601.

Gibbard, A. (1977). Manipulation of schemes that mix voting with chance. *Econometrica: Journal of the Econometric Society*, pages 665–681.

Grossman, S. J. and Hart, O. D. (1988). One share-one vote and the market for corporate control. *Journal of Financial Economics*, 20:175–202. The Distribution of Power Among Corporate Managers, Shareholders, and Directors.

Halaburda, H., He, Z., and Li, J. (2021). An economic model of consensus on distributed ledgers. Technical report, National Bureau of Economic Research.

Han, J., Lee, J., and Li, T. (2023). Dao governance. *SSRN Working Paper*.

Harris, M. and Raviv, A. (1988). Corporate governance: Voting rights and majority rules. *Journal of Financial Economics*, 20:203–235. The Distribution of Power Among Corporate Managers, Shareholders, and Directors.

Hashemi Joo, M., Nishikawa, Y., and Dandapani, K. (2020). Announcement effects in the cryptocurrency market. *Applied Economics*, 52(44):4794–4808.

He, P., Tang, D., and Wang, J. (2020). Staking pool centralization in proof-of-stake blockchain network. *SSRN Electronic Journal*.

John, K., Kogan, L., and Saleh, F. (2022). Smart contracts and decentralized finance. *SSRN Electronic Journal*.

John, K., Rivera, T. J., and Saleh, F. (2020). Economic implications of scaling blockchains: Why the consensus protocol matters. *SSRN Electronic Journal*.

Lalley, S. P. and Weyl, E. G. (2018). Quadratic voting: How mechanism design can radicalize democracy. In *AEA Papers and Proceedings*, volume 108, pages 33–37.

Medium (2021). Price oracle manipulation. https://extropy-io.medium.com/price-oracle-manipulation-d46fd413cc17.

Mueller, D. C. (1973). Constitutional democracy and social welfare. *The Quarterly Journal of Economics*, 87(1):60–80.

Pasdar, A., Dong, Z., and Lee, Y. C. (2021). Blockchain oracle design patterns. *arXiv preprint arXiv:2106.09349*.

Satterthwaite, M. A. (1975). Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of economic theory*, 10(2):187–217.

Tsoukalas, G. and Falk, B. H. (2020). Token-weighted crowdsourcing. *Management Science*, 66(9):3843–3859.

Xu, J. and Livshits, B. (2019). The anatomy of a cryptocurrency pump-and-dump scheme. *28th USENIX Security Symposium*.

Zhang, L. (2023). Robust (decentralized) oracle design. *Working Paper*.

A Proofs

A.1 Proof of Lemma 1

The expected payoff of voter j, given that he votes for conviction is

$$\mathbb{E}(Vote\ Conviction) = 0P(\mathcal{C} \setminus \mathcal{Q}) + 0P(\mathcal{Q}) - DT_i P(\mathcal{A} \setminus \mathcal{Q}) \tag{A.1}$$

Here the payoff is 0, except in case the malfeasant provider is aquitted, in which case the voter faces the drop in token value. Likewise the payoff if the participant votes for acquittal can be written

$$\mathbb{E}(Vote\ Acquittal) = 0P(\mathcal{C} \setminus \mathcal{Q}) - DT_i P(\mathcal{Q}) + (-DT_i)P(\mathcal{A} \setminus \mathcal{Q}) + \sum_{k \in \mathcal{K}} B_{i,k} P(k) \tag{A.2}$$

The difference in expected payoffs between voting for conviction and voting for acquittal is:

$$\Psi_i = 0P(\mathcal{C} \setminus \mathcal{Q}) + (DT_i)\,P(\mathcal{Q}) + 0P(\mathcal{A} \setminus \mathcal{Q}) - \sum_{k \in \mathcal{K}} B_{i,k} P(k). \tag{A.3}$$

