

Compressed to 0: The Silent Strings of Proof of Personhood¹

Puja Ohlhaber², Mikhail Nikulin³, Paula Berman⁴

Abstract

Experiments in Proof of Personhood—where each person has a single, unique identity—have increasingly been touted as a mechanism for tracing information provenance, distributing Universal Basic Income, and facilitating democratic governance over systems of artificial intelligence. This paper chronicles Idena’s experiment in Proof of Personhood from launch in August 2019 to a crisis in May 2022, which prompted a pivot towards a novel experiment in sublinear identity staking. We show how despite verifying humans, hidden pools rapidly emerged—some cooperative, but most controlled by “puppeteers” who, at best, remunerated participants for periodically proving their uniqueness in exchange for access to their secret keys and controlling their accounts. Instead of fostering an egalitarian network of unique identities, the protocol fractured into hidden subnetworks vying for control over an economic pie with economies of scale trending towards oligopoly, undermining the protocol’s security and ambitions for democratic governance (*one-person, one-vote*) and UBI rewards (*one-person, one-reward*). By giving humans economic incentives to periodically differentiate themselves *from* bots—even as low as \$2 to \$14 every few weeks—the protocol gave more informed, resourceful humans financial incentives to puppeteer less informed humans *like* bots. Notably, by May 2022, 23 entities constituting less than 0.6% of the network’s distinct entities controlled at least ~40% of accounts and the distribution of almost half (~48%) the network rewards. More striking, 3 entities controlled ~19% accounts and ~24% rewards. An off-chain system trending towards oligopoly subsumed an on-chain egalitarian system, quietly and opaquely. Achieving *de jure* sybil-resistance (filtering humans from bots) revealed a deeper challenge of *de-facto* sybil resistance (filtering humans acting *like* bots), which could not coherently or computationally be disentangled from the problem of collusion-resistance.

¹ We thank Glen Weyl, Joel Miller, Leon Erichsen, Andrew Miller, Alex Tabarrok, Nikete della Penna, Stephane Gosselin, Phil Daian, Christopher Goes, Allison Stanger, Henry Farrell, Vitalik Buterin, Sebastian Burgel, Matt Prewitt, Jack Henderson, Jasleen Malvai, Sophia Cossar, Barnabe Monnot, Omid Malekan, Awa Sun Yin and Zaki Manian for their conversations, review and thoughtful comments. All errors are our own. Please send comments and questions to compressed-to-0-authors@idena.io.

² Puja Ohlhaber is an independent researcher. She neither owns IDNA tokens nor was compensated financially for this paper. This paper should not be construed as legal advice.

³ Mikhail Nikulin is a Founder of Idena. He holds IDNA tokens.

⁴ Paula Berman is COO of RadicalxChange. She holds IDNA tokens.

TABLE OF CONTENTS

I.	INTRODUCTION	3
II.	OVERVIEW OF IDENA PROTOCOL	5
	<i>A. Filtering Humans from Bots</i>	5
	<i>B. Synchronous Participation & Periodic Re-authentication</i>	6
	<i>C. Stymying Private Key Sales with Identity Staking & Identity Slashing</i>	7
	<i>D. One-person, One-vote, One-reward....and One-node</i>	8
III.	PROTOCOL LAUNCH	9
	<i>A. Suspicious Transaction Patterns</i>	9
	<i>B. Muddied Waters: Puppeteering or Cooperation?</i>	11
	<i>C. Waking Up Puppets?</i>	14
IV.	PIVOT TOWARDS DELEGATION	15
	<i>A. Proliferation of Pools</i>	16
	<i>B. 3rd Party Key Access in Top Pools</i>	18
	<i>C. An Emergent Puppeteering Oligopoly in Top Pools</i>	19
	<i>D. From 3rd Party Key Access to Puppeteering</i>	21
	<i>E. Panic and Threat of Collapse</i>	25
V.	DISCUSSION	27
	<i>A. A Failure in Egalitarianism (one-person, one-reward, one-vote)</i>	27
	<i>B. Democratic Governance in a Network</i>	29
	<i>C. From Sybil-Resistance to Collusion-Resistance</i>	32
	<i>D. Dark DAOs and Voting Security</i>	35
VI.	CONCLUSION	36
	APPENDIX A (Methodology)	38
	APPENDIX B (Pool Analysis)	39

I. INTRODUCTION

As generative foundation models achieve faster-than-human proficiency in mimicking essential identity markers—such as writing style, vocal subtlety, and visual representation—Proof of Personhood protocols have been touted as a way to establish provenance and authenticity in the wake of deep-fakes, as well as facilitate democratic governance (*one-person, one vote*) and distribute a universal basic income (*one-person, one-reward*). Experiments in Proof of Personhood aim for “sybil-resistance,” where verified unique humans control corresponding unique accounts 1:1.

