Auer, Raphael A.; Monnet, Cyril; Shin, Hyun Song

Working Paper

Distributed Ledgers and the Governance of Money

Working Paper, No. 21.01

Provided in Cooperation with:
Study Center Gerzensee, Swiss National Bank

Suggested Citation: Auer, Raphael A.; Monnet, Cyril; Shin, Hyun Song (2021) : Distributed Ledgers and the Governance of Money, Working Paper, No. 21.01, Swiss National Bank, Study Center Gerzensee, Gerzensee

This Version is available at:
https://hdl.handle.net/10419/247258

Distributed Ledgers and the Governance of Money

Raphael Auer, Cyril Monnet and Hyun Song Shin

Working Paper 21.01

Distributed Ledgers and the Governance of Money*

Raphael Auer — BIS

Cyril Monnet — University of Bern and SZ Gerzensee

Hyun Song Shin — BIS

November 26, 2021

Blockchain technology breathes new life into the classical analysis of money as a substitute for a ledger of all past transactions. While it involves updating the ledger through a decentralized consensus on the unique truth, the robustness of the equilibrium that supports this consensus depends on who has access to the ledger and how it can be updated. Using a global game analysis of an exchange economy with credit, we solve for the optimal ledger design that balances security, scalability and decentralization. When intertemporal incentives are strong, a centralized ledger is always optimal. Otherwise, decentralization may be optimal.

JEL Codes: C72, C73, D4, E42, G2, L86.

Keywords: market design, money, distributed ledger technology, DLT, blockchain, decentralized finance, global game, consensus.

*We thank an anonymous referee of the BIS Working Paper Series, Joseph Abadi, Aleksander Berentsen, Aldar Chan, Francesca Carapella, Jon Frost, Rod Garratt, Piero Gottardi, Hans Gersbach, Hanna Halaburda, Ricardo Lagos, Jacob Leshno, Dirk Niepelt, Jean-Charles Rochet, Harald Uhlig, participants of the 2021 CBER Forum, CB&DC Virtual Seminar Series, 2020 CEBRA annual meeting, the 2020 Summer Workshop on Money and Payments, the 2020 ETH-Zurich Workshop on Future Money and seminar participants at the BIS, Yale University and the University of Zurich for comments. The views presented in this paper are those of the authors and not necessarily those of the Bank for International Settlements.

1 Introduction

Money is a social convention. People accept money in payments in the expectation that others will do so in the future. Within an equilibrium with monetary exchange, holding money is a record of goods sold or services rendered in the past. In this sense, money serves as a record-keeping device. In this spirit, Kocherlakota (1998) and Kocherlakota and Wallace (1998) showed in a simple setting that the social convention of money does almost as well as when agents have free access to a complete ledger of all past transactions in the economy. The motto is that "money is memory."

Advances in cryptography and digital technology have opened up the possibility of taking the idea of a complete ledger of past transactions more literally than just as a theoretical construct, and building a monetary system around such a ledger.

However, while advances in technology have brought the full ledger within the reach of every economic agent, there remains the crucial question of who should have the authority to update the ledger and how. After all, the economic value of the ledger is precisely that everyone carries around identical copies of the ledger. The decentralization agenda, as exemplified by the blockchain, is that this authority should no longer be in the hands of a central authority (eg, a central bank), but that it should be replaced by the consensus of network members. Truth is whatever is deemed to be so by the consensus of network members.

Distributed ledger technology (DLT) is an updating mechanism for the ledger that rests on a set of rules for network members that elicits the decentralized consensus on the unique, true history, without appeal to a trusted authority. Once consensus is reached, all network members then coordinate on using the unique, agreed version of the ledger. In this respect, decentralization evens up the playing field and equalizes the distribution of power.

The goal of decentralization is to prevent reaching consensus on anything but the correct history of transactions.The system is centralized when just one node has the authority to update the ledger and fully decentralized when all nodes in the system take part in updating the ledger. In between these two extremes, the system is "permissioned" when only a pre-

selected set of nodes can update the ledger (see Townsend (2020) for an introduction).[1] The consensus mechanism should specify rules that exclude coordination on all but the true state of the world. The idea is that decentralization makes the consensus mechanism more robust because the reliance on many nodes make it difficult that a conflicting entry will be accepted by all nodes at once.[2]

However the larger the ledger is and the more nodes there are in a network, the harder it becomes to update the ledger quickly. In short, the ledger is hard to scale. This conundrum introduces trade-offs sometimes known as the ledger's "scalability trilemma" (see ie Buterin (2021), and Figure 1). The trilemma is posed in terms of the challenge of attaining a ledger that is simultaneously decentralized, secure, and scalable.[3]

In view of the trilemma, the task is to find the sweet spot between decentralization, security and scalability. What is the optimal solution to the trade-offs involved? We address this question in what we believe is a first economic analysis of DLT in a monetary economy.

Our approach is one based on economics and game theory. We abstract from the details of the computing or cryptographic implementation and focus, instead, on the economic incentives underlying the consensus process. Concretely, we focus on the incentives facing the economic agents in fulfilling their tasks as validators whose role is to agree on the unique set of executed trades that will be written in to the ledger.

Our model has three building blocks, each addressing one side of the triangle. The first modeling block consists of an intertemporal model of exchange involving credit. This block is the basis to study the scalability of the ledger. The mechanism is scalable whenever

[1] A ledger is only fully decentralized if it allows free entry into validation – we analyze this case for completeness in Appendix E. The literature on the economics of Bitcoin and other anonymous cryptocurrencies is growing large (see i.e. Abadi and Brunnermeier (2018), Budish (2018), Koeppl and Chiu (2018), Schilling and Uhlig (2018), Auer (2019), and Biais et al. (2019)).

[2] For example, the consensus mechanism should sett out the detailed procedure on how to choose one ledger over another if a conflict were to arise between multiple versions of a ledger. Biais et al. (2019) show in the context of the Bitcoin's blockchain that in equilibrium members will coordinate on the longest chain.

[3] Buterin (2021) also discusses "sharding" as a solution to the blockchain trilemma, which is akin to parallel computing where different tasks of one single program are distributed on different nodes. However, even absent technical scale limitations, economic incentives limit the amount of trading: we below model a DLT without scale limitations, but still find the volume of trade to be limited as a higher volume of trade requires higher rents for the validators.

trades are at their optimal levels. Our economy has two types of infinitely lived agents, early and late producers. In each period, an early producer is randomly matched with a late producer and the pair engages in two subsequent production stages. In the early stage, the early producer produces goods for the late producer. In the late stage, the late producer should reciprocate and produce some goods for the early producer. We impose two main frictions on producers. First, there is private information: late producers can be faulty and cannot produce but early producers cannot tell the difference between faulty and other producers. Second, late producers cannot commit to reciprocating. Therefore, there is no trade unless a record-keeping device – the memory of our economy – tracks the actions of late producers and, in particular, whether the late producer has ever defaulted in the past. Just as in Kocherlakota (1998), rather than users owning and paying with monetary tokens, the ledger's memory of the production history suffices to allow for trustless exchange: it is well known (see e.g. Rocheteau and Nosal, 2017) that there is an equilibrium with trade when the trading history of a late producer is publicly and freely observable and automatically updates itself according to the behavior of the late producer. In turn, since the ledger's memory is the essential value underpinning of the economy, ensuring its integrity is the quintessential design issue to be solved.

The second modeling block endogenizes the process of updating the history of trades, that is the validation of records on the ledger and consensus as an equilibrium outcome. This block is the basis to analyze security, that is how strong consensus is. We assume that a number of agents known as "validators" are in charge of reading and updating the ledger of trade histories. In blockchain parlance, validators form the network of nodes that store and exchange the data file that constitute the ledger. Validators can be users of the credit system in which case we say the validation is internal, as opposed to external validation when validators are not users of the system. For each trade involving a late and early producer, some (enough) validators have to verify the history of the late producer and communicate the result to the early producer. The history is understood to be "good" (that is, without default) whenever a supermajority of validators say it is so. We model the strength of the security around the integrity of the ledger as the supermajority threshold for agreement that is needed to validate a new block in the chain. The most secure rule is to insist on unanimity,

so that everyone agrees. However, insisting on unanimity could be a recipe for gridlock and delay. We capture these inefficiencies by assuming that verifying histories has a known common cost, while the cost of communicating the result is idiosyncratic, reflecting, e.g. the possibilities of operational failures for some validators. Validators privately learn their cost of communicating histories, which consists of a common component and an idiosyncratic one. Since verification and communication are both costly activities, validators must be compensated for their efforts and they will expect a payment from the pair of early and late producers whose history they have to validate. Since validation requires a supermajority of validators, this structure gives rise to a global game that we analyze using the approach of Morris and Shin (1998, 2003). As we explained above, validators cannot be trusted, because (1) they cannot commit to verifying histories, so while messages sent by validators are observable, their checking is a costly non-observable action which raises a moral hazard problem and (2) they can accept side payments to record a false entry. These frictions imply that reaching consensus as an equilibrium is difficult.

Our third and final modeling block is the analysis of the optimal degree of decentralization, as a solution to the optimal design of the trade and validation mechanisms, where optimality is defined as the surplus from the trade net of the validation costs. The optimal mechanism chooses the number of validators, the supermajority threshold, the compensation of validators, as well as the trade allocation that maximize the gains from trade subject to incentive compatibility conditions.

Preview of results In this context, we can pose Buterin's scalability trilemma formally within our monetary economy. Scalability is only achieved by abandoning decentralization, while decentralization and security comes at the cost of scalability. Achieving all three goals is ruled out in equilibrium. The optimal solution that balances decentralization, scalability and security depends on the specific frictions that operate in the economy and their severity. We lay out the considerations for choosing the socially optimal point on the triangle that finds the sweet spot amid the trade-offs involved. Our main result is the derivation of a mapping from the terms of the trade-offs involved to the socially optimal point on the

Buterin triangle.[4]

A decentralized consensus mechanism sets up a public good contribution game, where greater security demands a higher supermajority threshold for the successful provision of the public good – a clean, reconciled ledger that everyone accepts. However, each validator needs to perform their assigned task of verifying the transactions. This task entails a small cost for each validator, and this cost is assumed to vary slightly across validators. These individual costs in performing the tasks necessary for consensus inject strategic uncertainty and set up a global game for the public good contribution game. Unless the rewards that accrue to the validators are sufficiently high, the validation protocol may not be followed in equilibrium. This is especially so when the stringency of the supermajority threshold is very high. Thus, a direct consequence of aiming for stronger security is that the validators need to be given a larger share of the social surplus in the form of rents. In turn, the high rents that the validators extract reduce the overall size of the pie in terms of the economic gains that arise from monetary exchange. Rents undermine scalability, but rents are necessary for security. We thus end up with a trade-off along one dimension of Buterin's trilemma.

Rents to validators can be reduced (and scalability enhanced) by moving away from the requirement of full decentralization. The extreme case is to rely on a single, trusted validator. If the single validator node can be trusted with managing the ledger, then security and scalability can be achieved. However, if the single validator cannot be trusted, better governance calls for the checks and balances that only a larger committee of validators can deliver. That nodes need to coordinate on a single ledger acts as a discipline device for each one of them. However, this brings us back full circle, as the robustness of the voting equilibrium requires that nodes be incentivized with sufficient rents.

We characterize the optimal degree of decentralization, security, and scalability by solving for the optimal trade size, supermajority, and number of validators, as well as whether internal or external validation is optimal. Naturally, the optimal solution is constrained by the incentives of late producers and validators. Validators should get a reward for validating transactions

[4]We thus present what we believe is a first economic analysis of permissioned DLT in a monetary economy. A notable exception - discussed below - is Amoussou-Guénou et al. (2019).

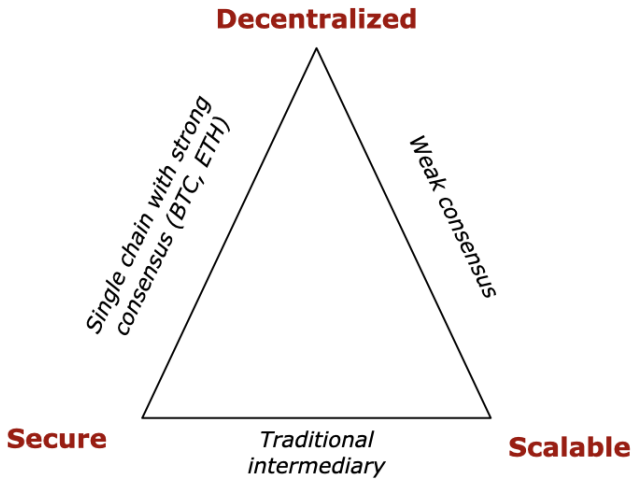


Figure 1: Buterin's trilemma

and that reward should be high enough to deter them from accepting a bribe. Therefore intertemporal incentives are key to characterize the optimal solution. When intertemporal incentives are strong in the sense that the present values of future rewards are high, validators can be trusted as they would have much to lose from accepting a bribe. In this case, a single validator who earns a large rent can be entrusted with managing the ledger. This validator can indifferently be a user of the credit system or not, and the size of each transaction is at the first best level.

The solution is more interesting when intertemporal incentives are low: Then it becomes too costly to prevent a single validator to accept a bribe. There are two solutions: either increase the number of validators or reduce the size of each transaction to lower the bribe size. The optimal solution plays on both margins. However, validators now need to reach a consensus according to some supermajority rule. Given their private costs, validators play a game that has the attributes of a public good provision game – the public good is provided if and only if a supermajority provides it – which we proceed to solve using global game methods (see Carlson and van Damme, 1993 and Morris and Shin, 1998, 2003). We show that validators can reach a given level of consensus as a unique, dominance solvable equilibrium if and only

if they earn a large enough reward that is above some threshold. Naturally, that threshold is increasing with the average cost of validation. Also, validators can reach a higher level of consensus – i.e. a higher supermajority – if they obtain higher rewards. Therefore, our model has the feature that consensus cannot be reached without validators earning rents (see also Abadi and Brunnermeier (2018)).

It is rather intuitive why decentralized consensus fails to be sustained when the rents accruing to validators falls below the threshold, even when validation would be a possible equilibrium in a complete information game. The reason is that the uncertainty surrounding the fundamental communication cost can reverberate throughout the validation process: validators may choose to abstain from validating a trade when, given their private cost, they believe that other validators will also abstain. However, there is an equilibrium where validators validate a trade as long as the (expected) reward is large enough relative to their (realized) private costs. Driving the idiosyncratic component of the cost to zero, we find the validation process "works" whenever the fundamental communication cost is below that level. In turn, this gives the probability that the validation process will be successful. That success probability falls with the supermajority threshold, but increases with the payments validators obtain. Therefore a higher supermajority requires higher rewards to validators in order to guarantee the same success probability. In other words, reaching a higher level of consensus among validators requires higher rents to be paid to validators.

Therefore, when intertemporal incentives are weak, decentralized validation by many validators is optimal and we can solve for the optimal supermajority threshold and the size of the transaction. Optimally, consensus is weaker and the transaction size is smaller as intertemporal incentives weaken. We also show that only internal validation can decentralize trade, while there is no trade with external validation.

Our findings suggest a number of initial conclusions. While it is costly to duplicate verification and communication across many validators, we find conditions under which many validators are better than one. To use Aymanns et al.'s (2020) terminology, we find conditions under which a (trading) platform should be vertically disintegrated – a group of agents should handle the interaction between users – rather than vertically integrated, when

a single intermediary has the monopoly over managing the interaction of the platform users. A central result is that there are economies of scope in trading and validation: achieving good governance and honest record-keeping is made easier by having validators who also participate in the market themselves and thus have an intrinsic interest in keeping it going smoothly. This also implies that validators should more often than not be selected from the set of market participants.

Our results on the supermajority are naturally dependent on the communication cost being stochastic and unknown. When the communication cost is common knowledge, unanimity is optimal and, to reduce the incentive to bribe validators, they should be sufficiently many. Hence it is typically sub-optimal to have a central validator whenever they have to satisfy incentive constraints and the communication cost is common knowledge.

However, using a mechanism with many validators gives rise to a free-rider problem in solving information frictions, similar to the one in the seminal paper of Grossman and Stiglitz (1976). Since verifying a label is costly, we show in the Appendix that, under some conditions, validators have an incentive to skip verification while still communicating a good label, which jeopardizes the legitimacy of the whole ledger. To resolve this free-rider problem and maintain the integrity of the ledger, we show that (absent a unanimity rule where all validators are pivotal) the allocation of validators should be dependent on which label they communicate to the ledger and how their communication compares with the supermajority. When the label they communicate differs from the supermajority (which is observable and verifiable), validators should be excluded from trading and validating in the future. In this context, we derive a folk's theorem of sort for validators; as validators become more patient, the free-rider problem has no bite and any allocation satisfying the validators' participation constraint can be implemented in our strategic set-up.

Relation to the literature A sizable literature analyzes the incentives of miners in Bitcoin and similar cryptocurrencies to follow the proof-of-work protocol.[5] Kroll et al (2013)

[5]The variant with staking one's cryptocurrency holding on the truth instead of costly computation, i.e. proof-of-stake, is attracting increased attention (see Abadi and Brunnermeier 2018, Saleh 2021, and Fanti et al. 2020). However, proof-of-stake can also be attacked via so called "long-run attacks" (see Deirmentzoglou

and Prat and Walters (2020) examine free entry and the dynamics of the "mining" market,[6] while Easley et al (2019) and Hubermann et al. (2021) examine the economics of the transaction market. Abadi and Brunnermeier (2018), Budish (2018) and Chiu and Koeppl (2019) show that ensuring the finality of transactions in Bitcoin is very costly as so-called "majority" or "history reversion" attacks are inherently profitable, while Auer (2019) examines whether the transaction market can generate sufficient miner income to ensure the finality.[7] Further to this, even in the absence of incentives to reverse history, sunspot equilibria can arise in proof-of work based blockchains (Biais et al 2019).[8]

The literature on validator incentives and design of permissioned versions of distributed ledgers is sparser.[9] Townsend (2020) focuses on an economics-based approach to the issue of distributed ledgers, exploring novel contracting possibilities enabled by DLT. In recent work, Bakos and Halaburda (2021) model an exogenously imposed fine on the validators in the case of an attack on a permissioned DLT system. In the analysis below, we instead derive this punishment endogenously as the loss from future expected rents as validator and participant in the system. Most closely related to our analysis is Amoussou-Guénou et al. (2019), who first modeled the interaction between validators as a game entailing non-observable effort to

et al. 2018 for a survey). Therefore, proof-of-stake implicitly assumes the existence of some overarching social coordination (see Buterin, 2014).