Voter i will only vote for acquittal, if $\Psi_i < 0$. To explicitly specify the bribe term, now consider the following dominant nash strategy from the perspective of the malefactor: offer individuals that vote for acquittal a contingent bribe that exactly match the payoff of conviction in every state and additionally comprises an $\epsilon \gtrsim 0$. Formally, such a bribe can be described as follows:

$$B_{i,k} = \begin{cases} \epsilon & \text{if } k \in \mathcal{C} \setminus \mathcal{Q} \\ DT_i + \epsilon & \text{if } k \in \mathcal{Q} \\ \epsilon & \text{if } k \in \mathcal{A} \setminus \mathcal{Q}. \end{cases} \tag{A.4}$$

In words: if the outcome is conviction, the individual who voted for acquittal was not pivotal ($k \in \mathcal{C} \setminus \mathcal{Q}$) will receive ϵ. If the outcome is acquittal and the individual was pivotal ($k \in \mathcal{Q}$), his payoff will additionally include a compensation for the loss in token value (which is assumed to occur in case of a false acquittal). Finally, if the outcome is acquittal

and the individual was not pivotal ($k \in \mathcal{A} \setminus \mathcal{Q}$), the payoff will be ϵ. This is the most efficient dominant nash-strategy, since the payouts of a conviction vote are only exceeded by ϵ.

To see this, simply compare the expected payoffs of the voters:

$$\mathbb{E}(Vote\ Conviction) = 0P(\mathcal{C} \setminus \mathcal{Q}) + 0P(\mathcal{Q}) - DT_i P(\mathcal{A} \setminus \mathcal{Q}) \tag{A.5}$$

$$\mathbb{E}(Vote\ Acquittal) = \epsilon P(\mathcal{C} \setminus \mathcal{Q}) + \epsilon P(\mathcal{Q}) + (-DT_i + \epsilon)P(\mathcal{A} \setminus \mathcal{Q}). \tag{A.6}$$

Evidently, the briber only needs to offer an infinitesimal amount ϵ to ensure $\mathbb{E}(Vote\ Conviction) - \mathbb{E}(Vote\ Acquittal) < 0$. In addition, since it is a dominant strategy for all participants to vote acquittal, $P(\mathcal{A} \setminus \mathcal{Q}) = 1$. Therefore, the total realized costs for the briber will be $n\epsilon$ and therefore arbitrarily close to zero.

A.2 Proof of Proposition 2

This proof is an application of an argument first made by Dal Bó (2007). Under a masked voting scheme, bribes cannot be conditioned on individual votes. This heavily restricts the outcome space $\mathcal{K}$, implying a substantial reduction of the set of bribing strategies available to the service provider. We therefore scrutinize strategies contingent on the voting power (token holdings) of network participants, which is transparent even when the actual votes are masked. To begin with, consider the following feasible strategy. Offering

$$\frac{1}{2}\left(D\sum_{j=1}^{n} T_j + \epsilon\right) \tag{A.7}$$

(with $\epsilon > 0$) to the majority of the voting power overcompensates their aggregate payoff in the conviction state and thus makes acquittal the dominant strategy.

A briber might be tempted to reduce the bribing costs associated with this strategy. He could do so by:

i Offering an individual tokenholder with voting power $\frac{T_i}{\sum_{j=1}^{n} T_j}$ less than DT_i. However, this will induce the tokenholder to vote for conviction.

ii Offering the strategy to participants which together hold less than half of the voting power. However, this will never lead to $P(\mathcal{A}) > 0$.

iii Turning to pivotal bribing strategies.

Due to masked voting, it is not possible for the bribing strategy to target pivotal individuals specifically. However, there might be pivotal strategies on the aggregate level. In this case the briber could offer a bribe only to a subset of voter and just pay them conditional on a pivotal aggregate outcome. To complete the proof, we need to show that this strategy does not work.