Idena is an open-source, Proof of Personhood blockchain that launched with the egalitarian goals of *one-person, one-vote, one-reward* in August 2019, before pivoting in May 2022 towards a novel experiment in “sublinear identity staking”—an intermediate between Proof of Personhood and Proof of Stake. This paper is an empirical study about Idena’s first experiment in Proof of Personhood, saving the pivot for our next essay. We show how despite filtering bots and verifying humans, pools rapidly emerged—some cooperative but most controlled by “puppeteers” who, at best, paid participants to periodically prove their uniqueness in exchange for access to their private keys and controlling their accounts. By giving humans economic incentives to differentiate themselves *from* bots—even as little as \$2 to \$14 UBI for 30 minutes of work every 1 to 3 weeks—Idena gave more informed, resourceful humans economic incentives to control less informed humans *like* bots and extract greater rewards. Notably, by May 2022, 23 entities constituting less than 0.6% of the network’s distinct entities controlled at least ~40% of accounts and the distribution of almost half (~48%) the network rewards. More striking, 3 entities controlled ~19% accounts and ~24% rewards, with a trendline towards oligopoly. Instead of an egalitarian network of unique identities, on-chain, *asocial* personhood credentials collapsed into off-chain *social* arrangements that obfuscated power at best, or reinforced it at worst.

Proof of Personhood protocols differ widely in their technological and political architecture: how they validate unique identities, achieve consensus (if there is a blockchain), govern the protocol, protect privacy, or conceptualize “personhood” philosophically. While *one-person, one-vote* has been a motivating use-case for Proof of Personhood, *one-person, one-reward* UBI is a more recent innovation accelerated by blockchains, which enable global distribution of transferable tokens. When referring to “Proof of Personhood,” this paper narrowly refers to the subset of blockchain protocols that seek to offer a more egalitarian alternative to Proof of Work and Proof of Stake; rather than confer votes and economic rewards to participants who *already* can afford to buy compute or stake, these protocols aim to verify unique humans controlling corresponding unique accounts, thereby enabling democratic processes and a more egalitarian rewards distribution (*one-person, one-vote, one-reward*).

Why Idena?

This paper is a case study in the unintended and unequal consequences of wedding egalitarian goals with economic and political incentives to verify biological uniqueness. We eschew a comparative overview of competing Proof of Personhood protocols (e.g., WorldCoin, Proof of Humanity, Humanode) in favor of an empirical audit of Idena. Through this deep-dive, we aim to set a standard for analysis and disclosure for global identity protocols, especially those that tout similar egalitarian ambitions. Idena offers a benchmark, having technically succeeded in

filtering bots, validating humans, and undermining account trading for the study’s period—a problem which beleaguers several protocols today. Instead, the protocol had to grapple with a deeper *social* problem: an ongoing economic *relationship* between a verified human participant and operator—a principal and agent, or puppet and puppeteer, depending on the depth of asymmetry and who captured the benefit of the bargain.

Roadmap

This paper starts technically and ends pragmatically in a discussion about power: we first explain how Idena verifies unique humans (Section II) and then chronicle the protocol’s outcomes, from launch in August 2019 (Section III) to delegation in March 2021 (Section IV), ending in a puppeteering crisis a year later. We then pivot to a discussion (Section V), where we match the egalitarian ambitions of democratic governance (*one-person, one-vote*) and UBI rewards (*one-person, one-reward*) against the reality of a protocol fractured into off-chain subnetworks vying for control over an economic pie and, by extension, participants’ time, attention and their accounts. We then extrapolate to the future; as humans further integrate biologically with information technology, the distinction between filtering humans from bots (*de jure* sybil-resistance) and filtering humans *acting like* programmable bots (*de facto* sybil-resistance) will blur more, if not collapse, underscoring a more foundational challenge than establishing biological uniqueness: establishing the *informational* uniqueness of participants—or the extent to which they cluster with the *same interests and biases*, leading to tacit collusion that risks monopoly and majoritarian capture. Rather than presume all participants to be informationally the *same* in *one* global identity game, we offer an alternative starting point: acknowledging informational *differences* which arise from talking and trading in markets and politics—our social ties—for *many* identity games as varied as the diversity of human associations.

Contributions

We offer unique contributions to fields at the intersection of decentralized identity, information theory, computational democratic governance, and voting security:

- **Empirical Study:** We present a first empirical analysis of a Proof of Personhood protocol, checking the egalitarian ambitions against the outcomes of a network trending towards oligopoly.
- **Puppeteering:** We offer definitions of “puppeteering,” distinguishing between “strong” and “semi-strong,” and contrast puppeteering against accountable principal-agent relationships.
- **Sybil-Resistance:** In the context of Proof of Personhood, we refine the concept of sybil-resistance to acknowledge two forms: *de jure* sybil-resistance (filtering humans from bots where each humans controls a corresponding account 1:1) and *de facto* sybil-resistance (filtering humans *acting like* programmable bots).
- **Collusion-Resistance:** We bridge “*de facto* sybil resistance” and “collusion resistance” as equivalent computational, or informational challenges.
- **Computational Democratic Governance:** We reframe the goals of democratic governance away from simple *one-person, one-vote* towards checking faction to surface broader public goods, which in turn requires acknowledging the informational uniqueness (or clusters) of participants.

- **Voting Security:** We highlight how undermining vote-buying *on-chain* doesn't solve for *off-chain* vote-buying in "meatspace," but may encourage it as a low-cost alternative.