[6]See also Cong et al. (2019) for an analysis of the concentration of mining and efficiency.

[7]Such attacks are outlined in Nakamoto (2008). In these, the majority of computing power is used to undo a transaction in the blockchain by creating an alternative transaction history that does not contain the transaction. It is noteworthy that other attacks on cryptocurrencies are possible, including the possibility of "selfish" mining analyzed in Eyal and Sirer (2014). Gervais et al. (2016) present a dynamic analysis of the costs and benefits of various attack vectors. Garatt and van Oordt (2020) examine the role of fixed cost of capital formation for the security of Proof-of-Work Based Cryptocurrencies, Böhme et al. (2015) and Schilling and Uhlig (2019) present discussions of broader economic implications and governance issues, respectively. Leshno and Strack (2020) present a generalization of such analysis, demonstrating that no other anonymous and proof-of-work-based cyrptocurrencies can improve upon the performance of Bitcoin. This can serve as a benchmark for the analysis at hand when comparing permissioned and permissionless market designs.

[8]See Carlstens et al. (2016) for a related argument based on simulations and Pagnotta (2021) for an examination of multiple equilibria in the presence of a feedback loop between blockchain security and cryptocurrency valuation. Halaburda et al. (2021) recently examined possible equilibria and their robustness if rational validators can send messages to selected recipients only or even send conflicting messages to different recipients.

[9]Applications of permissioned DLT are being explored for securities settlement systems, trade finance solutions, "stablecoins", and central bank digital currencies, see also Baudet et al. (2020), Arner et al. (2020), Auer et al. (2020), and Chiu and Koeppl (2019). In the Appendix, we describe the main features of permissioned DLT that we think any model of permissioned DLT should capture.

check transactions and costly voting. They also analyzed this game in terms of moral hazard and public good provision. Relative to their analysis, our contribution is to link the ledger validation game to monetary exchange, establish the uniqueness of the equilibrium via a global game approach, and characterize the optimal mechanism design, in particular in terms of the number of validators, size of transactions, and optimal supermajority voting threshold. In our work, all validators are profit-seeking, and the issue at heart is how the market can be designed so that profit-seeking validators actually verify the ledger and validate only correct histories.[10] In a related context, Halaburda et al. (2021), examine the incentives to follow the protocol using communication games and focussing on possible equilibria and their robustness if rational nodes can freely send messages to selected recipients only. The focus on dealing with free-riding and coordination relates to several classical strands of papers on the coordination with many actors. Reminiscent of Grossman and Stiglitz (1976), free riding can prevail in the case of multiple validators. Consistent with Biais et al. (2019) and Amoussou-Guénou et al. (2019) we also derive a folk theorem.

More narrowly, in the context of existing applications in decentralized finance, our model shows how the so-called "Oracle Problem" can be solved via incentive design. On platforms such as Ethereum, oracles serve as reference points for external information – such as asset prices, interest rate benchmarks, or other relevant variables such as the official inflation rate (See Xu et al. (2016) for an introduction to oracles). Since such information is used as an input to calculate the pay-outs of self-executing (i.e. smart) financial contracts, oracles are easy targets for manipulation.[11]

Our paper also has ramifications in the banking literature, starting with Diamond (1984) or Williamson (1986, 1987) where banks are modeled as a way to save on monitoring costs. Another approach, pioneered by Leland and Pyle (1977) and developed by Boyd and Prescott (1986) models banks as information-sharing coalitions. Gu et al. (2016) show that higher rents can discipline intermediaries, while Huang (2019) uses that model to study the optimal number of intermediaries when they have an incentive to divert deposits. A related analysis

[10]Note that Amoussou-Guénou et al. (2019) do not examine history reversion attacks; rather, byzantine attackers are assumed to attempt bringing the system to a halt for exogenous reasons.

[11]See for example Luu et al. (2016) and Froewis and Boehme (2017).

that focuses on the optimal composition of the money stock between inside and outside money can be found in Monnet (2006), Cavalcanti and Wallace (1999a, b), and Wallace (2005). Global games techniques have also been introduced in the banking literature to study the probability of a bank run occurring, eg by Rochet and Vives (2004) and Goldstein and Pauszner (2005).

In game theory, the literature on incentives with public and private monitoring is large and it is beyond the scope of this paper to summarize it (see Kandori, 2001 for an early survey and Rahman (2012) for a problem of private monitoring where the observation of the monitor(s) is not verifiable.)[12] In political science, the literature on the optimal committee size under different decision rules is large (see e.g. Gersbach et al. (2020) for a recent contribution and the references therein).

Section 2 lays down the basic set-up and characterizes benchmark allocations absent a record-keeping device and a freely accessible one. Section 3 defines incentive feasible allocation with DLT, and characterizes the optimal allocation including the optimal number of validators.

2 The model

Our model builds on Gu et al. (2013).[13] Time is discrete and infinite. $\beta \in (0,1)$ is the discount factor. Each period is divided in two distinct production/consumption stages, early and late with one good per stage, the "early good" and the "late good". Goods are non-storable across stage or across periods. There is a continuum of agents. Agents can be of three types, which are permanent. There is a measure one of early producers, a measure $1-f>0$ of late producers, and a measure f of faulty producers. Early producers cannot distinguish between late and faulty producers. Early producers can produce the early good that late and faulty producers like to consume. Late producers can produce the late good

[12] Rahman (2012) shows that sending a false positive to test the "attention" of the monitor can be optimal. In his words, "The principal allocates private information to provide incentives." While the principal in Rahman (2012) knows when the false positive is sent our planner does not know if the match involves a faulty producer (the false positive in our setup).

[13] Gu et al. (2013) borrows methodological elements from Lagos and Wright (2005). See also Williamson and Wright (2011), and Lagos et al. (2017).

that early producers like to consume. Faulty producers do not produce and there can only be gains from trade between early and late producers.

Preferences of early and late producers are represented by the following utility function, respectively[14]

$$\begin{aligned} U_e(x^e, y^e) &= x^e - y^e \\ U_\ell(x^\ell, y^\ell) &= u(x^\ell) - y^\ell \end{aligned}$$

where x^e (resp. x^ℓ) is the consumption of early (resp. late) producers, and y^e (resp. y^ℓ) is the production of early (resp. late) producers. The function $u(.)$ is continuous, increasing, concave, and $u(0) = 0$. We assume that there are gains from trade between early producers and productive late producers. That is, there is x such that $u(x) > x$. We denote by x^* the efficient allocation that solves $u'(x^*) = 1$. Finally, faulty producers derive utility ρx^ℓ from consuming x^ℓ of the late good. We assume $\rho < 1$ so there are no gains from trade between early producers and faulty producers.

Early and late producers meet pairwise at the start of the early production stage. The matching technology is such that nature selects a measure α of early producers, a measure $\alpha(1-f)$ of late producers and a measure αf of faulty producers. Nature also matches one of the early producer with either one of the late or faulty producers. All other producers remain unmatched for the period. Therefore, the probability of a match for any producer is α. The probability that a match involves a faulty producer is f and with complementary probability, the match involves a late producer. The match is maintained across both stages but it dissolves at the end of the later stage. Since being faulty is private information to that producer, a faulty producer will mimic a productive one. As a consequence, we can concentrate on allocations in matches that involve early and late producers only.

Feasibility and efficiency require that what is produced in each match is consumed, ie $x^e = y^\ell$ and $x^\ell = y^e$. Therefore, we can conveniently drop indices and use $x \equiv x^\ell = y^e$ and $y \equiv x^e = y^\ell$. Hence, an allocation is (x, y) where x denotes the production (consumption) of early

[14]Linear utility function for one of the agents (here early producers) allows us to get clean comparative statics, as would do quasilinear utility functions like $x^e - v(y^e)$.

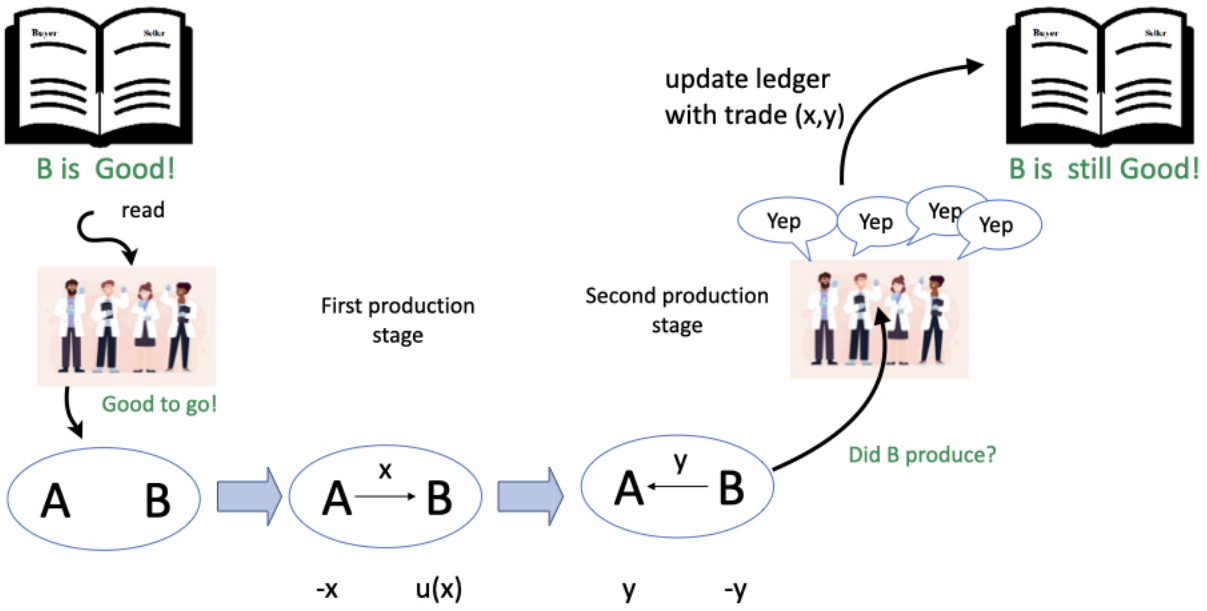


Figure 2: Timeline

(late) producers, and y denotes the production (consumption) of late (early) producers. In the below analysis, we will concentrate on symmetric and stationary allocations. Figure 2 sketches the timeline of our economy.

We can now define a trading mechanism. We restrict trading mechanisms to be in the class of coordination games: In each match the two agents announce a pair $(\tilde{x}, \tilde{y}) \in \mathbb{R}^2_+$. If both choices coincide, the early producer produces $x = \tilde{x}$ in the early stage and the late producer produces $y = \tilde{y}$ in the late stage (faulty producers never produce). In this case we say that the allocation (x, y) can be implemented.

If agents can commit, the set of allocations (x, y) that can be implemented is solely defined by the participation constraints of early and late producers. Late producers participate if and only if $u(x) \geq y$, while early producers participate if and only if $(1 - f)y \geq x$ since they can only consume if the producer is not faulty. Therefore all allocations (x, y) such that $(1 - f)u(x) \geq x$ can be implemented. If there are too many faulty producers and f is too large, then the efficient allocation $x = y = x^*$ is not implementable.

In addition to hidden state, we will assume late producers only have a limited ability to commit.[15] Absent commitment and any record-keeping technology, it is routine to show that late producers will never repay early producers and the only implementable allocation is autarky $(x, y) = (0, 0)$.

A ledger technology

To discipline late producers and to tell them apart from faulty producers, it is necessary to record the history of trades of these producers in a ledger.[16] Such a truthful record helps to discipline late producers by threatening the loss of future consumption in case they do not produce today (see Kocherlakota, 1998). Since the *actions* of early and late producers x_t and y_t may differ from their *words* $\tilde{x}_t$ and $\tilde{y}_t$, the ledger records the result of the coordination game as well as the actions of both producers for all pairs of early and late producers in all period t. The ledger records $\emptyset$ in case producers are not matched. Histories of producers can be conveniently summarized using two labels: good (G) or bad (B). We will also say that late producers can be in good or bad standing. A producer will be assigned label B whenever his actions differed from his words sometime in the past, irrespective of how long ago it was. Otherwise, a producer will be assigned label G. Notice that label B is an absorbing label and an agent carrying label B will never consume or produce.[17]

Then, an allocation (x, y) is implementable if it is incentive feasible (IF). That is, the allocation satisfies participation constraints and late producers have the incentive to repay. These two participation constraints are $u(x) \geq y$ and $y \geq x$, while late producers will have the

[15] For the sake of symmetry, we can also assume that early producers are unable to commit, but their incentive problem is straightforward because, if they do not produce, the late producer will immediately retaliate and will not produce either.

[16] As Ricardo Lagos pointed out to us, the identity of producers in the model is known so that a producer can be excluded from using the ledger. This is important as in BTC only wallet addresses are known and not their true owners. To be clear, here we think the ledger is set up by producers and their unique digital signature gives access to the system. No further digital signatures will be admitted to the system once it is launched, but they can be invalidated in case of misbehaviour.

[17] When an agent has label B, early producers rationally expect that either the agent is a faulty producer or the agent has defaulted in the past, and therefore announces $(\tilde{x}, \tilde{y}) = (0, 0)$ in the coordination game. Anticipating that all agents in the future will announce $(\tilde{x}, \tilde{y}) = (0, 0)$, an late producer with label B will not produce.

incentive to repay early producers if their "repayment constraint" holds,

$$-y + \beta\alpha\frac{u(x) - y}{1 - \beta} \geq 0.$$

If late producers do not produce, they are excluded from the economy, so the right-hand side of the constraint is zero. If they produce, they incur the production cost $-y$, then they are assigned a good label and can participate in trade in the future. The expected value of having a good label is equal to the lifetime gains from trade times their probability of trading α. Setting $y = x$, the set of IF allocations is summarized by x such that

$$\beta\alpha u(x) \geq (1 - \beta(1 - \alpha))x.$$

It should be clear that the efficient allocation x^* is implementable if β and α are large enough.

So far, our analysis has been routine because we have taken the functioning of the ledger as given. The objective of our paper is to endogenize the ledger updating process and explain the incentives problem arising from the mechanism used to update the ledger.

3 Permissioned Ledgers

We now endogenize the updating process of the ledger. We assume the ledger is managed by a measure V of "validators." Validators are rational agents entrusted with validating transactions and updating the ledger, who need incentives to verify transactions and update the ledger honestly. The resulting history should be trusted by all producers.

We consider two types of validator: In period 0, a measure V of validators can be selected from the set of late producers (internal validation) or from a set of agents who, as faulty producers, only consume both goods indifferently (external validation).[18] Internal validators can trade, but they cannot validate their own trades. Each period, all V validators work on validating each of the α match.[19] We assume a validator uses the same strategy across all

[18]The latter could include early producers as well as faulty producers.

[19]When validators are selected from late producers, $V \leq 1 - f$. Since there is a continuum of validators,

matches, as described below. Therefore, a validator either works on validating all matches or none.

There are two validation stages: in the early stage and in the late stage. In the early stage, the validation process consists of validating the label of producers (late and faulty). To validate the label, validators have to 1) verify the label of producers, and 2) send a message to the ledger. Validator i sends message $m_i^1 \in \{\emptyset, 0, 1\}$, where $\emptyset$ means the validator does not send a message, 0 means the validator sends message B and 1 means the validator sends message G. A consensus is reached and is communicated to the early producer whenever at least a fraction $\tau \in [0, 1]$ of validators send the same message, i.e. cast the same vote.

In the late stage, the validation process consists of confirming that late producers have produced according to plan, i.e. $\tilde{y} = y$.[20] To confirm production, validators have to 1) verify whether production took place according to plan, and 2) send a message to the ledger – if production took place according to plan and the late producers had label G, validators will communicate G, but B otherwise. We denote the message of validator i in the late stage as $m_i^2 \in \{\emptyset, 0, 1\}$. Again, a consensus is reached on the new label which is recorded on the ledger whenever at least τV validators cast the same vote. We emphasize that τ is a choice variable when designing the consensus algorithm.[21]

Validators incur verification and communication costs. We assume in the early stage that validators incur a linear utility cost $c_v \geq 0$ to verify the label of a producer and an idiosyncratic linear utility cost $c_{s,i}$ to send a message to the ledger (or to enter their information on the ledger). Since it is costly to send messages, we assume validators only send a message when they want to communicate that the late producer's label is G.[22]

In order to model the possibility of computer glitches and operational failures, we assume

there is also a measure V that will work on validating the trade involving a validator. In this sense, all statements should be qualified with "almost surely."

[20] In practice, this is when double-spending can happen.

[21] We assume the threshold τ is the same in the first and the second stage, but our analysis extends to cases where it differs across the stages.

[22] For tractability, we allow the communication costs to be negative so that we can use symmetric distributions with no mass point. The results would also obtain when we use distributions with a mass point at zero. We assume validators do not incur costs in the second stage. It is straightforward to extend the model to analyze this case too.

that the private cost of communicating a label $c_{s,i}$ that takes the form

$$c_{s,i} = c_s + \mu_i,$$

where c_s is a common component to all validators, while μ_i is the idiosyncratic element for validator i. The idiosyncratic element μ_i is uniformly distributed over the interval $[-\varepsilon, \varepsilon]$, where ε is a small positive number. For any two distinct validators $i \neq j$, μ_i is independent of μ_j. Finally, we suppose that c_s itself has a uniform ex ante distribution over $[\underline{c}_s, \bar{c}_s]$. Validators learn their cost ahead of the verification game and so ahead of verifying the label.