To incentivize any individual tokenholder j, the bribe B_j offered for his vote, contingent on a pivotal aggregate voting outcome for acquittal, must exceed the value loss in his token holdings:

$$B_j \geq DT_j. \tag{A.8}$$

Furthermore, the bribe has to be at least as large for any outcome where more than one tokenholder votes for acquittal. If not, tokenholder j votes for conviction. To achieve an aggregate acquittal outcome with such a pivotal strategy, the malfeasant service provider thus still has to bribe at least 50 percent of the overall DAO voting power. Accordingly, the total bribing costs are at least as large as in equation A.7:

$$\mathbf{B} = \sum_{j=1}^{n} B_j = \frac{1}{2}\left(D \sum_{j=1}^{n} T_j + \epsilon \right). \tag{A.9}$$

Now turn to the perspective of the briber. For the DAO to be collusion proof, the cheapest possible bribing strategy (A.7) must be more expensive than the stake of the briber that will be confiscated in case of conviction:

$$S \leq \frac{1}{2}\left(D \sum_{j=1}^{n} T_j + \epsilon \right) \tag{A.10}$$

In a nutshell, the optimal incentive-compatible stake can be implemented and the system is collusion proof, except for a few extreme cases. Those are given, if ceteris paribus,

i the token value response is negligible,

ii the total value of all network tokens $\left(\sum_{j=1}^{n} T_j\right)$ is small,

iii the largest token holdings commanded by any of the participants is dominant relative to the sum of all token holdings, i.e., if the token holdings are highly centralized.

A.3 Proof of Proposition 3

We solve a slightly modified version of **Problem 1** to account for stake adjustments. More precisely we merge the IC and bribing constraint together.

Problem 2

$$\begin{aligned}
&\arg\min_{S} \quad -rS \\
&s.t. \\
&R \geq G - SP(\mathcal{C}|\boldsymbol{B}) - P(\mathcal{A}|\boldsymbol{B})SD - \sum_{k\in\mathcal{K}} P(k|\boldsymbol{B})\left(\sum_{j=1}^{m} B_{j,k}\right) \qquad (IC\&BC) \\
&R - rS \;\geq\; 0 \qquad (PC)
\end{aligned}$$

The expected loss of the stake in case the provider engages in malfeasance is now weighted with the conditional probability of conviction, given the bribing strategy.

Again, a tokenholder will accept the bribe if $\mathbb{E}(Vote\ Conviction) - \mathbb{E}(Vote\ Acquittal)$. This means that a tokenholder j will accept the bribe, iff:

$$\sum_{k\in\mathcal{K}} B_{j,k}P(k) - DT_jP(\mathcal{Q}) - DT_jP(\mathcal{A}\setminus\mathcal{Q}) \geq -DT_jP(\mathcal{A}\setminus\mathcal{Q}), \tag{A.11}$$

i.e., the expected payoff in case the token holder votes for acquittal (left hand side) must be larger than or equal to the expected payoff in case he votes for conviction.

One characteristic of stochastic voting is that a 50% majority is not necessary to achieve consensus. Hence, the briber can decide for each participant individually, if he wants to offer that person a bribe or not. As a consequence, the conviction probability is no longer

binary $P(\mathcal{C}|\mathbf{B}) \in \{0,1\}$ as it was in the case of majority voting.

Stochastic voting implies that voter (token holder) j is pivotal with a **fixed** probability

$$P(Q) = \left(\frac{T_j}{\sum_{j=1}^{t} T_j} \right), \tag{A.12}$$

and a voter accepts the bribe if:

$$\sum_{k \in \mathcal{K}} B_{j,k} P(k) - DT_j \left(\frac{T_j}{\sum_{j=1}^{t} T_j} \right) \geq 0 \tag{A.13}$$

A malfeasant actor decides to offer exactly him a bribe (marginal decision) when:

$$DT_j \left(\frac{T_j}{\sum_{j=1}^{t} T_j} \right) \leq S_i \left(\frac{T_j}{\sum_{j=1}^{t} T_j} \right). \tag{A.14}$$

The left hand side represents the bribe a malicious provider has to offer to win over the token holder times the probability that this bribe will need to be paid. The right hand side is the expected loss of not bribing this token holder.