II. OVERVIEW OF IDENA PROTOCOL

A. Filtering Humans from Bots

The Idena Protocol is an open-source Proof of Personhood blockchain.⁵ To gain the status of a unique identity or "human" in Idena, users **synchronously** participate in a series of 4 consecutive validation ceremonies, progressively increasing their status to "human."⁶ Validation ceremonies are scheduled in advance (15:00 UTC⁷) every 1 to 3 weeks ("epoch") based on the network size; the larger the network, the longer the epoch.⁸ In each validation ceremony, participants solve 6 FLIP ("**Filter for Live Intelligent People**") puzzles within 2 minutes. A FLIP is a cognitive test, consisting of a series of photos generated by other participants that convey an intelligible human story in one configuration and are meaningless in another random configuration.⁹ FLIPS aim to be a reverse Turing test, verifying the presence of a human rather than a bot.¹⁰ For the period under analysis (August 2019 to May 2022), Idena's FLIPS were successfully AI-resistant, where bots did not succeed in generating fake accounts.¹¹ Other Proof of Personhood protocols prove uniqueness in other ways, relying on centralized identity verification

⁵ See Idena, "Whitepaper," Idena Docs, accessed December 2, 2023, <https://docs.idena.io/docs/wp/technology>. For the protocol's github, see Idena Network, "Idena," GitHub, accessed December 2, 2023, <https://github.com/idena-network>. For a survey of Proof of Personhood protocols, see Divya Siddarth, Sergey Ivliev, Santiago Siri, and Paula Berman, "Who Watches the Watchmen? A Review of Subjective Approaches for Sybil-Resistance in Proof of Personhood Protocols," *Frontiers in Blockchain vol. 3* (2020).

⁶ After receiving an invitation and completing one validation ceremony, a "candidate" account may upgrade to "newbie." After two ceremonies, the candidate may upgrade to "verified," and then after a fourth ceremony upgrade to "human." Throughout this paper we broadly refer to "verified humans," "verified accounts," "verified unique humans," or "unique identity status." Applied to the context of Idena, we are referring to participants who either achieve the "verified" or "human" status. Idena, "Validation Session," Idena FAQ, accessed December 2, 2023, <https://www.idena.io/faq#faq-validation-6>.

⁷ The scheduled time for validation ceremonies was changed in 2023 from 13:30 UTC to 15:00 UTC. Idena, "IIP-2: Change the time of validation ceremony," Idena Docs, accessed December 2, 2023, <https://docs.idena.io/docs/iip/iip-2>.

⁸ For the epoch formula, see Idena, "Validation Session," Idena FAQ, accessed December 2, 2023, <https://www.idena.io/faq#faq-validation-1>.

⁹ Participants create FLIPs in response to randomly generated two-word prompts, a subset of which are then randomly distributed to the network to solve in two minutes. In a subsequent "long phase," nodes form consensus on the correct answer for each flip, where strong and weak consensus flips (with 75%+ and 66%+ agreement, respectively, on what configuration was the "correct" answer) generates different rewards for the participants who answered consistent with the consensus opinion. See Idena, "Flip Challenge" Idena FAQ, accessed December 2, 2023, <https://www.idena.io/faq#faq-challenge-3>.

¹⁰ For an early formalism of "AI-hardness," see Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford, "CAPTCHA: Using Hard AI Problems for Security," *Advances in Cryptology — EUROCRYPT* (2003): 294-311, https://doi.org/10.1007/3-540-39200-9_18. See also Roman Yamolskiy, "AI-Complete, AI-Hard, or AI-Easy: Classification of Problems in Artificial Intelligence," Paper presented at the 23rd Midwest Artificial Intelligence and Cognitive Science Conference, Cincinnati, OH, USA, 2012.

¹¹ FLIP tests are intended to be "AI-hard" or "AI-resistant." Frontier AI models may challenge this boundary in the future. See Idena Network, "AI-Resistant CAPTCHAs: Are They Really Possible?" *Medium*, May 8, 2019, <https://medium.com/idena/ai-resistant-captchas-are-they-really-possible-760ac5065bae>. Idena has an ongoing bug bounty to develop an open source AI instrument for solving FLIPs. Idena, "FLIP Challenge," Idena Docs, accessed December 2, 2023, <https://docs.idena.io/docs/wp/flip-challenge>.

(vulnerable to censorship), peer-to-peer validation (vulnerable to AI deep-fakes), or biometric information (vulnerable to leaks by surveillance).¹²



Figure 1: Sample FLIP Test

B. Synchronous Participation & Periodic Re-authentication

Whereas FLIP tests aim to filter bot accounts, simultaneous validation ceremonies across all participants prevent participants from generating fake accounts, or “sybils,” because one person can only be in one place at one time.¹³ Given cognitive diversity, however, outliers may complete more than one validation ceremony within the allotted 2 minutes, but are nonetheless capped at 2 (rare) or 3 (very rare).¹⁴ Required periodic re-authentication every epoch (every 1 to 3 weeks) also increases the cognitive cost and lowers the chance of one person maintaining multiple accounts. Given this hard ceiling to multiple accounts—a ceiling that exists even if anecdotal estimates are