As Morris and Shin (2003) show, the key to the analysis is the characterization of the strategic uncertainty faced by players. Even if the idiosyncratic component is small relative to the other payoff parameters in the frame, the relative ranking of the costs injects strategic uncertainty in the coordination game. Even if remote, the possibility of computer glitches will imply that validation should not rely on unanimous agreement when there are many validators. In the sequel, the reader should think of $\varepsilon \to 0$.

To bring agents to become validators, they must make a positive expected profit from the validation process. Sending a message is verifiable, so validators who correctly sent message G are entitled to z^s units of the good in stage $s = 1, 2$ whenever enough validators agree that the label is G. Validators receive nothing if they do not send a message to the ledger (or if they send a bad message) and they cannot work in the late stage if they have not sent a message to the ledger in the early stage. We assume that validators receive these transfers at the end of stage 2 – once the dust settles – and validators value these transfers in an additive and linear way.[23] When a fraction $w^s \geq \tau$ of validators have sent a good message in stage $s = 1, 2$, early producers in the early stage have to produce $w^1 V z^1$ for each of the $w^1 V$ validators to get z^1, and symmetrically, in stage 2 a late producer has to produce $w^2 V z^2$ for

[23] A linear utility function allows us to abstract from possible insurance mechanism among validators. Also, we could assume that validators only consume the late good. In this case, the late producer produce $y + V z^2$, and the early producer consumes $y - V z^1$ and each validator collects $z^1 + z^2$. Since early producers are risk-neutral, it turns out this is equivalent to validators consuming both the early and the late goods. Also, at the cost of simplicity, we could assume that the utility of validators is $u(x + z^1 + z^2)$. Finally, we could assume that only a share τ of validators receive a reward in stage $s = 1, 2$ since only this number is necessary to reach an agreement.

each of the w^2V validators to get z^2.

Late producers can bribe validators to send a false message: A late producer who has label G when starting the period may get away with not repaying the early producer while keeping its label by making a side payment to τV validators in the late stage (after consuming in the first stage).[24] Validators who accept a bribe are caught with probability π in which case they lose the privilege to trade and validate in the future.[25]

3.1 Payoffs and incentive feasible allocations

Given a threshold τ and a measure of validators V assigned to validate each match, a stationary allocation is a list (x, y, z^1, z^2). An allocation is incentive feasible if it is feasible, it satisfies the incentive constraints of early and late consumers, given τ, the label of late producers is correctly communicated to the ledger, and validators have no incentive to tamper with the record of labels.

Given a stationary incentive feasible allocation (x, y, z^1, z^2), U_i is the expected discounted lifetime utility of late producer i satisfying

$$(1-\beta)U_i = \mathbb{E}_i \left\{ \mathbb{I}_{w_i \geq \tau} \alpha \left[u(x) - y - w_i V z^2 \right] \right\},$$

where $\mathbb{E}_i$ is the expectation operator of late producer i over w_i, the share of working validators. The late producer only trades if more than τV validators work and validate the trade. In this case the late producer gains $u(x) - y$ from trading but pays $w_i V z^2$ to the working validators. The expected discounted lifetime utility of an internal validator i with private

[24] A late producer who misbehaved at some time in the past and enters a period with a label B can bribe validators so as to obtain label G to get to consume. However, validators will not agree to a bribe in the early stage because (1) there is a possibility that the briber is a faulty late producer, but also (2) they would have to trust the late producer to pay the bribe in the late stage. However, if validators accept the bribe, the late producer has no incentive to make good on it.

[25] We assume it is costless for late producers to bribe validators because if the latter reject it, late producers can always revert to paying early producers as planned even when validators are caught cheating.

communication cost $c_{s,i}$ is $U_{IV}(c_{s,i})$ and satisfies

$$\begin{aligned} U_{IV}(c_{s,i}) &= \mathbb{E}_i\left\{\mathbb{I}_{w_i\geq\tau}\alpha\left[u(x)-y-w_iVz^2\right] \mid c_{s,i}\right\} && (1)\\ &\quad +\alpha\max\left\{0;\mathbb{E}_i\left[-c_v+(1-f)\left(\mathbb{I}_{w\geq\tau}\left(z^1+z^2\right)-c_{s,i}\right) \mid c_{s,i}\right]\right\}+\beta\mathbb{E}U_{IV} && (2)\end{aligned}$$

where $\mathbb{E}(\cdot \mid c_{s,i})$ is the expectation operator over the common communication cost c_s of validator i conditional on receiving signal $c_{s,i}$.[26] Since internal validators are selected from the set of late producers, they obtain the expected payoff of late producers. In addition they also get the expected payoff from validating a trade: Given their signal $c_{s,i}$, validators can choose to work or not. If they do not work, they get nothing. If they work, validators incur the verification cost c_v. If they verify and the producer has a good label (which happens with probability $1-f$), they incur the communication cost $c_{s,i}$, and get the reward z^1+z^2, but only when the trade is validated ($\mathbb{I}_{w\geq\tau}=1$). Otherwise they do not get a reward.

Similarly, the expected discounted lifetime utility of an external validator i with private communication cost $c_{s,i}$ is $U_{EV}(c_{s,i})$ and satisfies,

$$U_{EV}(c_{s,i}) = \alpha\max\left\{0;\mathbb{E}_i\left[-c_v+(1-f)\left(\mathbb{I}_{w\geq\tau}\left(z^1+z^2\right)-c_{s,i}\right) \mid c_{s,i}\right]\right\}+\beta\mathbb{E}U_{EV} \quad (3)$$

Participation constraints. An allocation (x,y,z^1,z^2) satisfies the participation constraints of validators, early and late producers whenever

$$\begin{aligned} \mathbb{E}\left[u(x)-y-wVz^2\right] &\geq 0 && (4)\\ \mathbb{E}\left[y-x-wVz^1\right] &\geq 0 && (5)\\ \mathbb{E}\left[-c_v+(1-f)\left(\mathbb{I}_{w\geq\tau}\left(z^1+z^2\right)-c_s\right)\right] &\geq 0 && (6)\end{aligned}$$

The expectation operator in (4) and (5) is again on the share of working validators w. Since late producers can refuse to become validators, the last constraint requires that validators expect to make a positive expected profit from the validation process and their expectation

[26] Validators have more information concerning the fundamental communication cost c_s, which they use to compute the probability that the trade be validated.

operator is over w and c_s.

Repayment constraints. Using a ledger, in equilibrium late producers who do not produce the announced amount y are detected by validators and thus assigned a label B, and are permanently excluded from all economic activities. Therefore, given the share of working validators is $w \geq \tau$, the repayment constraint of late producers and internal validators is respectively

$$-\left(y + wVz^2\right) + \beta U \geq 0. \tag{7}$$

$$-\left(y + wVz^2\right) + \beta \mathbb{E}U_{IV} \geq 0. \tag{8}$$

No bribe. If a validator accepts a bribe, we assume it is caught with probability $\pi \in [0, 1]$. In this case it loses its right to validate future transactions and to consume as a late producer. A validator prefers recording the truth to a false record when the late producer offers $\bar{z}$ iff

$$\underbrace{z^1}_{\text{producer does not pay } z^2} + \beta \mathbb{E}U_\upsilon \geq \underbrace{z^1 + \bar{z}}_{\text{producer bribes } \bar{z}} + (1-\pi)\beta \mathbb{E}U_\upsilon$$

where $\upsilon = IV, EV$ to denote internal of external validation, respectively. When a share w of validators are working on a match, the late producer in this match is willing to pay at most a total of $y + wVz^2$ to get away with production. Given that the ledger requires the agreement of at least τV validators to validate a transaction, the cheating late producer will pay $\bar{z} = (y + wVz^2)/(\tau V)$ to τV validators. Using $\bar{z} = (y + wVz^2)/(\tau V)$, a validator rejects the bribe whenever[27]

$$\pi \beta \mathbb{E}U_\upsilon \geq \frac{1}{\tau V}\left(y + wVz^2\right). \tag{9}$$

It remains to find a condition such that, given a threshold τ, the allocation allows for a truthful record of the ledger.

Validation threshold. Suppose there is a positive measure of validators in charge of verifying each match. Then the decision of an arbitrary validator to work or to shirk depends

[27] We assume validators act in unison: they either all accept the bribe, or they all reject it.

on the subjective probability this validator assigns to other validators working or shirking. Therefore, there are many possible equilibria depending on the original beliefs of validators. In other words, the uncertainty about the cost of other validators of communicating a label to the ledger may reverberate throughout the system and may jeopardize the validation process.

We assume the continuation payoff of validators is independent of the current validation results.[28] Also if validators do not work, they do not send any messages. We relax these assumptions in Section F, where we also specify the voting game that validators play in detail, and we state conditions under which validators do not send a message without working.

Now suppose the validation process requires unanimity. As soon as validators expect a positive measure to abstain from validating, they will also abstain even though they may have received a signal that the communication cost is small. As a consequence, we show that validation only occurs correctly when the validation rule is based on supermajority unless payments to validators are arbitrarily large. The higher the supermajority threshold, the more rents should accrue to validators in order to guarantee the integrity of the ledger. We show the following result in the Appendix,

Proposition 1. *Given τ, in the limit as $\varepsilon \to 0$, there is a unique dominance-solvable equilibrium where validators work if and only if the allocation (z^1, z^2) satisfies*

$$z^1 + z^2 \geq \frac{c_s + c_v/(1-f)}{1-\tau}. \tag{10}$$

The proof follows two steps. In the first step, we characterize equilibria when, for some common threshold c_s^*, validators employ switching strategies whereby they work if their cost $c_{s,i}$ below the threshold c_s^* and they shirk otherwise. A validator receiving a cost at threshold level c_s^* is indifferent between working and shirking.

It is well-known from the global game literature (e.g. Morris and Shin, 2003) that, for the marginal player whose cost is exactly equal to the threshold value c_s^*, the density over the

[28] For example, it is difficult to distinguish whether a validator shirked or his message technically failed to reach other validators. However, see Green and Porter (1991) and Monnet and Quintin (forthcoming) show how punishments followed by forgiveness may discipline agents who can hide behind the veil of "bad luck."

share of working agents is uniform over $[0, 1]$. Hence, the validator assigns a probability q to the event that a fraction q of the V validators will work. Since this validator is indifferent between working or not, and his subjective beliefs of the share of working validators is uniform, i.e. $g(\tau \mid c_s^*) = 1$, c_s^* solves

$$-c_v + (1-f)\left[(1-\tau)\left(z_1 + z_2 - c_s^*\right) + \tau\left(-c_s^*\right)\right] = 0.$$

When the noise vanishes, all individual costs necessarily converge to the common value c_s. Therefore, when $c_s \leq c_s^*$, all validators will work to validate a trade and the ledger will record labels correctly, while when $c_s > c_s^*$ none of them will. The allocation $z_1 + z_2$ logically affects the threshold value c_s^*: By increasing the validator's rents $z_1 + z_2$, the validation protocol can ensure that validation happens for higher levels of the communication cost.

Hence, the first step of the proof establishes that the validation game has a unique equilibrium in switching strategies. The second step of the proof establishes that this unique equilibrium in switching strategies is also the only strategy profile of the players that survives the iterated deletion of strictly dominated strategies. In other words, the game is dominance solvable.[29]

As a corollary, notice that the payment to validators, as measured by $z^1 + z^2$ is positively linked to the supermajority level τ when there are validation costs. Therefore, the ledger can only retain integrity when unanimity is required if payments to validators are arbitrarily large. Finally, given τ, the probability that the trade will go through when c_s is uniformly distributed is the probability that

$$c_s \leq (1-\tau)(z_1 + z_2) - \frac{c_v}{1-f} \equiv c_s^*.$$

If the ledger is required to allow *all* legitimate trades involving a producer with label G will

[29] Morris and Shin (2003) shows that a sufficient condition for dominance solvability in our setting is that the payoffs satisfy strategic complementarity – that is, the payoff to working is weakly increasing in the proportion of other validators who work. Since this condition is satisfied in our game, we can apply the global game results in Morris and Shin (2003) to conclude that our game is dominance solvable.

always go through, then $z_1 + z_2$ should be set to

$$z_1 + z_2 = \frac{1}{1-\tau}\left(\bar{c}_s + \frac{c_v}{1-f}\right) \equiv \frac{C}{1-\tau} \tag{11}$$

where $\bar{c}_s$ is the maximum possible communication cost. In this case, and given τ, validators will always work. Below, we assume this is the case. In the Appendix, we relax this assumption and we analyze the optimal validation protocol for any thresholds c_s^* defined by (10) holding with equality and we let the designer choose what c_s^* should be. Also, we derive sufficient conditions under which $c_s^* = \bar{c}_s$ is optimal.

We can now define incentive feasible allocations.

Definition 1. Given τ and V, an incentive feasible allocation is a list (x, y, z^1, z^2) that satisfies (4)-(9) and (11).

In the sequel we solve for the optimal design of the validation protocol, and how it affects incentive feasible allocations. We simplify matters further by assuming that the distribution for the communication cost c_s converges to the degenerate distribution that gives all the mass to just one point $\bar{c}_s$.

3.2 Degenerate distribution of communication costs

From now we consider the limiting case where $\varepsilon \to 0$, (11) holds, and the distribution of the communication costs converges to a degenerate distribution at $\bar{c}_s$. So unless they are bribed, all validators always verify the label, always verify that production took place according to plans, and always cast the right vote to the ledger. Therefore the share of working validators w converges to one and we can write (4)-(10) with $w_i \to 1$. In this section, we consider the incentives of a late producer to make side payments to validators so that they record a false trade, and we analyze the incentives of validators to accept that bribe.

We can further simplify the set of IF allocations by setting the participation constraint of early producers at equality.

Since the payment to validators should be minimized, (11) binds so validators earn $z^1 + z^2 = Z(\tau) = C/(1-\tau)$ and the participation constraint of validators (6) is always satisfied. Using $R(\tau)$ to denote the expected rent of validators,

$$R(\tau) \quad = \quad (1-f)\,\frac{\tau C}{1-\tau} \tag{12}$$

the set of IF allocations is characterized by[30]

$$\frac{\beta\alpha}{1-\beta}\left[u(x) - (x + VZ(\tau))\right] \quad \geq \quad x + VZ(\tau) \tag{13}$$

$$\pi\frac{\beta\alpha}{1-\beta}\left[u(x) - (x + VZ(\tau)) + R(\tau)\right] \quad \geq \quad \frac{1}{\tau V}\left(x + VZ(\tau)\right) \tag{14}$$

Constraint (13) says that late producers are better off retaining their good label by repaying x to early producers and $VZ(\tau)$ to validators, than getting a bad label and lose the expected lifetime discounted payoff from trading net of the payment to validators. Constraint (14) compares the payoff a validator would obtain from accepting the maximum bribe a late producer would offer $(x + VZ(\tau))/\tau V$, to the expected loss of accepting such a bribe, given they are caught with probability π. In addition to losing the expected lifetime discounted payoff from trading net of compensating validators, validators would also lose the validation rents they earn $R(\tau)$. Notice that since early producers have linear utility, their production $x + z^1$ is compensated by late producers through their production $y = x + z^1$. In this context, it does not matter who bears the cost to compensate validators $Z(\tau)$ and only the total cost $Z(\tau)$ matters for the allocation.[31] With external validation (14) simplifies to

$$\pi\frac{\beta\alpha}{1-\beta}R(\tau) \geq \frac{1}{\tau V}\left(x + VZ(\tau)\right) \tag{15}$$

Equation (15) says that losing future validation rents with probability π should be a larger cost to external validators than the benefit of accepting the largest bribe a late producer

[30] We can set the participation constraint for early producers (5) at equality and replace $y = x + Vz^1$. Also, since validators earn a rent, $U_{IV} \geq U$ and (8) is satisfied whenever (7) is.

[31] This is obviously not robust to a generalization of the preferences of early producers. In the Appendix, we solve the case when early producers have preferences for consumption represented by a strictly concave utility function.

would offer.[32]

It will be convenient to define the default factor as

$$\delta \equiv \frac{\pi\beta\alpha}{1-\beta}.$$

The higher δ is, the lesser are the incentives of validators to accept a bribe, either because they would lose a lot of trading opportunities (as captured by a large α) or because they would very likely be caught cheating (as captured by a high π) or because they care a lot about future income (as captured by a high β). We next focus on the number of validators V. The more validators there are, the higher the total production must be to make the payment to validators – this is the term $VZ(\tau)$ on the left-hand side of (13). However, given τ more validators also means that the bribe per validator diminishes, which relaxes the constraint of validators as shown on the RHS of the validator's incentive constraint (14) or (15). The optimal number of validators will trade off both effects. When the measure of validators is large, the repayment constraint of late producers becomes the binding one when $\pi \to 1$, so that validators never accept a bribe. This is intuitive: when V is large, no validator gets a very large bribe, but if the mechanism almost surely observes when they accept a bribe, they almost certainly loose the expected lifetime payoff from trading. Then, only the incentives of the late producer matter. However, note that if (13) is binding while (14) is slack, V can be reduced up to the point where (14) binds. We summarize the discussion above in the following result.

Lemma 1. *The distribution of the cost to remunerate validators $Z(\tau)$ among early or late producers is indeterminate. If π and V are large enough, validators will never accept a bribe.*

[32]Our mechanism that requires a supermajority τ of *all* validators implies that producers can offer a smaller bribe relative to a mechanism that would base consensus on a supermajority of *voting* validators (instead of all validators). Also, using a mechanism relying only on voting validators may be problematic because validators cannot be induced to vote when producers are faulty, as they do not expect any rewards.