Condition (A.14) is less likely to be fulfilled for voters with large token holdings, because the left hand side is quadratic in T_j. Under a stochastic voting system, the people most at risk of being bribed are the ones, who only hold small amounts of tokens. A malicious actor can therefore simply target all voters, for which $DT_j \leq S$. As long as at least one such "small" token holder exists, bribing cannot be avoided altogether. This is not a problem, as long as collusion is more expensive in expectation.

Denote with $\mathcal{V}$ the set of token holders, for which (A.14) holds. We can then write the conviction probability $P(C|\mathbf{B})$ as:

$$P(C|\mathbf{B}) = \frac{\sum_{j \notin \mathcal{V}} T_j}{\sum_{j=1}^{t} T_j}. \tag{A.15}$$

Plugging (A.15) into the (IC) of **Problem 1** yields:

$$R \geq G - S\frac{\sum_{j\notin\mathcal{V}} T_j}{\sum_{j=1}^{t} T_j} - \left(1 - \frac{\sum_{j\notin\mathcal{V}} T_j}{\sum_{j=1}^{t} T_j}\right) SD - \frac{\sum_{j\in\mathcal{V}} DT_j^2}{\sum_{j=1}^{t} T_j}$$

Not engaging in malfeasance (left hand side) has to lead to a larger expected payoff than malfeasance. The misconduct outcome (right hand side) consists of multiple parts. The first term represents the expected benefit of malfeasance. The second term consists of the conviction probability times the stake. The third term consists of the probability of not being convicted times the drop in token value of the retained stake. The forth term are the bribing costs under the optimal bribing strategy.

Hence, for sufficiently centralized tokenholder the system becomes collusion-proof.

A.4 Proof of Proposition 4

We are interested in a sufficient condition for a collusion-proof DAO, so we can ignore the first three summands on the right-hand side of (BC). Since

$$\frac{\partial c(m)}{\partial m} > 0, \tag{A.16}$$

we are essentially looking for a lower bound on m. First, we need to understand how m is linked to the concentration of the voting power:

Lemma 2 *The number of DAO members that a briber needs to contact to achieve collusion is inversely related to the degree to which voting power in the DAO is concentrated.*

Proof. We need to show that

$$m = f(H_V), \qquad \frac{\partial m(H_V)}{\partial H_V} < 0. \tag{A.17}$$

The malefactor decides on a bribing strategy,[35] given the degree of concentration of the voting power in the DAO as reflected by H_V. In a DAO with an absolute-majority decision

[35] A bribing strategy comprises the size of the individual bribes ($B_{j,k}$), the number of DAO members that are contacted with the offer (m), and the condition under which bribes will be paid ($k \in \mathcal{K}$ or $k \in \mathcal{A}$).

rule, the minimum number of individuals (including the malefactor) required to achieve a majority vote is:

$$m = min\left(\kappa \mid \frac{\sum_{j=1}^{\kappa} T_j}{\sum_{j=1}^{n} T_j} \geq \frac{1}{2}\right). \tag{A.18}$$

The malefactor needs to secure at least half of the voting power. It is rational for him to choose a strategy that minimizes the amount of people that he needs to contact to achieve this goal.

Now, rearrange (3) as follows:

$$\sum_{j=1}^{n} T_j = \sqrt{\frac{1}{H_V} \sum_{j=1}^{n} T_j^2} \tag{A.19}$$

and insert into (A.18) to obtain:

$$m = min\left(\kappa \mid \frac{\sum_{j=1}^{\kappa} T_j}{\sqrt{\sum_{j=1}^{n} T_j^2}} \sqrt{H_V} \geq \frac{1}{2}\right). \tag{A.20}$$

Evidently, m will be higher, the lower H_V, i.e., the less voting power is concentrated.