¹² Biometrics are a digital representation of a unique physical characteristic, such as fingerprints, facial patterns, and irises. Most biometrics are static—with facial recognition being a notable exception that can change with age, injury, or surgery. Static representations present a security challenge; if an attacker acquires this data, they can misuse it indefinitely because biometrics are mostly immutable (i.e. you can’t change your fingerprint or retinal patterns easily). Thus, biometrics as credentials are problematic because they are leaky, vulnerable to unauthorized capture with surveillance technology, a risk that only increases with better sensor technology. See Bruce Schneier, “Tigers Use Scent, Birds Use Calls — Biometrics Are Just Animal Instinct,” *The Guardian*, January 8, 2009, <https://www.theguardian.com/technology/2009/jan/08/identity-fraud-security-biometrics-schneier-id>. Worldcoin’s World ID attempts to mitigate leak risk by relying on cryptography to store only an “iris code” (not the scan itself), though participants may opt in for data custody of their iris scan. Notably, to loot wallet funds, an attacker has to acquire a participant’s private key, which is not linked (and unrelated) to their iris data. An attacker who possesses a participant’s iris data (because of leaks), however, may attempt to renew credentials (reissuing a new WorldID and revoking the old) presuming a participant may re-issue their World ID by returning to an orb and the attacker has advanced spoofing technology that circumvents the Orb’s security measures (e.g., Lidar, heat sensors). See n. 65 & n. 66. According to the company’s White Paper, “even though significant effort has been spent on raising the security bar of the Orb, it is expected that the Orb may get spoofed or compromised by determined actors.” Worldcoin, “White Paper: Limitations,” accessed December 2, 2023, <https://whitepaper.worldcoin.org/limitations>. For a discussion of how the orb works, see Worldcoin, “Opening the Orb: A look inside Worldcoin’s biometric imaging device,” *Worldcoin Blog*, January 27, 2023, accessed December 2, 2023, <https://worldcoin.org/blog/engineering/opening-orb-look-inside-worldcoin-biometric-imaging-device>.

¹³ For a discussion of attack vectors on Idena, see Jordi Subirà-Nieto, “Security of Proof-of-Personhood: Idena,” supervised by Bryan Ford, Louis-Henri Merino, and Haoqian Zhang, *Decentralized Distributed Systems Laboratory—EPFL*, June 11, 2021, https://www.epfl.ch/labs/dedis/wp-content/uploads/2021/07/report-2021-1-jordi-iden_a_report.pdf.

¹⁴ Our estimates of 2-3 maximum accounts is based on anecdotal evidence. We welcome a study in the cognitive distribution of accounts.

challenged in the future— we characterize Idena’s sybil-resistance as “semi-strong,” filtering bots from humans and verifying mostly “unique” humans, except for cognitive outliers.¹⁵

C. *Stymying Private Key Sales with Identity Staking & Identity Slashing*

Validated accounts earn rewards by participating in periodic ceremonies every epoch (**validation ceremony rewards**)¹⁶ and by running a node and producing blocks as a validator (**mining rewards**).¹⁷ Rewards are IDNA tokens, split evenly between the validation ceremony and mining rewards pies. Whether earned through validation or mining, **20% of all earned rewards are automatically locked as “identity stake” while the remaining 80% are unencumbered in the account’s wallet as “transferable rewards.”** Importantly, when a participant unlocks and withdraws their locked identity stake, their status as a unique identity in Idena is lost (or “killed”), along with the privilege of being a validator that can earn mining awards.¹⁸ Thus, locked identity stake undermines credible sales of private keys because sellers always have a financial incentive to withdraw their locked stake simultaneously or immediately *after* a sale.¹⁹ The older the account, the larger the identity stake, and the more powerful the financial incentive to not buy or sell an identity.

Periodic account re-authentication also discourages account trading by imposing ongoing cognitive costs to the buyer who periodically must re-validate the account. Missing a series of validation ceremonies (or failing them), results in a progressive degradation of account status over several stages until the account is “killed”—or

¹⁵ Another characterization for “semi-strong sybil-resistance” is “semi-unique” verified identities. To the extent that a minority of participants have 2 accounts (or very rarely 3), Idena has verified semi-unique biological humans. Nonetheless, semi-strong sybil-resistance might be desirable to eschew single identities that become a target for theft, coercion and cancellation. See Vitalik Buterin, “Progress,” *Vitalik Buterin’s Blog*, <https://vitalik.eth.limo/general/2019/11/22/progress.html> (noting “it should be much harder to get two identities than one, but making it impossible to get multiple identities is both impossible and potentially harmful even if we do succeed.”)

¹⁶ As an account increases in status from “newbie” to “verified” to “human,” it enjoys increasing privileges and opportunities to earn rewards. “Newbies” may engage in mining and earn validation ceremony rewards, but lack voting power, which is conferred only to “verified” and “human” accounts. Validation ceremony rewards are distributed across several activities: successfully completing a ceremony, submitting qualifying FLIPs, and invitations that convert to candidate accounts. Thus, while mining rewards are split evenly across all validated accounts (whether “newbie” or “human”), “human” accounts may earn more validation rewards (have more FLIPs and invitations). See “Economy,” Idena FAQ, accessed December 2, 2023, <https://www.idena.io/faq#faq-economy-6>.

¹⁷ 100 accounts are randomly called on to vote on proposed blocks as a committee, whereby blocks only get validated with a supermajority vote from a committee. Only accounts which have mining nodes are selected to a committee. Failing to vote as a node, or staying silent, costs an inactivity penalty of 8 hours of mining without accruing mining rewards. See Idena, “Staking,” Idena Website, accessed December 2, 2023, <https://www.idena.io/staking>.

¹⁸ Replenishing withdrawn identity stake does not reinstate “human” status. Instead, reviving a lost account requires participation in at least 3 consecutive validation ceremonies. See “Economy,” Idena FAQ.