3.3 Optimal design

Since the probability of a productive match is $1-f$ in each period, the objective function of a planner equals the gains from trade net of the validation costs. A planner chooses the trading size x, the type of validation (internal or external), the number of validators and the threshold τ to solve

$$\alpha(1-f)\max_{x,V\geq 0,\tau\in[0,1]}\{u(x)-x-VC\} \tag{16}$$

subject to (13) and (14) for internal validation, or (15) for external validation.

Conveniently, the size of the payment to validators $Z(\tau)$ only matters for incentives but has no impact on the objective function: it is a mere transfer between producers and validators. Also note that whether validation is internal or external does not affect the social planer's objective function (16). However, comparing (14) and (15) shows that as long as trade is sufficiently beneficial, or $u(x) > (x + VZ)$ then (14) is always satisfied whenever (15) is. In other words, the set of incentive feasible allocations is larger with internal validation, and internal validation weakly dominates external one. But not only this. As we show below, for the case of external validation, whenever trade can be supported, centralization is optimal. Next, we start by considering the optimal design with external validation.

Optimal design with external validation. With external validation the only relevant constraints are (13) and (15), which we can write respectively, as

$$u(x)-x-\frac{VC}{1-\tau} \geq 0 \tag{17}$$

$$\left[\delta(1-f)\tau^2-1\right]VC \geq (1-\tau)x \tag{18}$$

It is easy to show that the validators' incentive constraint (18) always binds,[33] so we can use it to replace for the total cost of validation, VC in the late producer's repayment constraint

[33]Suppose it does not at the solution $(\tilde{x},\tilde{\tau},\tilde{V})$. Then reduce V until it does. This increases the objective function (16), while relaxing the PC of late producers. So $(\tilde{x},\tilde{\tau},\tilde{V})$ could not be the solution, a contradiction.

(17) and the objective function. The problem of the planner then becomes

$$\alpha(1-f)\max_{x,\tau\in[\bar{\tau},1]}\left\{u(x)-x\left[\frac{\delta\left(1-f\right)\tau^{2}-\tau}{\delta\left(1-f\right)\tau^{2}-1}\right]\right\}$$

subject to

$$u(x)-x\frac{\delta\left(1-f\right)\tau^{2}}{\delta\left(1-f\right)\tau^{2}-1}\quad\geq\quad 0.$$

It is straightforward to verify that both the objective function and the left-hand side of the constraint are increasing in τ. Therefore, with external validation, the solution is $\tau = 1$ and $V \to 0$, and the optimal trading size is $x = x^*$ since, with $\tau = 1$, it solves the first order condition. Inspecting (18), notice that a necessary and sufficient condition for the existence of external validation is that intertemporal incentives are strong enough, in the sense that $\delta(1-f) > 1$. Otherwise, the only incentive feasible allocation is autarky.

Optimal design with internal validation. Next, consider the optimal design with internal validation. As we show in the Appendix, the social planner's objective function (16) is decreasing in V, and so we obtain

Lemma 2. *With internal validation, the incentive constraint of validators (14) always binds while the incentive constraint of producers (13) never binds.*

Replacing the expression for the validation rent in (14), and re-arranging, we obtain:

$$\delta\left[u(x)-x\right]\geq\frac{1}{\tau}\left[\frac{x}{V}+Z(\tau)\right]-\delta\left[\left(1-f-V\right)Z(\tau)-(1-f)C\right] \tag{19}$$

Since the objective function (16) is independent of τ, the planner chooses τ to minimize the right hand side of (19). Two forces are at play. On the one hand, increasing τ reduces the maximum bribe size per validator. On the other hand, increasing τ increases the payment to validators $Z(\tau)$ to ensure that validators indeed verify and validate. When intertemporal incentives are strong, so that $1 \leq \delta\left(1-f-V\right)$, this second effect reduces the right-hand side of (19): Validators have much to lose by accepting a bribe. In this case, as with external

validation, it is optimal to set $\tau = 1$, even if $Z(\tau) \to \infty$. Alternatively, suppose intertemporal incentives are weak, $1 > \delta(1 - f - V)$, then the optimal $\hat{\tau}$ trades-off the lower bribe size with the higher payment to validator and it solves[34]

$$\frac{1-\hat{\tau}}{\hat{\tau}} = \sqrt{\frac{[1-\delta(1-f-V)]VC}{x+VC}} \tag{20}$$

The optimal threshold $\hat{\tau}$ is decreasing in V but increasing in x. $\hat{\tau}$ is also increasing in π, β or α, as captured by δ. The intuition is apparent from (13): When π, β or α increase, the net rent $R(\tau) - VZ(\tau) > 0$ of validators becomes more important for the incentives of validators relative to the bribe size $(x + VZ)/\tau V$, either because they have a higher chance of losing it – when π increases – or because they have a higher lifetime discounted value – when β or α increases. So following a rise in δ, the planner increases the net rent of validators by increasing τ. So unlike in traditional models of limited commitment, higher trustworthiness as captured by a higher δ implies more rents to validators.

Importantly, while external validation would only implement autarky when intertemporal incentives are weak, internal validation can do much more. This is intuitive: With internal validation, the planner can use the value of trading in the future to discipline validators, thus allowing the planner to choose lower rents (and therefore a lower feasible bribe) validators obtain from being able to manage the ledger.

The optimal choice of x and V trades off several effects. First, increasing x toward x^* brings additional gains from trade, but at the cost of increasing the bribe size per validator which tightens their incentive constraint. Second, increasing V relaxes the incentive constraint of validators, but increases the cost of validation. We can now summarize these considerations in our main result.

Proposition 2. *The constrained optimal solution $(\hat{x}, \hat{V}, \hat{\tau})$ solves (19) at equality and (20)–(24) and is characterized by four regions:*

1. [centralized system] *If $\delta > 1/(1-f)$, internal and external validation can implement the optimal allocation, which is arbitrarily close to the one with a single validator, $\hat{V} \to 0$,*

[34] It is easy to verify that when $1 > \delta(1 - f - V)$, the second order condition for a minimum is satisfied.

$\hat{\tau} \to 1$ *and* $\hat{x} \to x^*$*, but it requires arbitrarily large payments, i.e.* $Z(\hat{\tau}) \to \infty$*, while* $\lim_{\hat{\tau}\to 1} V(\hat{\tau})Z(\hat{\tau}) = \frac{x^*}{\delta(1-f)-1}$.

2. [partially distributed system] *If* $\bar{\delta} < \delta \leq 1/(1-f)$*, only internal validation can decentralize trade. The constrained optimal number of validators is* $\hat{V} > 0$*, and only a supermajority* $\hat{\tau} < 1$ *is optimal. Each validator receives a finite payment* $Z(\hat{\tau}) < \infty$*. The constrained optimal allocation is* $\hat{x} < x^*$.

3. [fully distributed system] *If* $\delta_0 < \delta \leq \bar{\delta}$*, only internal validation can decentralize trade. All late producers are validators* $\hat{V} = 1 - f$*, and* $\hat{\tau} = \left(1 + \sqrt{\frac{(1-f)C}{x+(1-f)C}}\right)^{-1}$*. The constrained optimal allocation is* $\hat{x} < x^*$.

4. [no trade] *If* $\delta \leq \delta_0$*, there is no validation protocol that can decentralize trade.*

Proposition 2 states that the optimal design of the ledger requires centralized validation by a single validator only when validators are sufficiently trustworthy and there are relatively few faulty producers. However, moving toward centrality necessarily imposes a move toward unanimity. Reaching consensus is costly because it requires an arbitrarily large payment to the single authority managing the ledger. In reality, feasible payments may be bounded and in such a case, a single validator will never be optimal. Still, the single validator case offers a useful benchmark that illustrates the forces driving the solution toward centrality. When intertemporal incentives become weak, the single authority can be more easily convinced to do wrong and the optimal validation protocol moves away from centrality. The centrifugal forces manifest themselves also in a reduction in trade size, and a departure from unanimity or a high supermajority rule. Both margins reduces the feasible bribe to validators and therefore their incentives to do wrong.

Underlying the results in Proposition 2 are the following comparative statics (details of the calculations are in the Appendix),

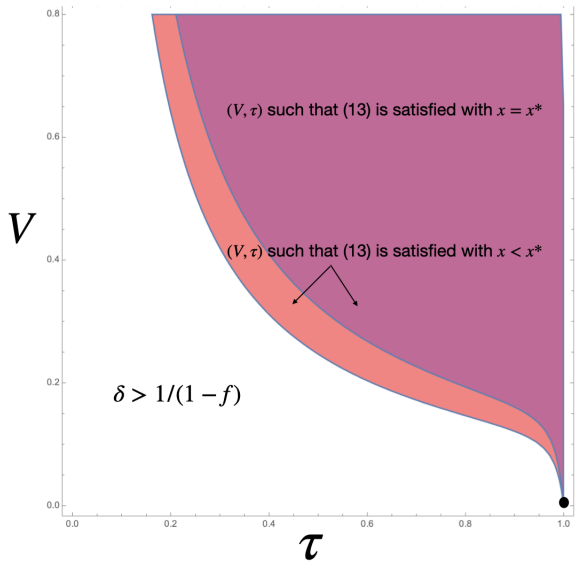
- x and τ are (weakly) increasing with δ (i.e. β, π, and α) but decreasing with validation costs C.
- V is (weakly) decreasing with δ (i.e. β, π, and α) but increasing with validation costs C.

It is rather intuitive that the optimal number of validators (V) decreases with δ, while the optimal trade size x increases with δ: Everything else constant, when validators are more patient, or when they trade more often, they have more to lose from wrongdoings. As a result, they are less likely to do so and their incentive constraint is loosened as δ increases. The planner can then increase the size of each trade x and reduce the number of validators. The choice of V and x indirectly impacts the optimal supermajority threshold, and we have explained above why the planner increases the net rent of validators by choosing a larger τ, as δ increases. Both direct and indirect effects are reinforcing each other, and the total effect of increasing δ is to increase τ. Finally, notice that when the system becomes fully distributed with $V = 1 - f$, the supermajority threshold converges to $1/2$ from above as $x \to 0$ although we have not imposed the constraint $\tau \geq 1/2$.

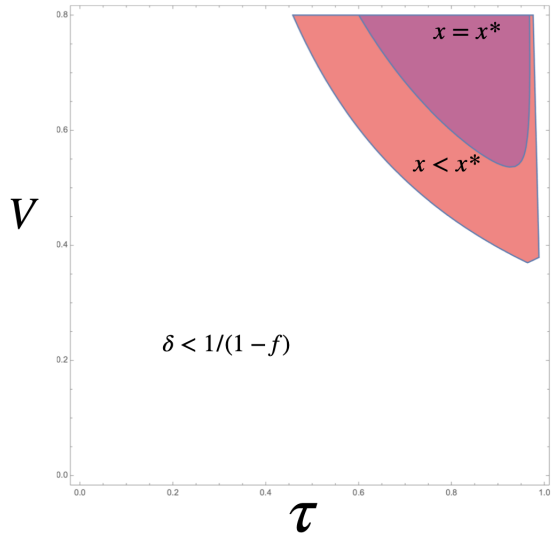
Centrifugal forces also include the costs of validation C. This may be surprising as increasing the number of validators also increases duplication costs to the detriment of social welfare. However, validation costs are reducing the overall lifetime discounted surplus of validators, who, as a result are more easily convinced to do wrong. Then it is optimal to increase the number of validators so that (given x) each of them can only be offered a smaller bribe.

Reaching consensus with a higher number of validators however does not come for free: As C is higher, validators are more likely to believe that fewer other validators will work and maintaining the same level of consensus requires a larger rent be paid to each validator as Proposition 1 shows. To maintain the legitimacy of the ledger while keeping costs in check, the threshold τ should fall.

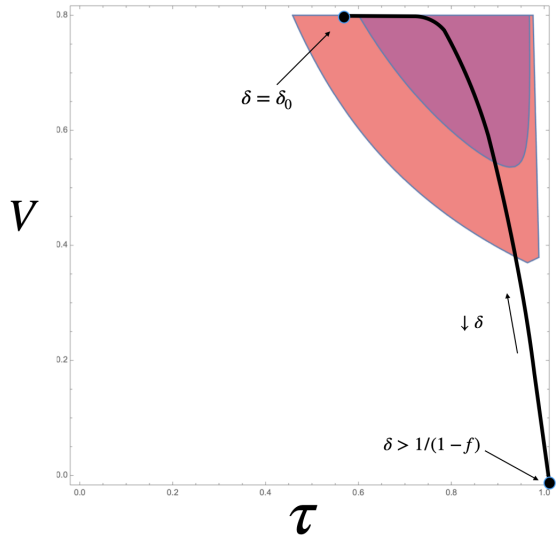
Figure 3 illustrates Proposition 2. Given some $x \leq x^*$ and the default factor δ, Figure 3 shows the set of (V, τ) for which the IC of validators (19) is satisfied. When δ is relatively high, $x = x^*$ and $(V, \tau) = (0, 1)$ satisfy (19) and this is the best design for the system. In this case, the solution is given by the black dot in Figure 3a. In contrast, Figure 3b shows incentive feasible allocations when δ is relatively smaller: Then $x = x^*$ and $(V, \tau) = (0, 1)$ is no longer incentive feasible. The figure shows that implementing $x = x^*$ is feasible for some (V, τ), but only for relatively large V. This is costly and the planner does better by choosing a lower x which decreases the bribe size and allows it to select fewer validators (lower the



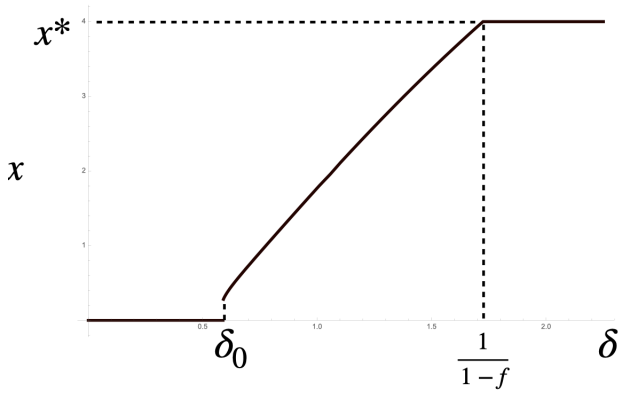
(a) $\delta(1-f) > 1$, IF choice of (V,τ) for $x = x^*$, $x < x^*$.



(b) $\bar{\delta} < \delta < 1/(1-f)$ IF choice of (V,τ) for $x < x^*$, $x = x^*$



(c) Optimal solution for (V,τ) as δ falls



(d) Optimal solution for x

Figure 3: Incentive feasible choice of (V,τ)

number of validators V), as shown by the red area in Figure 3b. Reducing x slightly below x^*, the planner only makes a second order loss, but realizes a first order gain as V decreases, thus reducing the overall validation costs. Therefore, as δ decreases, x declines and (V, τ) moves up along the black curve, as shown in Figure 3c. As δ falls below the threshold $\bar{\delta}$ toward δ_0, $V = 1 - f$ and τ moves westward on the black line until it reaches $1/2$ and $x = 0$. For levels of δ below δ_0, there is no equilibrium with trade. Finally, as we move to the left along the black curve, the trade size x decreases toward zero.

4 Conclusion

In this paper, we have presented an economic analysis of decentralized ledger technology in an economy where money is essential. To our knowledge, our analysis is the first economic analysis of DLT in such a context. It links a ledger validation game to monetary exchange, establishes the uniqueness of the equilibrium via a global game approach, and characterizes the optimal mechanism design, examining the optimal supermajority voting rule, number of validators, and size of transactions, thus addressing Buterin's scalability trilemma.

We believe our analysis is a timely one, as DLT is rapidly becoming an industry standard for digital currencies and in other applications. In particular, our results can shed light on the burgeoning literature on central bank digital currency and stablecoins insofar as it gives conditions under which a central authority should manage the ledger of transactions.[35] The economic discussion of technology and the economics of central bank digital money has thus far centered on the balance sheet effects and related systemic implications.[36] Here, we focus not on balance sheets and the issue of how the value of a currency can be guaranteed (central backing is of the essence for a CBDC irrespective of our analysis), but on the governance of money when it is used in exchange as the societal record-keeping device.

Of course, we have made simplifying assumptions in order to better grasp the basic economics of memory. Future work should relax some of these. For instance, we have assumed that

[35] See the stock-taking exercise of pursued technological designs in Auer et al. (2020).

[36] See among others Andalfatto (2018), Brunnermeier and Niepelt (2019), Fernández-Villaverde et al. (2020), Keister and Monnet (2020).

one individual can only have one account, so that the reputation of the individual and their account are intertwined. However, ledgers only record transactions for one account and it is usually difficult to trace the identity of the owner of the account. However, our analysis would extend directly to reputation on accounts rather than on individuals.[37]

Further - to simplify the analysis - we have assumed that validators all agree to accept bribes in unison. It would be interesting to also study the cooperative games between validators in more detail. We have also taken as given that agents use a private permissioned ledger as they want to preserve their anonymity in trades. Tirole (2020) and Chiu and Koeppl (2020) make progress on this front. In order to better compare the different types of ledger, future work should also include the benefit from preserving anonymity. Also our mechanism design approach implies that we have ignored every industrial organization aspect of DLT, which might be significant if this technology were to be widely adopted in the future.

[37]For an analysis of why it is important to distinguish individuals from accounts, see Li and Wang (2019).

Appendix

A The technology and economics of permissioned distributed ledgers

How does the model developed above relate to the actual underlying technologies? In all DLT applications, the aim is to achieve agreement on the state of some data in a network of nodes that constantly exchange the underlying data file, the ledger. For example, in the case of the permissionless cryptocurrency Bitcoin, the state is the current ownership of bitcoins and the network includes anyone who downloads the blockchain.[38] In a permissioned application in the field of trade finance, the state might be the delivery status of a set of shipments and respective payments, and the network includes anyone authorized to access the system. The aim is to achieve agreement on the state of some data in a network of nodes that constantly exchange the underlying data file, the ledger. For example, in a trade finance application, the state might be the delivery status of a set of shipments and respective payments, and the network includes anyone authorized to access the system.