■

Next, we determine how the voting power must be allocated to minimize H_V for a fixed $m = \bar{m}$:

Lemma 3 *The lowest possible H_V for a given $m = \bar{m}$ is reached for a uniform distribution of the voting power among the smallest group of network participants that controls least 50 percent of the overall voting power.*

Proof. Consider the following situation: At least 50 percent of the voting power is uniformly distributed among $\bar{m}$ participants. These are the network participants, which the malevolent actor aims to contact. Call a tokenholder from this set a K_1 participant. The rest of the voting power is uniformly distributed among the remaining $(n-\bar{m})$ participants with $(n-\bar{m}) \geq \bar{m}$. Call a tokenholder from this set a K_2 participant. For ease of exposition

and w.l.o.g., now consider the special case where the two groups K_1 and K_2 hold exactly $\frac{1}{2}$ of the voting power:

$$H_V = \bar{m}\left(\frac{1}{2\bar{m}}\right)^2 + (n-\bar{m})\left(\frac{1}{2\,(n-\bar{m})}\right)^2. \tag{A.21}$$

Notice how any deviation from the status quo distribution increases H_V. If, e.g., we increase the voting power of one of the K_1 participants by ϵ at the expense of another K_1 participant, we obtain the following new value H'_V:

$$H'_V = \left(\frac{1+\epsilon}{2\bar{m}}\right)^2 + \left(\frac{1-\epsilon}{2\bar{m}}\right)^2 + (\bar{m}-2)\left(\frac{1}{2\bar{m}}\right)^2 + (n-\bar{m})\left(\frac{1}{2\,(n-\bar{m})}\right)^2. \tag{A.22}$$

The net change is:

$$H'_V - H_V = \left(\frac{1+\epsilon}{2\bar{m}}\right)^2 + \left(\frac{1-\epsilon}{2\bar{m}}\right)^2 - 2\left(\frac{1}{2\bar{m}}\right)^2, \tag{A.23}$$

which can be reduced to

$$H'_V - H_V = \frac{2\epsilon^2}{(2\bar{m})^2} > 0. \tag{A.24}$$

The same argument can be made for any change of the voting power of K_2 participants.[36] Hence, the outlined situation with uniformly distributed voting rights in each of the two groups is associated with the minimal Herfindahl index H_V^{min} for a given network size n (and fixed $\bar{m}$). [37] ■

Finally, we express $\bar{m}$ as a function of a given network size $\bar{n}$ and concentration of the voting power $\bar{H}_V$. From (A.21) we get:

$$\bar{H}_V = \frac{1}{4\bar{m}} + \frac{1}{4(\bar{n}-\bar{m})} = \frac{\bar{n}}{4\bar{m}(\bar{n}-\bar{m})} \tag{A.25}$$

Rearranging delivers a quadratic equation:

$$\bar{m}^2 - \bar{n}\bar{m} + \frac{\bar{n}}{4\bar{H}_V} = 0. \tag{A.26}$$

[36] Moreover, this logic also applies to those cases in which the distribution of voting power between the K_1 participants and the K_2 participants is not 50/50.

[37] Similarly, the maximum Herfindahl index H_V^{max} obtains, if the 50 percent voting power of the K_1 participants is maximally concentrated inside the group, which is substantively no different from the case where $\bar{m} = 1$.

We are interested in the following root of (A.26):

$$\bar{m} = \frac{\bar{n}}{2} - \frac{1}{2}\sqrt{\bar{n}^2 - \frac{\bar{n}}{\bar{H}_V}}. \tag{A.27}$$

For the specific the network size $\bar{n}$ and degree of concentration $\bar{H}_V$, the service provider needs to contact and bribe at least $\bar{m}$ participants to be acquitted. Therefore, (A.27) describes a lower bound for m:[38]

$$m \geq \frac{n}{2} - \frac{1}{2}\sqrt{n^2 - \frac{n}{H_V}}. \tag{A.28}$$

[38]Evidently, the lower bound is inversely related to H_V. From **Lemma 3** we know that, for a given number of network participants n, the smallest $H_V (= H_V^{min})$ and therefore the largest lower bound for m is associated with a uniform distribution of voting rights (in the each of the two groups K_1 and K_2).