¹⁹ For a discussion of *partial* key encumbrances using trusted execution environments, see the section “Dark DAOs and Voting Security.” Our observation was that off-chain purchases of people’s time (and accounts) was the preferred (and perhaps easier) alternative than partial key encumbrances for vote-buying.

the identity “**slashed**.”²⁰ When slashed, 100% of the identity stake is burned (though transferable awards remain available) and the account loses its unique identity status, along with eligibility to earn rewards.

D. *One-person, One-vote, One-reward....and One-node*

Whereas Proof of Work and Proof of Stake require control over compute and stake respectively to participate in blockchain consensus, Idena requires a verified human account running their own mining node.²¹ The rationale behind accounts running their own nodes was the network would gain an excessive, redundant number of nodes for greater throughput (increasing speed)²² along with more diverse node operators less likely to collude in a 51% attack (increasing security).²³ Politically and economically, each verified account running their own node was also necessary to achieve egalitarianism,²⁴ enabling UBI rewards (*one-node, one-mining reward*) and egalitarian governance (*one-node, one-vote*), where a supermajority of nodes vote on-chain by upgrading their software (or “hard-fork”) to instantiate protocol changes. Notably, if each account failed to run their own node, governance and mining rewards from a fixed economic pie would concentrate in node operators, rather than distributing equally across human accounts, undermining egalitarianism.

²⁰ Depending on whether the account misses a validation ceremony or fails it, an account moves through different statuses: from “human” or “verified” to “suspended” to “zombie” to eventually losing the account, or being killed. See “Economy,” Idena FAQ.

²¹ In Idena, nodes are randomly selected to propose blocks and vote on proposed blocks to earn block rewards and risk identity slashing (losing their unique identity status and identity stake) or penalties (e.g., loss of mining rewards) for inactivity or protocol deviations, like proposing two blocks instead of one. See “Economy,” Idena FAQ. In Proof of Work, “miners” compete with compute to determine the next block by way of solving cryptographic puzzles. In Proof of Stake, “validators” are selected based on their staked collateral to propose or attest to new blocks and risking the loss of their stake for dishonest or strategic behavior. For a discussion of how Proof of Personhood seeks to be a different alternative than both Proof of Work and Proof of Stake, see Maria Borge, Eleftherios Kokoris-Kogias, Philipp Jovanovic, Linus Gasser, Nicolas Gailly, and Bryan Ford, “Proof-of-Personhood: Redemocratizing Permissionless Cryptocurrencies,” *2017 IEEE European Symposium on Security and Privacy Workshops* (2017): p. 23-26, <https://doi.org/10.1109/EuroSPW.2017.46>. See also Satoshi Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System,” 2008, <https://bitcoin.org/bitcoin.pdf>; Vitalik Buterin, “Why Proof of Stake,” *Vitalik Buterin’s Blog*, November 6, 2022, <https://vitalik.eth.limo/general/2020/11/06/pos2020.html>.

²² Node-based governance and economic awards also dovetailed protocol’s architectural goals to increase the blockchain throughput by leveraging sharding of the network with a redundantly high number of nodes. Idena aimed to address the blockchain scalability trilemma with a redundantly high number of independent nodes split into shards. See Vitalik Buterin, “Sharding,” *Vitalik Buterin’s Blog*, April 7, 2021, <https://vitalik.eth.limo/general/2021/04/07/sharding.html>. See also Idena, “Technology: Scalability,” Idena Docs, accessed December 2, 2023, <https://docs.idena.io/docs/wp/technology#scalability>.

²³ A 51% attack refers to a scenario in which a single entity or colluding group controls the majority of the network’s computational power in Proof of Work or a majority of the network’s stake in Proof of Stake. This control can be leveraged to manipulate the blockchain’s consensus and commit fraudulent transactions, double-spend coins, prevent certain transactions from being confirmed, or even rewrite parts of the blockchain’s history. In the context of Idena’s Proof of Personhood (PoP), where the premise is one unique human having a corresponding unique account and node, a 51% attack happens when one entity or colluding group controls a majority of unique nodes, by way of controlling/influencing unique accounts and their nodes (or, after delegation, a combination of solo account nodes and pool nodes).

²⁴ Savvy readers will acknowledge, however, there already were inequalities in rewards based on the different statuses in Idena (e.g., “verified” or “human”) and what they could earn in validation ceremony rewards. See n. 16.

PROTOCOL LAUNCH (August 2019)

A. Suspicious Transaction Patterns

The Idena protocol launched in August 2019. Anyone could join the protocol so long as they participated in the synchronous validations ceremonies (solving FLIP tests), held periodically at 15:00 UTC. Accounts steadily increased, reaching over 6,000 accounts within the first 18 months, generally among blockchain enthusiasts who learned about the protocol by way of word-of-mouth, articles, and blog posts. But as the network grew, on-chain data began to show a curious and increasing phenomena; some wallets were sending their unlocked, transferable rewards to the same address. Moreover, these rewards never returned on-chain back to participants, but instead proceeded to exchanges.