The set of rules by which agreement is achieved is called the consensus algorithm (or mechanism). This is a computer protocol, specifying the conditions under which a ledger is considered as valid and, importantly, it also guides how to choose between multiple versions of a ledger should conflicts ever emerge. The consensus mechanism sustains decentralized exchange whenever it creates incentives for everyone to follow the rules of its protocol.

Permissioned DLT applications generally feature two kinds of actors, "users" and "validators". On the one side, users want to transact, whereas validators are those who can include (either alone or as a group, depending on the precise consensus algorithm) new transactions in ledger updates. We note that there is nothing that bars a validator from also being a user. In fact, as we showed in the theoretical analysis above, it is optimal to draw validators from the set of users because, as users of the DLT, validators are interested in preserving its integrity.

The validation process involves both the verification of transactions and the casting of a vote to the ledger. Verification means to read the ledger and analyze

[38] A blockchain is a ledger is composed of files (called blocks, each containing a number of transactions) that are sequentially chained (thus resulting in a chain of blocks).

whether it is consistent with the current state (ie the version of the ledger that resulted from a previous instance of running the consensus mechanism). Voting is the process by which one or many validator(s) agree to write an authorized transaction into the ledger to update it. Below we will model these two steps of the validation processes in detail.

To exemplify these actors and actions, consider a simple transfer of a token from user L to E in a DLT application such as in Corda's R3, Hyperledger Fabric, or Quorum, or alternatively in the Diem Blockchain of the Diem Association's global stablecoin proposal.[39] User L initiates the transaction via its digital signature. The set of validators V (called "notaries" in Corda) can verify that L is indeed the current owner of the token and that the transfer authorization is authentic. The verification process includes verifying whether L is authorized to transfer the token and verifying whether L has already spent the token. For the former verification step, each validator needs to check the digital signature of the token owner L, as well as, the entire chain of transactions involving this token. If all signatures from the point of issuance to that of the transfer to L are valid, the token is authentic. For the latter verification step, each validator needs to check the ledger to see if L still owns the token, i.e. for the absence of an older transaction transferring the token to somebody else.

Each validator can then broadcast its vote to include the transaction in the next update of the ledger. In this process, the key issue concerns the conditions under which the votes can be considered as being sufficient for E to assume that the transaction has indeed occurred, i.e. that the transfer is and will continue to be included in the consensus version of the ledger. This is achieved via a voting-based consensus protocol, for example Practical Byzantine Fault Tolerance.[40] It typically involves the following steps: after verification, validators

[39]To be precise we focus on the details of the validation steps in Corda. In Hyperledger Fabric and Quorum, the relevant logical steps of the transaction are the same, although the technical implementation differs somewhat. Technically, whereas Corda like Bitcoin follows a UTXO model where verification of a transaction involves tracing a token all the way back to its origin, in Hyperledger Fabric and Quorum – as in the cryptocurrency Ethereum – a transaction resembles the account-based system where transaction include an update on the balance of accounts. Tokens are not native units of Hyperledger Fabric or Quorum, but can be constructed or emulated on them.

[40]Indeed, many of today's sandbox-style applications of Corda R3 feature only one notary, and once this notary has signed a transaction into the ledger, it becomes final. As we show below, this is the exact equivalent of today's model of centralised exchange. Looking ahead, Corda (2020) argues that while Byzantine Fault Tolerance-type of consensus mechanism is most commonly used, the platform is pluggable with different consensus mechanisms.

communicate their votes to each other. Once a share of votes exceeding the pre-set threshold – for example, 50% plus 1 – of votes have been communicated, the transaction is understood to be valid under the consensus rule. After this has happened E can, with some but not full certainty, assume that the transaction is final.

The key underlying economic problems – which we analyze in the main part of the paper – concern not the technical implementation, but the underlying incentives of the validators that are needed to sustain honest exchange. The first aspect is that there is no technical way to force a validator to sign any given transaction. Validation may be costly – especially if this involves costly monitoring of off-chain events. Validators hence need to be incentivized to actively verify and vote on transactions. The second aspect is that nothing can technically prevent a validator from voting on multiple ledgers with conflicting histories. The latter opens up the possibility of a "history-reversion attack", in which a first consensus ledger emerges that includes a given transaction, but later a second consensus ledger emerges that does not include it.

A history-reversion attack is one in which the payment for a merchandise first enters the consensus ledger to be sub-sequentially excluded in a future consensus ledger. In the above-stated transfer from L to E, such a fraudulent attack would take place in the following way. L and E agree on an exchange of merchandise from E to L in exchange for the payment of a token from L to E. L initiates the token transfer and after E observes that the transfer has become part of the consensus ledger, it releases the merchandise. Once this has happened, L bribes a sufficient number of validators to vote on a conflicting ledger that does not contain said transfer, thus effectively undoing the transaction.[41]

Such history-reversion attacks are becoming increasingly frequent for permissionless cryptocurrency (see i.e. Shanaev et al. 2020), and, going forward, a mechanism must be found to ensure that it does not happen in permissioned DLT. For this mechanism, importantly, although validator action is not enforceable, it is observable to the participants of the system,

[41] As multiple versions of the ledger can exist, a consensus protocol also must specify a rule to distinguish among them. This tends to be the one version with the most votes. If the consensus rule was 50% plus 1, 50% plus 2 votes are needed for a successful attack. Axiomatically, a history reversion attack thus can only succeed if at least some validators – the exact number depending on the supermajority rule – validate conflicting histories.

and the protocol can thus reward and punish certain validator actions. On the one side, it can reward active participation (also to reimburse for the cost incurred during verification). On the other side, it can punish malfeasance via exclusion from the set of future validators. If validators earn fee income in excess of operation costs, supporting a history-reversion attack thus has the cost of forgoing the net present value of validator profits.

In line with these considerations, we model the two-step verification process and the underlying incentives by assuming 1) some users are faulty, and 2) users are not trustworthy. To model step 1) we abstract from the idea of users owning a token. As in Kocherlakota (1998), we rather consider users having a label related to their history of trade. In our model, a simple label turns out to be a set of statistics sufficient to summarize the history of trades. A good label stands for the fact that L is authorized to trade, very much as a token was acquired in a legitimate way. So in our model, validators verify the entire history of trade as summarized by the label, and communicate the result by sending a signal to the ledger. Our model takes step 2) into consideration by letting L choose to "double spend" by obtaining goods in exchange for the promise to produce later (akin to the promise to deliver the coin) but then not delivering on that promise in due time.

B Proof of Proposition 1

Proposition. *[Proposition 1] Given τ, in the limit as $\varepsilon \to 0$, there is a unique dominance-solvable equilibrium where validators work if and only if the allocation (z^1, z^2) satisfies*

$$1 - \tau \geq \frac{c_s + c_v/(1-f)}{z^1 + z^2} \tag{21}$$

Let $z = z^1 + z^2$. Assume that each validator receives private cost

$$c_{s,i} = \gamma_s + \varepsilon_i$$

where ε_i is uniformly distributed over $[-\varepsilon, \varepsilon]$. Given a validation threshold τ, the expected

payoff of validator i is

$$-c_v + \begin{cases} (1-f)\left(-c_{s,i} + z + \beta U_V\right) + f\beta U_V & \text{if } \int_{i=1}^{V} m_i \mathbb{I}_{m_i \neq \emptyset} di > \tau V \\ (1-f)\left(-c_{s,i}\right) + \beta U_V & \text{otherwise} \end{cases}$$

The expected payoff of a validator is given by (33) when the measure of validators sending a message is higher than τV. We can rewrite the expected payoff as

$$\beta U_V + (1-f)z \times \begin{cases} -\frac{\tilde{c}_v^1}{1-f} - \tilde{c}_{s,i} + 1 & \text{if } \int_{i=1}^{V} m_i \mathbb{I}_{m_i \neq \emptyset} di > \tau V \\ -\frac{\tilde{c}_v^1}{1-f} - \tilde{c}_{s,i} & \text{otherwise} \end{cases} \tag{22}$$

where we have normalized the cost by the validator's rent, as $\tilde{c}_{\cdot,i} = c_{\cdot,i}/z$. This normalized cost is necessarily lower than 1 and it is uniformly distributed since γ_s is uniformly distributed. Note that if a validator expects $\int_{i=1}^{V} m_i \mathbb{I}_{m_i \neq \emptyset} di < \tau V$, this validator will not even verify the label in the first place. The structure of the above payoff is the same as the one in the public good game analyzed in Morris and Shin (2002) and their results extend almost directly. We repeat their argument here for completeness.

Let w be the random variable $\int_{i=1}^{V} m_i \mathbb{I}_{m_i \neq \emptyset} di/V$ measuring the fraction of working validators. This is a random variable because the communication strategy m_i of each validator i depends on their communication cost. Also this random variable belongs to the interval $[0, 1]$. The distribution of w is important because it gives the probability that a trader with label G is able to trade and compensate validators, and in turn whether it is worth it in expected terms for a validator to verify and communicate the label to the ledger. Let $g(w \mid \tilde{c}_{s,i})$ be the subjective density over w for a validator with private cost $\tilde{c}_{s,i}$ and total verification and communication cost $-\frac{\tilde{c}_v}{1-f} - \tilde{c}_{s,i}$. We conjecture that validators adopt a switching strategy whereby they work whenever their total cost is lower than some level $C^* \equiv \frac{\tilde{c}_v}{1-f} + \tilde{c}_{s,i}^*$. Since the normalized cost is uniformly distributed over the interval $\left[\frac{\gamma_s - \varepsilon}{z}, \frac{\gamma_s + \varepsilon}{z}\right]$, the total cost is also uniformly distributed over $\left[\frac{\frac{c_v^1}{1-f} + \gamma_s - \varepsilon}{z}, \frac{\frac{c_v^1}{1-f} + \gamma_s + \varepsilon}{z}\right]$, and validators with $C_i < C^*$ are

working. Therefore, the measure of working validators is

$$w = \frac{\frac{\tilde{c}_v^1}{1-f} + \tilde{c}_s^* - \frac{\frac{c_v^1}{(1-f)} + \gamma_s - \varepsilon}{z}}{2\frac{\varepsilon}{z}} = \frac{c_s^* - (\gamma_s - \varepsilon)}{2\varepsilon}$$

So for some $q \in [0,1]$, there is a value for the **common** communication cost $\gamma_s(q)$ such that $w = q$. This is

$$\gamma_s(q) = c_s^* + \varepsilon - 2\varepsilon q$$

Hence, $w < q$ iff $\gamma_s > \gamma_s(q)$. We now need to find the probability that $\gamma_s > \gamma_s(q)$. Considering the validator with total cost C^*, the posterior density over γ_s conditional on his communication cost being c_s^* is uniform over the interval $[c_s^* - \varepsilon, c_s^* + \varepsilon]$. Hence, the probability that $\gamma_s > \gamma_s(q)$ is

$$\frac{c_s^* + \varepsilon - \gamma_s(q)}{2\varepsilon} = \frac{c_s^* + \varepsilon - (c_s^* + \varepsilon - 2\varepsilon q)}{2\varepsilon} = q.$$

Therefore

$$G(w < q \mid c_s^*) = q,$$

so that by differentiation, for all w

$$g(w \mid c_s^*) = 1$$

and the density over w is uniform at the switching point $\tilde{c}_s^*$. Hence, the probability that the validation process will fail is

$$G(\tau) = \int_0^\tau g(w \mid \tilde{c}_s^*) dw = \tau.$$

The validator with private cost c_s^* is indifferent between working and shirking. Therefore the

switching point $\tilde{c}_s^*$ solves

$$\begin{aligned}
-\frac{c_v}{(1-f)z} + \int_\tau^1 (1-\tilde{c}_s^*)\, g(\tau \mid \tilde{c}_s^*) d\tau + \int_0^\tau (-\tilde{c}_s^*)\, g(\tau \mid \tilde{c}_s^*) d\tau &= 0 \\
1 - G(\tau \mid \tilde{c}_s^*) - \tilde{c}_s^* &= \frac{c_v}{(1-f)z} \\
1 - \tilde{c}_s^* - \frac{c_v}{(1-f)\left(z - c_v^2 - c_s^2\right)} &= G(\tau \mid \tilde{c}_s^*) \\
1 - \tau - \frac{c_v}{(1-f)z} &= \tilde{c}_s^* = \frac{c_s^*}{z}
\end{aligned}$$

Hence, as $\varepsilon \to 0$, all validators with signal $c_{s,i} \leq c_s^*$ will work while all other validators will shirk. The probability that the validation process succeeds is $1-\tau$. Then, as $\varepsilon \to 0$, all validators will work whenever $c_s \leq c_s^*$ and they will all shirk whenever $c_s > c_s^*$. The argument to show uniqueness is standard from Morris and Shin (2003) given the payoff of any validators (22) to communicating with the ledger is increasing in the measure of validators who also communicate with the ledger.

Therefore, the probability that a trade validation process goes through is the probability that $c_s \leq c_s^*$, or

$$\int_{\underline{c}_s}^{c_s^*} \frac{dc_s}{\bar{c}_s - \underline{c}_s} = \frac{(1-\tau)z - \frac{c_v}{(1-f)} - \underline{c}_s}{\bar{c}_s - \underline{c}_s}.$$

C Proof of Lemma 2

Proof. We first show that (14) must bind. Suppose it does not. The objective function is decreasing in V and (13) is relaxed as V is lowered. So if (14) does not bind, it is optimal to set $V = 0$, and the solution is $\tilde{x} > 0$ which is the solution to $\max [u(x) - x]$ subject to $\frac{\beta\alpha}{1-\beta}[u(x) - x] \geq x$. However, since $\tilde{x} > 0$ and $\tau \leq 1$, it is clear that (14) cannot be satisfied when $V = 0$. Therefore (14) must bind. Suppose now both (13) and (14) bind. Then, given τ, the solution is given by

$$\begin{aligned}
\frac{\beta\alpha}{1-\beta}\left[u(x) - (x + VZ(\tau))\right] &= (x + VZ(\tau)) \\
\tau V \pi \frac{\beta\alpha}{1-\beta}\left[u(x) - (x + VZ(\tau)) + R(\tau)\right] &= (x + VZ(\tau))
\end{aligned}$$

Hence from the first equation,

$$x + VZ(\tau) = \frac{\beta\alpha}{1-\beta+\beta\alpha}u(x)$$

and from the second,

$$\begin{aligned}
\tau V\pi\frac{\beta\alpha}{1-\beta}\left[u(x) - \frac{\beta\alpha}{1-\beta+\beta\alpha}u(x) + R(\tau)\right] &= \frac{\beta\alpha}{1-\beta+\beta\alpha}u(x) \\
\tau V\pi\frac{\beta\alpha}{1-\beta}\left[\frac{1-\beta}{1-\beta+\beta\alpha}u(x) + R(\tau)\right] &= \frac{\beta\alpha}{1-\beta+\beta\alpha}u(x) \\
(1-\tau V\pi)\,\beta\alpha u(x) &= \tau V\pi\frac{\beta\alpha}{1-\beta}\,(1-\beta+\beta\alpha)\,R(\tau) \\
u(x) &= \frac{\tau V\pi}{1-\tau V\pi}\left(1+\frac{\beta\alpha}{1-\beta}\right)R(\tau)
\end{aligned}$$

Hence

$$\begin{aligned}
x + VZ(\tau) &= \frac{\beta\alpha}{1-\beta+\beta\alpha}u(x) \\
x + VZ(\tau) &= \frac{\beta\alpha}{1-\beta+\beta\alpha}\frac{\tau V\pi}{1-\tau V\pi}\left(\frac{1-\beta+\beta\alpha}{1-\beta}\right)R(\tau) \\
x + VZ(\tau) &= \frac{\tau V\pi}{1-\tau V\pi}\frac{\beta\alpha}{1-\beta}R(\tau)
\end{aligned}$$

The problem of the planner then becomes

$$\begin{aligned}
\alpha(1-f)\left\{u(x) - x - V\left(Z(\tau) - \frac{R(\tau)}{1-f}\right)\right\} &= \\
\alpha(1-f)\left\{\frac{\tau V\pi}{1-\tau V\pi}\left(1+\frac{\beta\alpha}{1-\beta}\right)R(\tau) - \frac{\tau V\pi}{1-\tau V\pi}\frac{\beta\alpha}{1-\beta}R(\tau) + V\frac{R(\tau)}{1-f}\right\} &= \\
\alpha(1-f)\left\{\frac{\tau\pi}{1-\tau V\pi} + \frac{1}{1-f}\right\}VR(\tau) &= \\
\alpha(1-f)\left\{\pi\frac{\tau}{1-\tau V} + \frac{1}{1-f}\right\}V\left(\frac{\tau(1-f)\bar{c}_s + \tau c_v}{1-\tau}\right)
\end{aligned}$$

This is strictly increasing in both V and τ. Hence the solution is $V = 1-f$ and $\tau = 1$. However, this implies $u(x) \to \infty$ and $x \to \pm\infty$. If $x \to -\infty$, we get a contradiction. If

$x \to +\infty$, the planner's objective function is

$$\alpha(1-f)\left\{u(x) - x - V\left(\bar{c}_s + c_v + \frac{f}{1-f}c_v\right)\right\} \to -\infty$$

which cannot be optimal. Therefore, (13) and (14) cannot both be binding. This shows that only (14) binds. □

D Proof of Proposition 2

Proof. We concentrate on the results for internal validation. The first order conditions with respect to x (23) and the one with respect to V (24) are

$$\begin{aligned} [u'(x) - 1](1 + \delta\lambda) - \lambda\frac{1}{\hat{\tau}V} &= 0 \quad (23) \\ \left[\frac{R(\hat{\tau})}{1-f} - Z(\hat{\tau})\right] + \lambda\frac{x}{\hat{\tau}V^2} - \delta\lambda Z(\hat{\tau}) - \lambda_{1-f} &\leq 0 \quad (24) \end{aligned}$$

where λ is the Lagrange multiplier on the validators' IC constraint, and λ_{1-f} is the one on $V \leq 1 - f$.