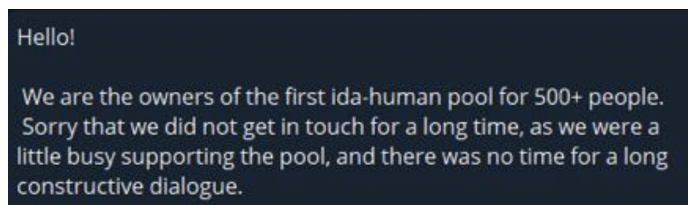
Transaction	From	To	Amount	Timestamp	Type
0xfa636...	0x89e32A...	0x989DAF...	71.01842	12/16/2020 19:44:41	Send
0xa9a8...	0x87F2e0...	0x989DAF...	73.099762	12/16/2020 19:44:41	Send
0x9912...	0x74f783F...	0x989DAF...	66.748138	12/16/2020 19:44:41	Send
0xcd53...	0x70C9D4...	0x989DAF...	66.577984	12/16/2020 19:44:41	Send
0x4e86...	0x7FB8CA...	0x989DAF...	77.373542	12/16/2020 19:44:41	Send
0x0a83...	0x791604...	0x989DAF...	69.21582	12/16/2020 19:44:41	Send
0x2b00...	0x8E3046...	0x989DAF...	67.30252	12/16/2020 19:44:41	Send
0x8674...	0x6Dd53a...	0x989DAF...	74.495269	12/16/2020 19:44:41	Send
0x7959...	0x71daF7c...	0x989DAF...	20.551793	12/16/2020 19:44:41	Send
0xf8898...	0x6B4EFD...	0x989DAF...	68.097209	12/16/2020 19:44:41	Send
0x6ec3...	0x85f398...	0x989DAF...	70.538272	12/16/2020 19:44:41	Send
0x4827...	0x6e2Feb...	0x989DAF...	71.30269	12/16/2020 19:44:41	Send
0x90e5...	0x6cB154...	0x989DAF...	65.422287	12/16/2020 19:44:41	Send
0xa2f1c...	0x849EEC...	0x989DAF...	70.063577	12/16/2020 19:44:41	Send
0xb772...	0x7dBF34...	0x989DAF...	69.857895	12/16/2020 19:44:41	Send
0xb189...	0x7A78a19...	0x989DAF...	72.340046	12/16/2020 19:44:41	Send
0xe35ac...	0x727702...	0x989DAF...	75.501131	12/16/2020 19:44:41	Send
0x2e8a...	0x8D2Bb0...	0x989DAF...	71.741447	12/16/2020 19:44:41	Send
0x95bc...	0x78D7b4...	0x989DAF...	72.295027	12/16/2020 19:44:41	Send

Figure 2: Snapshot of transactions
from accounts to a Russian account in December 2020 ([see blockchain explorer](#))²⁵

Blocks of one-way transfers at the *same* time to the *same* wallet implied automation, which would require 3rd party access to a participant's private keys. Either participants had unwittingly ceded their keys to a 3rd party, or never had them. The latter suspicion was confirmed when forked versions of Idena's software surfaced, where participants

²⁵ The last transaction of this transferable wallet (0x989daf4e639ea7438029fdbd3b04c79553f7164c) was a transfer to another Russian transferable wallet 0xDDDD06adBF37d5F7997E61e410d567DDC56AE79E. The 0x989 account was later delegated to two known Russian pools 0xDDDDaDDB856901ac3e2251b8234EfeaB2188b22A and 0xDDDDcFdCC512FacD27038BA958742E81e2982cB, before it was finally terminated. See Appendix B for further discussion.

were accessing validation ceremonies through software clients that masked private keys controlled by a 3rd party.²⁶ By December 2022, one participant admitted to running a “human pool,” finding over 500 participants willing to get paid in local currency in exchange for performing validation ceremonies, or finding others to do the same.



```
Hello!  
  
We are the owners of the first ida-human pool for 500+ people.  
Sorry that we did not get in touch for a long time, as we were a  
little busy supporting the pool, and there was no time for a long  
constructive dialogue.
```

Figure 3: Private message from a Russian puppeteer to Idena Team (Dec 2, 2020)

These transaction patterns, messages, and forked versions of software were all suggestive of “human farms” or “**puppeteering**,” where high-information operators, at best, pay low-information participants —“puppets”—to periodically perform validation ceremonies and verify their uniqueness in exchange for access to their private keys and controlling their accounts. Puppets were either unaware of their account’s private keys (“strong puppets”), or “knew” their private keys but were unaware of their significance within the protocol (“semi-strong puppets”).²⁷

Puppeteering was not a traditional (or *de jure*) sybil attack, where one person, a sybil, fakes many accounts, typically by way of bots.²⁸ Validation ceremonies had *succeeded* in filtering out bots and authenticating flesh-and-blood humans. Instead, puppeteering was a *de facto* sybil attack of humans *acting like* programmable bots—lacking some combination of information and control to be considered an intentional “agent” acting with knowledge and consent.²⁹ Because puppets ceded their private keys (or never had them), they lacked *control*. And

²⁶ For an example of a forked version of Idena client modified by a pool operator see “Commit History of ‘Idena-Mirror’ Repository,” GitHub, accessed December 2, 2023, <https://github.com/haritowa/idenamirror/commits/master>. Some changes aim to hide access to the private key or “remove dangerous buttons.”

²⁷ Puppeteering also includes scenarios where pool operators *don’t* pay participants (e.g., coercion). But not all cases of coercion are puppeteering. For example, a high-information participant may be coerced to share the keys despite understanding the significance. The best evidence to support coercion was a photograph of child puppets (see Fig. 7) by a pool operator, however, it’s unclear if the children were paid or merely thought they were playing a game.

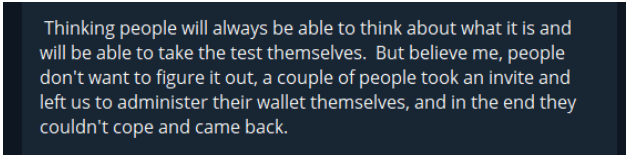
²⁸ The phrase “sybil-attack” in a blockchain refers to when an attacker may create multiple fake identities to gain disproportionate control or influence over a network. See J.R. Douceur, “The Sybil Attack,” in *Peer-to-Peer Systems*, ed. P. Druschel, F. Kaashoek, and A. Rowstron, Lecture Notes in Computer Science, vol. 2429 (Berlin, Heidelberg: Springer, 2002), 251–260. The threshold for “fake” in Proof of Personhood are personhood authentications or attestations from verification ceremonies. In Proof of Work and Proof of Stake, the thresholds for “fake” are compute or stake, respectively. In Proof of Work, the cost of proposing a share of the blocks is proportional to hashing power; specifically, the probability of a miner successfully solving the cryptographic puzzle (and thus proposing the next block) is proportional to their hashing power relative to the total hashing power of the network. In Proof of Stake Ethereum, the cost of creating a validator is 32 eth, and one person with 32*n* eth can create *n* validators. Until recently, the separate literatures on “sybil-resistance” (assuming almost everyone is honest and few adversarial) and “false-name proofness” (assuming no one honest, all utility maximizing) have studied counter-measures. For a discussion of how to unify these literatures with the key parameter of the relative cost of faking an identity, see Bruno Mazon, and Nicolás Della Penna, “The Cost of Sybils, Credible Commitments, and False-Name Proof Mechanisms,” last modified June 29, 2023, <https://doi.org/10.48550/arXiv.2301.12813>. We thank Nikete della Penna for his contribution to this footnote.

²⁹ Although we describe a participant with high-information and control as an “agent,” this is limited to the context of Idena. A person’s agency is not either-or, but context-specific. A person can be an “agent” in *local* contexts with relatively high information and control (parent, local church leader, team manager, informed district voter) and yet be a “*de facto* sybil” (acting like a bot) in more *socially distant*

because puppets were remunerated off-chain in local currency, it was unclear if they were *informed* about the protocol, understood the significance of their validation exercise, or were aware of the puppeteer’s cut from their rewards. In at least two confirmed locations of puppeteering, Indonesia and Russia, the median hourly wage was estimated at \$0.72 and \$2.18, respectively.³⁰ Depending on the fluctuating token price, epoch rewards of a unique account ranged from an hourly wage of \$6.40 at minimum to a median hourly wage of \$56.³¹ Conservatively ignoring the maximum possible earnings, and assuming puppets were paid at least local market rates, puppeteers could capture anywhere between ~\$4 to ~\$55 per hour, or 2x to 55x a puppet’s market wages.

B. Muddied Waters: Puppeteering or Cooperation?

While concerns around puppeteering grew, pool operators began to offer a counter-perspective; they were offering a *service* to consensual participants and being remunerated. Some, including “awakened” puppets, were coming back.



Thinking people will always be able to think about what it is and will be able to take the test themselves. But believe me, people don't want to figure it out, a couple of people took an invite and left us to administer their wallet themselves, and in the end they couldn't cope and came back.

Figure 4: Private message from a Russian puppeteer to Idena Team (Dec 2, 2020)

Earning rewards on Idena presented a number of hassles: continuously running a node on a personal computer for mining rewards, maintaining stable internet connections for validation ceremony rewards, updating software, exchanging rewards into local currency—to name a few. In exchange for ceding exclusive control over private keys, pools could coordinate these tasks better than participants on their own, without interruption, minimizing the risk of identity slashing for failing a validation ceremony or penalties for failing to attest to a block as a validator.

and even nested contexts (feuding distant relatives, Vatican politics, mid-manager of a large company, indifferent national voter). See also n. 50.

³⁰ The median income per day in Russia is \$17.41 and in Indonesia, \$5.74. “Median Income by Country,” *WiseVoter*, accessed December 3, 2023, <https://wisevoter.com/country-rankings/median-income-by-country>. Assuming 8 hour workdays, this translates to an hourly wage of \$2.18, and \$0.72 in Russia and Indonesia, respectively. Another statistic for comparison is average monthly earnings; in Russia, this was \$729 USD in Aug 2023, and in Indonesia, \$192 USD in Dec 2022. See “Monthly Earnings,” *CEIC Data*, accessed December 3, 2023, <https://www.ceicdata.com/en/countries>. Assuming 173 working hours in a month, this translates to an average hourly wage of \$4.21 and \$1.10 in Russia and Indonesia, respectively.

³¹ From January 2020 to May 2022, epoch rewards fluctuated between \$1.60 to up to \$98, with a mean reward of \$18, median reward of \$14 (and standard deviation of 14). Even taking a conservative estimate of \$1.60 every epoch for roughly 15 minutes of “work” (including setup, the 2 minute validation ceremony, and sign-off) this translated to a wage of \$6.40 per hour. Taking the median estimate of \$14 every epoch, this translated to a \$56 hourly wage. See Fig. 16 for the fluctuating value of the token and earnings over time.