We first consider the solution when $\hat{\tau} \to 1$. Then

$$\frac{R(\hat{\tau})}{1-f} - Z(\hat{\tau}) = -(\bar{c}_s + c_v) - \frac{f}{1-f}c_v$$

while $Z(\hat{\tau}) \to \infty$. We now show that λ and V both converge to zero.

From (24) we obtain either $\lambda \to 0$ and/or $V \to 0$. It is clear that if $V > 0$ and $\lambda > 0$ then the LHS of (24) is necessarily negative as $\tau \to 1$ so that $V = 0$, a contradiction. Now rewrite (24) as

$$\frac{(\bar{c}_s + c_v) + \frac{f}{1-f}c_v}{\left[\frac{x}{\hat{\tau}V^2} - \delta Z(\hat{\tau})\right]} = \lambda$$

Since $\lambda \geq 0$ we obtain $x \geq \tau\delta V^2 Z(\hat{\tau})$. Since x is bounded from above but $Z(\hat{\tau}) \to \infty$, we must have $V \to 0$ as $\hat{\tau} \to 1$. Further, since (13) never binds, it must be that $VZ(\tau)$

converges to a positive constant.[42] Therefore,

$$\frac{\lambda}{V} = \frac{(\bar{c}_s + c_v) + \frac{f}{1-f} c_v}{\left[\frac{x}{\hat{\tau}} - \delta V^2 Z(\hat{\tau})\right]} V \overset{\hat{\tau}\to 1}{\to} 0.$$

Then (23) implies $x \to x^*$. When the solution for $\hat{V}$ is interior, we can simplify (23) and (24) to obtain

$$\frac{\left[Z(\hat{\tau}) - \frac{R(\hat{\tau})}{1-f}\right]}{\left[\frac{x}{\hat{\tau}V^2} - \delta Z(\hat{\tau})\right]} = \lambda$$

and

$$u'(x) - 1 = \frac{\lambda}{1 + \delta\lambda} \frac{1}{\hat{\tau} V}$$

or

$$u'(x) = 1 + \frac{V Z(\hat{\tau})}{x} - \frac{R(\hat{\tau}) V}{x(1-f)(1+\delta\lambda)}$$

The right hand side of the above equation is always higher than 1, so that generically $x < x^*$. The constrained optimal solution is $(\hat{x}, \hat{V})$ that solves (23), (24) and (19) holds with equality. $\hat{\tau}$ is given by (20).

Consider the case where $V = 1 - f$. Using (20), it is easy to check that $\tau = \left(1 + \sqrt{\frac{C}{x+C}}\right)^{-1}$.

[42] From (19) holding at equality,

$$\begin{aligned}
\delta\left[u(x) - x\right] &= \left(\frac{1}{\tau V}\right)\left[x + V Z(\tau)\right] - \delta\left[R(\tau) - V Z(\tau)\right] \\
\delta V\left[u(x) - x\right] &= \left(\frac{1}{\tau}\right) x + V Z(\tau)/\tau - \delta\left[1 - f - V\right] V Z(\tau) - V\delta\left[(1-f)\bar{c}_s + c_v\right]
\end{aligned}$$

and taking the limit as $\tau \to 1$ (which implies $V(\tau) \to 0$), since $u(x) - x$ is bounded,

$$\begin{aligned}
0 &= x^* + \lim_{\tau\to 1} V(\tau) Z(\tau) - \delta\left[1 - f - V(\tau)\right] V(\tau) Z(\tau) \\
0 &= x^* + \left[1 - \delta(1-f)\right] \lim_{\tau\to 1} V(\tau) Z(\tau) + \delta \lim_{\tau\to 1} V(\tau)^2 Z(\tau)
\end{aligned}$$

Suppose $V(\tau)^2 Z(\tau)$ would converge to a strictly positive constant. Then $\lim_{\tau\to 1} V(\tau) Z(\tau)$ would converge to $+\infty$ which would violate the equality above. Hence,

$$\lim_{\tau\to 1} V(\tau) Z(\tau) = \frac{x^*}{\delta(1-f) - 1}.$$

When $V = 1 - f$ so that $\lambda_{1-f} > 0$, the first order condition gives

$$\begin{aligned} [u'(x) - 1](1 + \delta\lambda) - \lambda\frac{1}{\hat{\tau}V} &= 0 \quad (25) \\ \left[\frac{R(\hat{\tau})}{1-f} - Z(\hat{\tau})\right] + \lambda\frac{x}{\hat{\tau}V^2} - \delta\lambda Z(\hat{\tau}) &> 0 \quad (26) \end{aligned}$$

Using (25) to eliminate λ, the definition $R(\tau)$ and $Z(\tau)$, as well as the expression for τ, (26) becomes

$$\left(\sqrt{\frac{(1-f)C}{x+(1-f)C}}\right)\left\{\frac{1+\sqrt{\frac{(1-f)C}{x+(1-f)C}}}{(1-f)}\right\}\left[\frac{x}{(1-f)C} - \frac{1}{[u'(x)-1]}\right] > \delta \quad (27)$$

where x is given by the incentive constraint of validators holding at equality, which we can write as

$$\delta\left[u(x) - x - (1-f)C\right] = C\left(1 + \sqrt{\frac{(1-f)C}{x+(1-f)C}}\right)\left[\frac{x}{(1-f)C} + \frac{(1+\sqrt{\frac{(1-f)C}{x+(1-f)C}})}{\sqrt{\frac{(1-f)C}{x+(1-f)C}}}\right] \equiv H(x)$$

It is tedious to check that $H'(x) > 0$ and using the implicit function theorem that $dx/d\delta > 0$. The left hand side of (27) is decreasing in x if

$$\frac{1}{u'(x)} + \sqrt{\frac{\rho}{xu'(x)}(1-f)C} < 1,$$

where $\rho = -u''(x)x/u'(x)$. This will hold if C and x are small enough and $\rho \geq 1$ so that $xu'(x)$ is decreasing in x. Assuming this is the case, since $dx/d\delta > 0$, x declines as δ decreases. So the LHS of (27) increases when δ decreases and the condition will be satisfied for δ low enough and below some $\bar{\delta}$.

Finally, as δ keeps decreasing below $\bar{\delta}$, $x \to 0$ so that $H(x) \to 4C$, while $\delta\left[u(x) - x - (1-f)C\right] < 0$. Therefore, there is a δ_0 below which there is no $x \in \mathbb{R}$ that satisfies the incentive constraint of validators. □

E Permissionless validation with free entry

In this section, we analyze the case where any agent can become a validator, although at a cost. The fixed cost of entry is set to ε that agents pay once and for all — for instance, buying the necessary computer equipment and connecting to the network. Since the validation is permissionless, as we assume that given there are V validators, one is selected at random to validate a trade. With α trade to validate, a validator has a probability α/V of being selected to validate a trade. As in the paper, any validator incurs the cost c_v to validate the trade and c_s to write and make the new ledger available to the community of validators. In exchange the validators gets a fee z as a compensation for the work. We set $C = c_s + c_v/(1-f)$.

Given V, the expected payoff of a validator then is $\bar{U}_F$ defined by

$$(1-\beta)\bar{U}_F = \frac{\alpha}{V}(1-f)\,(z-C)$$

Participation implies the restriction, $\bar{U}_F \geq 0$. Free entry implies validators will enter as long as $-\varepsilon + \bar{U}_F \geq 0$. This equation holding with an equality pins down the number of validators in equilibrium,

$$V = \frac{\alpha(1-f)\,(z-C)}{(1-\beta)\varepsilon} \tag{28}$$

We assume that validators can be prevented from downloading the blockchain if they are caught cheating (this is the best case scenario for permissionless validation). Validators do not accept a bribe $\bar{z}$ whenever

$$\beta\bar{U}_F \geq \bar{z} + (1-\pi)\beta\bar{U}_F.$$

Using $\bar{z} = y + z$ as well as $\bar{U}_F = \varepsilon$, we obtain

$$\pi\beta\varepsilon \geq (y+z)\,.$$

Using the PC of early producers $y \geq x$ at equality, an allocation is incentive feasible whenever

it satisfies,

$$u(x) - x - z \geq 0 \tag{29}$$

$$z \geq C \tag{30}$$

$$\pi\beta\varepsilon \geq x + z \tag{31}$$

The planner wants to maximize

$$\alpha(1-f)\max_{x,z}\{u(x) - x - V\left(C + (1-\beta)\varepsilon\right)\}$$

subject to the three constraints above. Using (28), we can easily see that the planner's objective function is decreasing in z, so the planner will set $z = C$ (which implies $V \to 0$) and being constrained by (31), the allocation will be

$$x = \begin{cases} x^* & \text{if } x^* < \pi\beta\varepsilon - C \\ \pi\beta\varepsilon - C & \text{otherwise} \end{cases}$$

Hence permissionless ledgers will implement the efficient allocation x^* as long as it is small enough relative to the entry cost into validation.

F The validation game

In this section, we specify the details of the validation game for internal validation. We analyze a free-rider problem inherent to the validation protocol: validators have incentives to abstain from verifying a label, but still send the message that the label of the producer is G. The severity of this free rider problem could undermine the existence of an equilibrium with trade, as the ledger would lose integrity.[43]

We keep some of the features of the optimal allocation. In particular, as payments are indeterminate, we simplify notation by using $z = z^1 + z^2$. Also, recall that as $\varepsilon \to 0$

[43] See also Amoussou-Guénou, et al. (2019).

and absent the free-rider problem, all validators should be expected to work as long as the communication cost is lower than some threshold (that we set at $\bar{c}_s$).

F.1 The free-rider problem

In this section, we describe the validation game that validators play in detail. In the first stage, a strategy for validator i consists of a verification strategy, $\nu_i \in [0,1]$, a voting strategy $\sigma_i \in [0,1]$ which is the probability to send a message and the validator i choice of message $m_i \in \{\emptyset, 0, 1\}$ to send. We call *shirkers* those validators who do not verify labels, and we call *workers* those validators who do. Define $\mathbf{m} = (m_1, m_2, ..., m_V)$ and $\mathcal{I}(\mathbf{m}) = 1$ if $\int_{i=1}^{V} m_i \mathbb{I}_{m_i \neq \emptyset} di > \tau V$ and $\mathcal{I}(\mathbf{m}) = 0$ otherwise.[44]

The public history at the end of period t consists of the public history h_t at the end of period $t-1$, as well as the result of the validation process $\mathcal{I}(\mathbf{m})$ and the production of the late producer, that we can summarize with the label of the producer $\ell \in \{G, B\}$.[45] We focus on strategies for validators that depend only on the public history and the information acquired during the current period. The equilibrium concept is Bayesian perfection: strategies are Nash equilibrium given the information validators have and validators are Bayesian so that they update their belief using Bayes' rule.

Recall that validators are dealing with legitimate late producers with probability $1-f$ and the probability that the producer is faulty is f. Also, recall that validators who do not send a message are not entitled to a payment. Given the allocation (x, y, z) is incentive feasible – so that a late producer who is found to have label G will produce for the early producer –

[44] Again we assume that $m_{i,k} = \emptyset$ counts as $m_{i,k} = 0$.

[45] Since we concentrate on incentive feasible allocations (x, y, z) the producer's label is a sufficient statistics for the outcome in a match because a late producer with label G will produce so that the early producer will also produce, while a late producer with label B is not expected to produce so that the early producer will not produce.

the expected payoff of a working validator from validating the transaction is

$$
\begin{aligned}
-c_v + (1-f) & \left\{ \begin{array}{c} \sigma_i(G)\left(-c_s + E_i\left[\mathcal{I}(\mathbf{m})z + \beta U_V(h_t,(\mathcal{I}(\mathbf{m}),G)) \mid m_i = 1\right]\right) \\ +(1-\sigma_i(G))E_i\left[\mathcal{I}(\mathbf{m})\times 0 + \beta U_V(h_t,(\mathcal{I}(\mathbf{m}),G)) \mid m_i = \emptyset\right] \end{array} \right\} \\
+f & \left\{ \begin{array}{c} \sigma_i(B)\left(-c_s + E_i\left[\mathcal{I}(\mathbf{m})\times 0 + \beta U_V(h_t,(\mathcal{I}(\mathbf{m}),B)) \mid m_i = 1\right]\right) \\ +\left(1-\sigma_i(B)\right)E_i\left[\mathcal{I}(\mathbf{m})\times 0 + \beta U_V(h_t,(\mathcal{I}(\mathbf{m}),B)) \mid m_i = \emptyset\right] \end{array} \right\}
\end{aligned} \tag{32}
$$

where $U_V(h_t,(\mathcal{I}(\mathbf{m}),\ell))$ is the continuation payoff of the validator given the new history $(h_t,(\mathcal{I}(\mathbf{m}),\ell))$ and $\ell \in \{G,B\}$. $U_V(\cdot) = U_V$ whenever the continuation payoff does not depend on the validator's actions. Also, we assume that if the majority agrees that the producer has label G, then validation and communication happens in the second stage.

We now explain the different elements in (32). In the early stage, the working validator incurs cost c_v to verify the label. If the label is G (which happens with probability $1-f$) the validator sends message $m_i = 1$ with probability $\sigma_i(G)$ and nothing otherwise (again, we anticipate that not sending messages $m_i = \emptyset$ is better than sending $m_i = 0$, since it communicates the same information at a lower cost). If the index $\mathcal{I}(\mathbf{m}) = 1$,[46] the match is validated and trade can take place. Then validators who sent a message get z from the late producer, and they verify production takes place in stage 2 and communicate the result to the ledger. If $\mathcal{I}(\mathbf{m}) = 0$, the transaction is not validated and working validators get nothing. E_i is the expectation of validator i over the index function $\mathcal{I}(\mathbf{m})$ given the validator's information summarized by message m_i. With probability f, the working validator learns the producer's label is B. Then with probability $\sigma_i(B)$ the validator sends message $m_i = 1$ but he expects to receive zero, even if (at least) τ validators send $m = 1$ because he knows the buyer has a bad label and will not produce. With probability $1-\sigma_i(B)$, validator i sends no message (or message 0), and does not expect any payments. In any case, the working validator knows production will not take place in stage 2 and so he does not verify or communicate anything in stage 2. Notice that in this section, the expected future payoff of validators $U_V(h_t,(\mathcal{I}(\mathbf{m}),\ell))$ where $\ell \in \{G,B\}$ only depend on public history and so can vary depending on the outcome of the validation process. It should be obvious that $\sigma_i(B) = 0$: a worker will not send

[46] $\mathcal{I}(\mathbf{m}) = 1$ if at least $\tau - 1$ other validators send $m = 1$ if $m_i = 1$, and at least τ other validators send $m = 1$ if $m_i = \emptyset$.

message $m_i = 1$ for a producer with label B.

The expected payoff of a shirker is

$$\bar{\sigma}_i \left\{ \begin{array}{c} (1-f)\left(-c_s^1 + E_i\left[\mathcal{I}(\mathbf{m})z + \beta U_V(h_t, (\mathcal{I}(\mathbf{m}), G)) \mid \bar{m}_i = 1\right]\right) \\ +f\left(-c_s^1 + E_i\left[\mathcal{I}(\mathbf{m}) \times 0 + \beta U_V(h_t, (\mathcal{I}(\mathbf{m}), B)) \mid \bar{m}_i = 1\right]\right) \end{array} \right\}$$
$$+ (1-\bar{\sigma}_i) \left\{ \begin{array}{c} (1-f)E_i\left[\mathcal{I}(\mathbf{m}) \times 0 + \beta U_V(h_t, (\mathcal{I}(\mathbf{m}), G)) \mid \bar{m}_i = \emptyset\right] \\ +fE_i\left[\mathcal{I}(\mathbf{m}) \times 0 + \beta U_V(h_t, (\mathcal{I}(\mathbf{m}), B)) \mid \bar{m}_i = \emptyset\right] \end{array} \right\}$$

A shirker does not know the buyer's label before sending their message $\bar{m}_i$. Again, given index $\mathcal{I}(\mathbf{m})$, the future payoff of validators is the same for all validators irrespective of their action. Note that a shirker only shirks in period 1. Since she observes the index $\mathcal{I}(\mathbf{m})$, she can learn the label of the producer and she can verify production and send a message relative to that production in period 2 only if the label is G.

Now suppose $\tau < 1$. We can show that in equilibrium, not all validators will be working/cooperating.

Lemma 3. *Suppose validation does not require unanimity $\tau < 1$. There is an equilibrium where **all** validators work whenever*

$$f \geq \frac{c_v}{c_s}.$$

Proof. Suppose all validators are working and send message $m = 1$ if the label is G and do not send a message otherwise. Since $\tau < V$, any validator i is not pivotal, because changing the value of one message will not change the overall index value. Then the value of working and sending $m = 1$ if the label is G and $m = \emptyset$ otherwise is

$$-c_v + (1-f)\left(-c_s + z + \beta U_V(h_t, (1, G))\right) + f\beta U_V(h_t, (0, B)) \tag{33}$$

while the expected value of shirking is

$$\bar{\sigma}_i \left\{(1-f)\left(-c_s + z + \beta U_V(h_t, (1, G))\right) + f\left(-c_s + \beta U_V(h_t, (0, G))\right)\right\}$$
$$+ (1-\bar{\sigma}_i)\left\{(1-f)\beta U_V(h_t, (1, G)) + f\beta U_V(h_t, (0, B))\right\}$$

So a shirker sends $m = 1$ whenever the expected payment is greater than the cost of always sending a message:

$$(1 - f)z \geq c_s.$$

Using (10) at equality to replace for z, a shirker sends $m = 1$ whenever

$$\tau \geq f - \frac{c_v}{c_s}.$$

So when $\tau < f - \frac{c_v}{c_s}$, shirkers prefer to send no message and they never get a payment. So in this case, working always give a higher payoff to validators than shirking (and not sending a message). If $\tau \geq f - \frac{c_v}{c_s}$, shirkers are better off sending a message. Then working gives a higher payoff than shirking (and sending a message) when $f \geq \frac{c_v}{c_s}$. □

Stated slightly differently, Lemma 3 says that there is a free-rider problem whenever $f < c_v/c_s$. This is intuitive: when $fc_s < c_v$, free-riders who expect at least τ validators to work save the verification cost c_v but incur the cost of sending a message c_s when they should not send it (when the producer is faulty). As a corollary, we deduce that incentives to free-ride are high whenever $c_s \to 0$, because the cost of sending a message when one should not is negligible.

Notice that the only punishment that free-rider incurs is the cost of sending a message when the producer is faulty in which case they will not receive a payment. We now look at other forms of punishments. First, the worst punishment when payoffs can only depend on public history, is that the system shuts down if, collectively, validators make a mistake. This means that $U_V(h_t, (1, B)) = 0$. A late producer with label B will not produce and so the system will detect that the validation process was flawed. However, a late producer with label G who was assigned the wrong label will not produce (because the early producer will not produce) and so will not be distinguished from a late producer with label B. In this case the system cannot detect the flawed validation. So we must have $U_V(h_t, (0, G)) = U_V(h_t, (0, B))$. We define a *uniform* mechanism as one that gives validators the same continuation payoff following these "observationally equivalent" outcomes and when the validation process resulted in a correct

outcome, that is

$$U_V(h_t,(0,G)) = U_V(h_t,(0,B)) = U_V(h_t,(1,G)) \equiv U_{Vt}.$$

Following the steps in the proof of Lemma 3, we can conclude that uniform mechanisms do not relax the free-rider problem.

F.2 Individual mechanisms

We now define an individual mechanism as one where both the current payoff and the continuation value depend on the publicly observable action of validators. To be precise, we consider that the ledger assigns label B to a validator who is caught sending a message that differs from the "supermajority" τ of validators. As a result, this validator loses the ability to validate but also the opportunity to trade in the future. Such mechanisms specify individual continuation values $U_V(h_t,(\mathcal{I}(\mathbf{m}),\ell);m)$ as a function of the result of the validation process $\mathcal{I}(\mathbf{m})$, whether the producer produced or not $\ell \in \{G,B\}$ and the validators' message m. A validator goes against the majority whenever $\mathcal{I}(\mathbf{m}) \neq m$. In this case, the worse punishment is the level of utility the validator would obtain in permanent autarky. So we set

$$U_V(h_t,(\mathcal{I}(\mathbf{m}),\ell);m) = \begin{cases} 0 & \text{if } \mathcal{I}(\mathbf{m}) \neq m, \\ U_{Vt} & \text{otherwise.} \end{cases}$$

Then we show

Lemma 4. *Using an individual mechanism and $\tau < 1$, there is an equilibrium where all validators work whenever*

$$f \geq \frac{c_v}{c_s + \beta U_{Vt}}$$

Proof. Suppose at least τV validators work. If one of the remaining validators shirks, he

obtains expected payoff

$$\begin{aligned}\bar{\sigma}_i \{(1-f)(-c_s + z + \beta U_V(h_t,(1,G),m=1)) + f(-c_s + \beta U_V(h_t,(0,B),m=1))\} \\ + (1-\bar{\sigma}_i)\{(1-f)\beta U_V(h_t,(1,G),m=0) + f\beta U_V(h_t,(0,B),m=0)\} \quad = \\ \\ \bar{\sigma}_i\{-c_s + (1-f)z\} + \bar{\sigma}_i(1-f)\beta U_{Vt} + (1-\bar{\sigma}_i)f\beta U_{Vt}\end{aligned}$$

If the buyer has a good label, all other validators send $m_i = 1$, so $\mathcal{I}(\mathbf{m}) = 1$ irrespective of the decision of the shirker. However, the shirker only gets the reward if he also sends $m = 1$. If he sends a message when the producer has a bad label, he does not receive a reward and he gets the autarkic payoff in the future. Therefore, a shirker sends a signal ($\sigma_i = 1$) whenever

$$-c_s + (1-f)(z + \beta U_{Vt}) > f\beta U_{Vt}$$

and does not otherwise.

The expected utility of a working validator (who sends signal G if the producer has label G and nothing if the producer has label B) is as before,

$$\begin{aligned}-c_v + (1-f)(-c_s + z + \beta U_V(h_t,(1,G),m=1)) + f\beta U_V(h_t,(0,B),m=0) \quad = \\ -c_v + (1-f)(-c_s + z) + \beta U_{Vt}\end{aligned}$$

The remaining validator will work whenever

$$\begin{aligned}-c_v + (1-f)(-c_s + z) &> -c_s + (1-f)z - f\beta U_{Vt} \\ -c_v + (1-f)(-c_s + \hat{z}) &> -c_s + (1-f)\hat{z} + (1-f)\beta U_{Vt} - \beta U_{Vt} \\ c_s + \beta U_{Vt} - c_v &> (1-f)(c_s + \beta U_{Vt}) \\ 1 - \frac{c_v}{c_s + \beta U_{Vt}} &> (1-f)\end{aligned}$$

So if $(1-f) \geq \frac{f(c_s+\beta U_{Vt})}{(z-c_s+\beta U_{Vt})}$, the remaining validator works whenever

$$1 - \frac{c_v}{c_s + \beta U_{Vt}} > (1-f) \geq \frac{f\left(c_s + \beta U_{Vt}\right)}{(z - c_s + \beta U_{Vt})},$$

and he would send a signal if he were to shirk.

However, if $-c_s + (1-f)\left(z + \beta U_{Vt}\right) < f\beta U_{Vt}$ we have $\bar{\sigma}_i = 0$ and in this case the remaining validator decides to work whenever

$$\begin{aligned} -c_v + (1-f)\left(-c_s + z\right) &> -(1-f)\beta U_{Vt} \\ (1-f)\left(-c_s + z + \beta U_{Vt}\right) &> c_v \\ 1 - f &> \frac{c_v}{z - c_s + \beta U_{Vt}}. \end{aligned}$$

So if

$$\frac{f\left(c_s + \beta U_{Vt}\right)}{z - c_s + \beta U_{Vt}} > (1-f) > \frac{c_v}{z - c_s + \beta U_{Vt}}$$

the remaining validator works (and he would not send a signal if he were to shirk). Notice that this case is only possible if $f\left(c_s + \beta U_{Vt}\right) > c_v$, or

$$1 - \frac{c_v}{c_s + \beta U_{Vt}} > 1 - f$$

Therefore, combining both conditions, validators prefer to work whenever

$$1 - \frac{c_v}{c_s + \beta U_{Vt}} > 1 - f > \frac{c_v}{z - c_s + \beta U_{Vt}}$$

Replacing z using (11), we obtain

$$\begin{aligned} (1-f)(z - c_s + \beta U_{Vt}) &> c_v \\ (1-f)\left[\frac{1}{1-\tau}\left(c_s + \frac{c_v}{1-f}\right) - c_s + \beta U_{Vt}\right] &> c_v \\ \frac{\tau}{1-\tau}\left((1-f)c_s + c_v\right) + (1-f)\beta U_{Vt} &> 0 \end{aligned}$$

which always true. Hence, validators prefer to work whenever

$$f > \frac{c_v}{c_s + \beta U_{Vt}}$$

This concludes the proof. □

When validators work, they lose c_v, but they send the right signal. When they don't work, either they prefer to never send a signal, or they send a signal. When (uninformed) shirkers prefer not to send a signal they effectively vote that the producer has label B. Therefore they often get it wrong when there are many good producers. In this case, validators prefer working than shirking whenever their loss c_v is less than the expected net loss of not sending a signal when they should $(1-f)(z - c_s + \beta U_{Vt})$. But if (uninformed) shirkers prefer to send a signal, they will get it wrong with probability f in which case they lose $c_s^1 + \beta U_{Vt}$. So they prefer to work whenever the expected loss of sending a wrong signal $f(c_s + \beta U_{Vt})$ is higher than the verification cost.

Notice that validators working can now be an equilibrium even if $c_s = 0$. So we have the following Folk's theorem

Lemma 5. *[**"Folk" Theorem**] Let $\beta \to 1$. Using an individual mechanism and $\tau < 1$, there is an equilibrium where all validators work whenever the late producer's participation constraint (13) is satisfied,*

$$u(x) > x + VZ(\tau)$$

Proof. The existence of the equilibrium requires

$$f \geq \frac{c_v}{c_s + \beta U_{Vt}}$$

In equilibrium, $\beta \to 1$ implies that $\beta U_V \to \infty$ as long as $u(x) > x + VZ(\tau)$. Then $\frac{c_v}{c_s + \beta U_{Vt}} \to 0$. Therefore, validators work whenever $f \geq 0$. So validators always work as long as $u(x) > x + VZ(\tau)$. □

G Case with non-degenerate cost distribution

In this Appendix, we consider the case where the distribution of the common cost c_s is non degenerate. We show that a weak sufficient condition for the planner to choose $c_s^* = \bar{c}_s$ defined as $c_s^* \equiv (1-\tau)(z^1+z^2) - \frac{c_v}{1-f}$ is

$$\frac{\bar{c}_s + \frac{c_v}{1-f}}{(1-f)(\bar{c}_s - Ec_s)} \geq \delta.$$

In this case, all validators will work, irrespective of their private communication costs.

Let c_s^* be defined as above. So validators only verify a trade whenever $c_s \leq c_s^*$. When c_s is uniformly distributed over $[0, \bar{c}_s]$ the probability that validators verify a trade is simply the probability that $c_s \leq c_s^*$, that is $c_s^*/\bar{c}_s$. Validators verify whenever $c_s \leq c_s^*$ and these validators take home

$$Z(\tau) \equiv \frac{c_s^* + c_v/(1-f)}{1-\tau}$$

When $z^1 = Z$, the participation constraint of validators (47) is always satisfied

$$\mathbb{E}_{c_s \leq c_s^*}\left[-c_v + (1-f)(Z - c_s)\right] \geq 0$$

Let $R(\tau, c_s)$ be the expected rent of validators when the fundamental communication cost is $c_s \leq c_s^*$,

$$\begin{aligned} R(\tau, c_s) &\equiv (1-f)\left[Z(\tau) - c_s\right] - c_v \\ &\equiv (1-f)\left[\frac{c_s^*}{1-\tau} - c_s\right] + \frac{\tau c_v}{1-\tau} \end{aligned}$$

We can set the participation constraint for early producers (44) at equality and replace $y = x + wVZ(\tau)$. Then the set of IF allocations is characterized by

$$\delta \frac{c_s^*}{\bar{c}_s}\left[u(x) - (x + VZ(\tau))\right] \geq (x + VZ(\tau)) \tag{34}$$

$$\delta \int_0^{c_s^*} \left[u(x) - (x + VZ(\tau)) + R(\tau, c_s)\right] \frac{dc_s}{\bar{c}_s} \geq \frac{1}{\tau V}(x + VZ(\tau)) \tag{35}$$

Notice that validators only work with probability $\frac{c_s^*}{\bar{c}_s}$. Therefore, producers (including validators) can trade only with probability $\frac{c_s^*}{\bar{c}_s}$.

G.1 Optimal design

We need to adapt the objective function to the new setup. Since the probability of a productive match is $1-f$ in each period, the objective function of a planner is the sum of the early and late producers' utility when they trade, and the expected rent of validators from operating the ledger for a measure α of trades,

$$\int_0^{c_s^*} \left\{\alpha(1-f)\left[u(x)-y+\left(y-x-Vz^1\right)\right]+\alpha V R(\tau,c_s)\right\}\frac{dc_s}{\bar{c}_s},$$

or replacing y and z^1, as well as $Z(\tau)-R(\tau,c_s)/(1-f)=c_s+\frac{c_v}{1-f}>0$, a planner chooses the trading size x, the number of validators V, the threshold τ, and the threshold c_s^* to solve

$$\alpha(1-f)\max_{x,V\geq 0, c_s^*\leq \bar{c}_s, \tau\in[0,1]} \int_0^{c_s^*}\left\{u(x)-x-V\left(c_s+\frac{1}{1-f}c_v\right)\right\}\frac{dc_s}{\bar{c}_s}$$

subject to (53) and (54). Using the same steps as the simpler case, we can show that (53) is always slack when (54) holds. Re-arranging the constraint,

$$\pi\frac{\beta\alpha}{1-\beta}\frac{c_s^*}{\bar{c}_s}\left[u(x)-x\right]\geq\left(\frac{1}{\tau V}\right)\left(x+VZ(\tau)\right)-\frac{\pi\beta\alpha}{1-\beta}\int_0^{c_s^*}\left(R(\tau,c_s)-VZ(\tau)\right)\frac{dc_s}{\bar{c}_s} \tag{36}$$

Since the objective function is independent of τ, the planner will choose τ to minimize the right hand side of (36). The first order condition for the optimal threshold $\hat{\tau}$ gives

$$\frac{1-\hat{\tau}}{\hat{\tau}}=\sqrt{\frac{\left[1-\delta\frac{c_s^*}{\bar{c}_s}\left(1-f-V\right)\right]\left(c_s^*+\frac{c_v}{(1-f)}\right)}{\frac{x}{V}+c_s^*+\frac{c_v}{(1-f)}}} \tag{37}$$

This threshold is well defined if

$$1>\delta\frac{c_s^*}{\bar{c}_s}\left(1-f-V\right)$$

and otherwise $\hat{\tau} = 1$.

Then it is useful to look at the first order conditions in detail. When λ is the Lagrange multiplier on the validators' IC constraint, the first order conditions with respect to x, V and c_s^* respectively are -

$$\begin{aligned}
[u'(x) - 1](1 + \delta\lambda)\frac{c_s^*}{\bar{c}_s} - \lambda\frac{1}{\hat{\tau}V} &= 0 && (38)\\
\int_0^{c_s^*}\left[\left(c_s + \frac{1}{1-f}c_v\right)\right]\frac{dc_s}{\bar{c}_s} + \lambda\frac{x}{\hat{\tau}V^2} - \delta\lambda\frac{c_s^*}{\bar{c}_s}Z(\hat{\tau}) &= 0 && (39)\\
\left\{u(x) - x - V\left(Z(\tau) - \frac{R(\tau, c_s^*)}{1-f}\right)\right\}\frac{1}{\bar{c}_s} && && (40)\\
+\lambda\left\{\begin{array}{c}\delta\left[\{u(x) - x - VZ(\tau)\} + R(\tau, c_s^*)\right]\frac{1}{\bar{c}_s}\\ +\delta\int_0^{c_s^*}\left[-V\frac{\partial Z(\tau)}{\partial c_s^*} + \frac{\partial R(\tau, c_s)}{\partial c_s^*}\right]\frac{dc_s}{\bar{c}_s} - \frac{1}{\tau V}V\frac{\partial Z(\tau)}{\partial c_s^*}\end{array}\right\} &\geq 0 && (41)
\end{aligned}$$

We now determine conditions so that the solution is $c_s^* = \bar{c}_s$. Suppose this is the case. Then the expression in $\{.\}$ in (41), the second part of the FOC which is multiplied by λ (which pertains to the behavior of the IC when the planner increases c_s^*) is

$$\delta\frac{1}{\tau V}x\frac{1}{\bar{c}_s} + \frac{1}{\tau(1-\tau)}\left\{\delta\left[\bar{c}_s + \frac{c_v}{1-f}\right]\frac{1}{\bar{c}_s} - 1\right\} + \delta\int_0^{\bar{c}_s}\left[(1-f-V)\frac{1}{1-\tau}\right]\frac{dc_s}{\bar{c}_s}$$

Hence, if

$$\delta\left[1 + \frac{c_v}{\bar{c}_s(1-f)}\right] \geq 1$$

then the LHS of the IC is increasing with c_s^* and it is optimal to set $c_s^* = \bar{c}_s$, as long as the objective function is also increasing in c_s^* when evaluated at $\bar{c}_s$, that is

$$u(x) - x - V\left(Z(\tau) - \frac{R(\tau, \bar{c}_s)}{1-f}\right) \geq 0$$

From (36)

$$\delta\left[u(x) - x + \underbrace{\int_0^{\bar{c}_s} R(\tau, c_s)\frac{dc_s}{\bar{c}_s}}_{=ER(\tau, c_s)} - VZ(\tau)\right] = \frac{1}{\tau V}\left(x + VZ(\tau)\right)$$

Hence,

$$\begin{aligned} u(x) - x - V\left(Z(\tau) - \frac{R(\tau, \bar{c}_s)}{1-f}\right) &= \\ \frac{1}{\delta\tau V}\left(x + VZ(\tau)\right) + V\frac{R(\tau, \bar{c}_s)}{1-f} - ER(\tau, c_s) &= \\ \frac{1}{\delta\tau V}x + V\frac{R(\tau, \bar{c}_s)}{1-f} + \left[\frac{1}{\delta\tau} - (1-f)\right]\left(\bar{c}_s + \frac{c_v}{1-f}\right) + (1-f)Ec_s + c_v & \end{aligned}$$

Since $\tau \leq 1$, and a sufficient condition for the RHS to be positive is

$$\begin{aligned} \frac{1}{\delta}\left(\bar{c}_s + \frac{c_v}{1-f}\right) - (1-f)\left(\bar{c}_s + \frac{c_v}{1-f}\right) + (1-f)\left(Ec_s + \frac{c_v}{1-f}\right) &\geq 0 \\ \frac{1}{\delta}\left(\bar{c}_s + \frac{c_v}{1-f}\right) - (1-f)\left(\bar{c}_s - Ec_s\right) &\geq 0 \\ \frac{\bar{c}_s + \frac{c_v}{1-f}}{(1-f)\left(\bar{c}_s - Ec_s\right)} &\geq \delta. \end{aligned}$$

Therefore, f large enough (for example) would allow the inequality to be satisfied. Also, if Ec_s is close enough to $\bar{c}_s$. In those cases, $c_s^* = \bar{c}_s$, and the solution if given by the FOC of the planner's problem,

$$\begin{aligned} \left[u'(x) - 1\right](1 + \delta\lambda) - \lambda\frac{1}{\hat{\tau}V} &= 0 \\ Ec_s + \frac{1}{1-f}c_v + \lambda\frac{x}{\hat{\tau}V^2} - \delta\lambda Z(\hat{\tau}) &= 0 \end{aligned}$$

together with the binding IC,

$$\delta\left[u(x) - x\right] = \left(\frac{1}{\tau V}\right)\left(x + VZ(\tau)\right) - \delta\int_0^{\bar{c}_s}\left(R(\tau, c_s) - VZ(\tau)\right)\frac{dc_s}{\bar{c}_s} \tag{42}$$

H Concave utility for early producers

In this Appendix, we lay down the analysis when early producers also have a concave utility function. We analyze the case when the distribution of the communication cost is degenerate at $\bar{c}_s$. Participation constraints are

$$u(x) - y - Vz^2 \geq 0 \tag{43}$$

$$v(y) - x - Vz^1 \geq 0 \tag{44}$$

$$-c_v + (1-f)\left(z^1 + z^2 - c_s\right) \geq 0 \tag{45}$$

From the PC of early producers holding at equality,

$$y = v^{-1}(x + Vz^1) \equiv \Phi(x + Vz^1)$$

where Φ is increasing and convex, and the PCs become

$$u(x) - \Phi(x + Vz^1) - Vz^2 \geq 0 \tag{46}$$

$$-c_v + (1-f)\left(z^1 + z^2 - c_s\right) \geq 0 \tag{47}$$

Repayment constraints:

$$-\left(\Phi(x + Vz^1) + Vz^2\right) + \beta U \geq 0. \tag{48}$$

$$-\left(\Phi(x + Vz^1) + Vz^2\right) + \beta \mathbb{E}U_V \geq 0. \tag{49}$$

No bribe.

$$\pi\beta U_V \geq \bar{z}$$

When a share w of validators are working on a match, the late producer in this match is willing to pay at most a total of $y + wVz^2$ to get away with production. Given the ledger requires the agreement of at least τV validators to validate a transaction, the cheating late producer will pay $\bar{z} = (y + Vz^2)/(\tau V)$ to τV validators. Using $\bar{z} = (y + wVz^2)/(\tau V)$, a

validator rejects the bribe whenever[47]

$$\pi\beta U_V \geq \frac{1}{\tau V}\left(y + Vz^2\right). \tag{50}$$

$$\pi\beta U_V \geq \frac{1}{\tau V}\left[\Phi(x + Vz^1) + Vz^2\right]. \tag{51}$$

Validation threshold.

$$z^1 + z^2 \geq \frac{c_s + c_v/(1-f)}{1-\tau}. \tag{52}$$

Since the payment to validators should be minimized, (11) binds so validators take home

$$z^1 + z^2 = Z(\tau) \equiv \frac{\bar{c}_s + c_v/(1-f)}{1-\tau}$$

Cheapest to deliver.

Given $Z(\tau)$ the planner will choose z^1 and z^2 that minimizes the total cost of delivering the amount $Z(\tau)$. That is the planner will choose z^1 and z^2 to solve

$$\min_{z^1 \geq 0} \Phi(x + Vz^1) + V\left(Z(\tau) - z^1\right)$$

with FOC,

$$\Phi'(x + Vz^1) - \lambda - 1 = 0$$

where λ is the Lagrange multiplier in the minimization problem. Hence, $z^1 = 0$ whenever $\Phi'(x) \geq 1$.

When $z^1 + z^2 = Z$, the participation constraint of validators (47) is always satisfied. Let $R(\tau)$ be the expected rent of validators,

$$\begin{aligned} R(\tau) &\equiv (1-f)\left[Z(\tau) - (\bar{c}_s + c_v)\right] - fc_v \\ &= \frac{\tau(1-f)\bar{c}_s + c_v}{1-\tau} - c_v^1 \end{aligned}$$

[47] We assume validators act in unison: they either all accept the bribe, or they all reject it.

We can set the participation constraint for early producers (44) at equality and replace $y = \Phi\left(x + VZ(\tau)\right)$. Since validators earn a rent, $U_V \geq U$ and (49) is satisfied whenever (48) is. Then the set of IF allocations is characterized by

$$\frac{\beta\alpha}{1-\beta}\left[u(x) - \Phi\left(x + Vz^1\right) - Vz^2\right] \geq \Phi(x + Vz^1) + Vz^2 \tag{53}$$

$$\pi\frac{\beta\alpha}{1-\beta}\left[u(x) - \Phi\left(x + Vz^1\right) - Vz^2 + R(\tau)\right] \geq \frac{1}{\tau V}\left(\Phi(x + Vz^1) + Vz^2\right) \tag{54}$$

References

Abadi, Joseph and Markus Brunnermeier (2018) "Blockchain Economics," NBER Working Papers 25407.

Amoussou-Guenou, Yackolley, Bruno Biais, Maria Potop-Butucaru, and Sara Tucci-Piergiovanni (2019) "Rationals vs Byzantines in Concensus-based Blockchains," Research report, HAL ID: hal-02043331.

Andolfatto, David (2020) "Assessing the Impact of Central Bank Digital Currency on Private Banks," *The Economic Journal*, ueaa073.

Arner, Douglas W., Raphael Auer, and Jon Frost (2020) "Stablecoins: risks, potential and regulation," *Financial Stability Review* 39, 95-123.

Auer, Raphael (2019) "Beyond the doomsday economics of 'proof-of-work' in cryptocurrencies", BIS Working Papers, no. 765, January.

Auer, Raphael and R Boehme (2020): "The technology of retail central bank digital currency", BIS Quarterly Review, March, p. 85-97.

Auer, Raphael, Giulio Cornelli, and Jon Frost (2020) "Rise of the central bank digital currencies: drivers, approaches and technologies", BIS Working Papers, no. 880, August.

Aymanns, Christoph, Mathias Dewatripont, and Tarik Roukny (2020) "Vertically Disintegrated Platforms" SSRN, February.

Bakos, Yannis and Hanna Halaburda (2021) "Trade-offs in Permissioned vs Permissionless Blockchains: Trust and Performance," SSRN, February.

Baudet Mathieu, George Danezis, Alberto Sonnino (2020) "FastPay: High-Performance Byzantine Fault Tolerant Settlement" arXiv:2003.11506v2 [cs.CR].

Biais, Bruno, Christophe Bisière, Matthieu Bouvard, and Catherine Casamatta (2019) "The blockchain folk theorem", *The Review of Financial Studies*, vol. 32, n. 5, May 2019, pp. 1662–1715.

Boar, Codruta and Andreas Wehrli (2021) "Ready, steady, go? - Results of the third BIS survey on central bank digital currency", BIS papers 114.

Bonneau, Joseph (2016) "Why buy when you can rent?" International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg.

Boyd, John and Prescott, Edward (1986) "Financial Intermediary Coalitions," *Journal of Economic Theory*, 38, 211–232.

Brunnermeier, Markus and Dirk Niepelt (2019) "On the Equivalence of Private and Public Money," *Journal of Monetary Economics* 106, 27-41.

Budish, Eric (2018) "The economic limits of bitcoin and the blockchain", NBER Working Papers, no 24717, June.

Buterin, Vitalik (2021) "Why sharding is great: demystifying the technical properties", available at https://vitalik.ca/general/2021/04/07/sharding.html

Cavalcanti, Ricardo and Neil Wallace (1999a), "A Model of Private Bank Note Issue," *Review of Economic Dynamics,* 2, 104–136.

Cavalcanti, Ricardo and Neil Wallace (1999b), "Inside and Outside Money as Alternative Media of Exchange," *Journal of Money, Credit, and Banking*, 31, 443–457.

Calle, George and Daniel Eidan (2020) "Central Bank Digital Currency: an innovation in payments", Whitepaper, R3, April

Carlsson, Hans and Eric E. van Damme (1993) "Global Games and Equilibrium Selection,"

Econometrica 61, 989- 1018.

Chiu, Jonathan and Thorsten Koeppl (2017) "The Economics of Cryptocurrencies - Bitcoin and Beyond," No 1389, Working Papers, Queen's University, Department of Economics.

Chiu, Jonathan and Thorsten Koeppl (2019) "Blockchain-Based Settlement for Asset Trading," *Review of Financial Studies* 32, 1716-1753.

Chiu, Jonathan and Thorsten Koeppl (2020) "Payments and the D(ata) N(etwork) A(ctivities) of BigTech Platforms," Mimeo Queen's University.

Committee for Payments and Markets Infrastructure (2017) "Distributed ledger technology in payment, clearing and settlement - an analytical framework", February

Cong Lin William, Zhiguo He and Jiasun Li (2020) "Decentralized Mining in Centralized Pools," *The Review of Financial Studies*, hhaa040, https://doi.org/10.1093/rfs/hhaa040.

Corda (2020) "Corda OS 4.6 Developer Documentation", https://docs.corda.net/docs/corda-os/4.6/key-concepts-notaries.html, accessed on 11.11.2020.

Deepesh Patel and Emmanuelle Ganne (2019) "Blockchain & DLT in trade: a reality check," World Trade Organisation, November.

Diamond, Doug (1984), "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51, 393–414.

Eyal I., Sirer E.G. (2014) "Majority Is Not Enough: Bitcoin Mining Is Vulnerable," in: Christin N., Safavi-Naini R. (eds) Financial Cryptography and Data Security. FC 2014. Lecture Notes in Computer Science, vol 8437. Springer, Berlin, Heidelberg

Fernandez-Villaverde, Jesus, Daniel Sanches, Linda Schilling, and Harald Uhlig (2020) "Central Bank Digital Currency: Central Banking For All?," NBER Working Papers 26753.

Froewis, Michael and Rainer Boehme (2017) "In code we trust? Measuring the control flow immutability of all smart contracts deployed in Ethereum," in J Garcia-Alfaro, G. Navarro-Arribas, H Hartenstein, and J Herrera-Joancomarti (eds), Data privacy management, cryptocurrency and blockchain technology, Springer, pp. 357-72.

Frost, Jon, Hyun-Song Shin and Peter Wierts (2020) "An early stablecoin? The Bank of Amsterdam and the governance of money," BIS Working Papers, no. 902.

Garatt, Rodney and Maarten R.C. van Oordt (2020) "Why Fixed Costs Matter for Proof-of-Work Based Cryptocurrencies," Bank of Canada Working Paper 2020-27.

Giulia Fanti, Leonid Kogan, and Pramod Viswanath (2019) "Economics of Proof-of-Stake Payment Systems," Mimeo, University of Illinois.

Gersbarch, Hans, Akaki Mamageishvili, and Oriol Tejada (2020) "Appointed Learning for the Common Good: Optimal Committee Size and Efficient Rewards," CEPR Working Paper DP15311.

Gervais, Arthur, Ghassan O Karame, Karl Wüst, Vasileios Glykantzis, Hubert Ritzdorf, Srdjan Capkun (2016) "On the security and performance of proof of work blockchains," CCS, Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, October

Goldstein, Itay and Ady Pauszner (2005) "Demand–Deposit Contracts and the Probability of Bank Runs," *Journal of Finance* 60, 1293-1327.

Graeber, David Debt: the first 5,000 years, Brooklyn: Melville House, 2011,

Gu Chao, Fabrizio Mattesini, Cyril Monnet, and Randall Wright (2013) "Banking: A New Monetarist Approach," *Review of Economic Studies* 80, 636-662.

Green, Edward, and Robert Porter (1984) "Noncooperative Collusion under Imperfect Price Information," *Econometrica* 52, 87-100.

Grossman, Sanford J., and Joseph E. Stiglitz (1976) "Information and competitive price systems," *The American Economic Review* 66, 246-253.

Halaburda, Hanna, Zhiguo He, and Jiasun Li (2021) "An Economic Model of Consensus on Distributed Ledgers," mimeo University of Chicago.

Huang, Angela (2019) "On the Number and Size of Banks: Efficiency and Equilibrium," Mimeo, National University of Singapore.

Huberman, Gur, Jacob Leshno, and Ciamac C. Moallemi (2021) "Monopoly without a monopolist: An economic analysis of the bitcoin payment system," *Review of Economic Studies*, forthcoming.

Judmayer, Aljosha, Nicholas Stifter, Philipp Schindler, and Edgar Weippl (2018) "Pitchforks in Cryptocurrencies," In Data Privacy Management, Cryptocurrencies and Blockchain Technology, pp. 197-206. Springer, Cham, 2018.

Kandori, Michihiro (2001) "Introduction to Repeated Games with Private Monitoring," *Journal of Economic Theory* 102, 1–15.

Keister, Todd and Cyril Monnet (2020) "Central Bank Digital Currency: Stability and Information," Mimeo, Rutgers University.

Kocherlakota, Narayana (1998) "Money Is Memory," *Journal of Economic Theory* 81, 232-251.

Kocherlakota, Narayana and Wallace, Neil (1998) "Incomplete Record-Keeping and Optimal Payment Arrangements," *Journal of Economic Theory* 81(2), 272-289.

Kroll, Joshua A., Ian C. Davey, and Edward W. Felten (2013) "The economics of Bitcoin mining, or Bitcoin in the presence of adversaries," Proceedings of WEIS. Vol. 2013.

Lagos, Ricardo and Randall Wright (2005) "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy* 113, 463-484.

Lagos, Ricardo, Guillaume Rocheteau, and Randall Wright (2017) "Liquidity: A New Monetarist Perspective," *Journal of Economic Literature* 55, 371-440.

Leland, H. E. and Pyle, D. H. (1977) "Informational Asymmetries, Financial Structure and Financial Intermediation," *Journal of Finance*, 32, 371–387.

Leshno, Jacob, and Philipp Strack (2020) "Bitcoin: An Axiomatic Approach and an Impossibility Theorem," *American Economic Review: Insights* 2, 269-86.

Li, Yiting, and Chien-Chiang Wang (2019) "Cryptocurrency, Imperfect Information, and Fraud," Munich Personal RePEc Archive (MPRA) Paper No. 94309.

Luu, Loi, Duc-Hiep Chu, Hrishi Olickel, Prateek Saxena and Aquinas Hobor (2016) "Making smart contracts smarter," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 254-69.

Mitchell-Innes, A "The Credit Theory of Money", The Banking Law Journal, vol 31, December 1914, pp 151–68.

Monnet, Cyril (2006) "Private vs Public Money," *International Economic Review* 47(3), 951–960.

Monnet, Cyril and Erwan Quintin (forthcoming) "Optimal Financial Exclusion," *American Economic Journal: Microeconomics.*

Morris, Stephen and Shin, Hyun Song (1998) "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks," *American Economic Review* 88, 587-97.

Morris, Stephen, and Hyun Song Shin (2002) "Measuring Strategic Uncertainty," manuscript London School of Economics.

Morris, Stephen, and Hyun Song Shin (2003) "Global Games: Theory and Applications," in Advances in Economics and Econometrics: Proceedings of the Eighth World Congress of the Econometric Society, vol. 1, edited by Matthias Dewatripont, Hansen, Lars P. and Stephen J. Turnovsky, chap. 3, pp. 56–114. Cambridge University Press.

Pagnotta, Emiliano (2021) "Decentralizing Money: bitcoin prices and blockchain security" *The Review of Financial Studies* (forthcoming).

Prat, Julien, and Benjamin Walter (2018) "An Equilibrium Model of the Market for Bitcoin Mining," CESifo Group Munich No. 6865.

Rahman, David (2012) "But Who Will Monitor the Monitor?" *American Economic Review* 102, 2767-97.

Rochet, Jean-Charles and Xavier Vives (2004) "Coordination Failures and The Lender of Last Resort," *Journal of the European Economic Association* 2, 1116–1147.

Rocheteau, Guillaume and Ed Nosal (2017) "Money, Payments, and Liquidity," MIT Press.

Saleh, Fahad (2021) "Blockchain Without Waste: Proof-of-Stake" *Review of Financial Studies*, Forthcoming.

Shanaev, Savva Arina Shuraeva, Mikhail Vasenin, Maksim Kuznetsov (2020) "Cryptocurrency Value and 51% Attacks: Evidence from Event Studies" *The Journal of Alternative Investments*, forthcoming.

Schilling, Linda and Harald Uhlig (2019) "Some Simple Bitcoin Economics," *Journal of Monetary Economics* 106, 16-26.

Sveriges Riksbank (2020) "The Riksbank's e-krona pilot", February.

Teutsch, Jason, Sanjay Jain, and Prateek Saxena (2016) "When cryptocurrencies mine their own business," International Conference on Financial Cryptography and Data Security. Springer, Berlin, Heidelberg.

Tirole, Jean (2020) "Public and Private Spheres and the Authentic Self.," Mimeo, Toulouse School of Economics.

Townsend, Robert M. (2020) "Distributed Ledgers: Design and Regulation of Financial Infrastructure and Payment Systems", MIT Press (2020), ISBN electronic: 9780262361194.

Wallace, Neil (2005) "From Private Banking To Central Banking: Ingredients of a Welfare Analysis," *International Economic Review*, Vol. 46, No. 2, 619–632.

Williamson, Stephen (1986) "Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing," *Journal of Monetary Economics*, 18, 159–179.

Williamson, Stephen (1987) "Financial Intermediation, Business Failures, and Real Business Cycles," *Journal of Political Economy*, 95, 1196–1216.

Williamson, Stephen and Randall Wright (2011) "New Monetarist Economics: Models," in Handbook of Monetary Economics, vol. 3A, B. Friedman and M. Woodford, eds, Elsevier, 25-96.

Xu, Xiwei, Cesare Pautasso, Liming Zhu, Vincent Gramoli, Alexander Ponomarev, An Binh Tran, and Shiping Chen (2016) "The blockchain as a software connector," in Proceedings of

the 13th Working IEEE/IFIP Conference on Software Architecture (WICSA), pp. 182-191.