We employ a lot of elderly people, they are validated, and receive a decent reward, given that they do not create flips, do not maintain servers, and do not take the risks of a fall in the rate. they would never have been able to figure out all the complexities of ida themselves. At the same time, we leave ourselves a small part of the income, since a lot of expenses go to pay for the creation of flips, the server, and so on.

We are aware of the existence of many groups similar to ours, who simply have not yet reached this size.

Figure 5: Private message from a Russian puppeteer to Idena Team (Dec 2, 2020)

If puppeteering was one extreme, then cooperative, voluntary pools were another possible extreme—coordinated not by puppeteers but by *intentional* participants who voluntarily shared information and control (including their private keys) to a pool operator who could better run their nodes, coordinate validation ceremonies and distribute their rewards, thereby capitalizing on *economies of scale*.³²

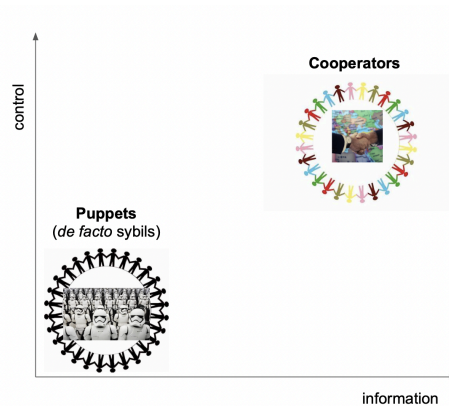


Figure 6: Range of Puppeteering and Cooperation

Yet, without knowing off-chain interactions with the operator or other circumstantial evidence (e.g., telegram chats, photos, etc.), it was impossible to decipher the nature of pools. **Theoretically, the same transaction pattern of blocks of one-way transfers at the same time to the same wallet were consistent with puppeteering and voluntary cooperation. Both extremes differed not in their coordination *on-chain*, but in the distribution of information and control—or *power*—*off-chain*.** In “puppeteering,” operators have *control* (private keys), more *information* (how the protocol works, the social network of puppets, how to cash out rewards into local currency), and capitalize on their asymmetry to control accounts in exchange for minimum wage payments, while maximally extracting to their own advantage. In contrast, “cooperation” has greater symmetry in

³² A possible example of a cooperative pool is 0xb0C3fD00cCd9CEAf17dad2524212021953D6ce0B. See “Idena Blockchain Pool Address 0xb0C3fD00cCd9CEAf17dad2524212021953D6ce0B - Size History,” Idena Blockchain Scanner, accessed December 3, 2023, <https://scan.idena.io/pool/0xb0C3fD00cCd9CEAf17dad2524212021953D6ce0B#sizeHistory>.

information and control.³³ Participants may know how the protocol works and control their keys, but nonetheless find it mutually beneficial to pool resources and delegate control (including their secret keys) to an operator for greater rewards with economies of scale, holding operators accountable off-chain.

To muddy the waters, a pool didn't have to be *either-or*. Even if cooperators voluntarily delegated control (including their secret keys) to an operator, they were also vulnerable to the same principal-agent asymmetries of puppets and at risk of sliding under the thumb of a puppeteer.³⁴ Depending on the off-chain governance to audit and hold the operator *accountable*, pools could slide along the spectrum between “puppeteered” and “cooperative.” Some pools might be more corporate with fiduciary-like duties, or democratic with greater checks and balances to correct for asymmetries, delivering participants more value (a greater benefit of the bargain). Whereas other pools might be more autocratic, extracting with minimum payments and maximum asymmetry. And then there were family-sized pools with strong ties and repeat interactions that might split rewards according to their needs as a group. Similarly, whether a participant was a “puppet” or “cooperator” wasn't *either-or*, but could change over time, depending on what they *knew*, what they could *do*, and more fundamentally to whom they were *socially tied* from talking and trading.³⁵

³³ In game theory, “cooperation” typically refers to the challenge of individuals or entities working together to achieve a collective benefit, particularly in overcoming the temptation to pursue self-interest at the expense of the group. “Coordination,” in contrast, focuses on the strategic alignment of decisions among parties to reach mutually beneficial outcomes, often in situations where multiple equilibria exist and the primary challenge lies in selecting and adhering to a common strategy. Whereas cooperation requires *enlarging* people's motivations, coordination does not. See Michael Suk-Young Chwe, *Rational Ritual: Culture, Coordination, and Common Knowledge* (Princeton University Press, 2001). Throughout this paper, we refer to “puppeteering” and “cooperation” as both instances of social “coordination.” We use “coordination” as a broader term to acknowledge the deeper social relationships of information and control that influence payoffs and shape motivations. For a study of the nuanced relationship between cooperation and coordination in the context of collective action and resource management, see Elinor Ostrom, “Governing the Commons: The Evolution of Institutions for Collective Action,” *Cambridge University Press*, (1990) (showing how local groups can effectively manage common resources, like fisheries or grazing lands, a third way outside traditional market or centralized state mechanisms, or public and private property).

³⁴ Information and control are two different sides of the same coin. Greater control (e.g., over a communication channel) generates an information edge, and vice-versa, more information generates a control edge. For example, puppeteers had an initial information edge which allowed them to bribe and take control over puppet accounts on-chain, by communicating with participants off-chain. Another classic example is authoritarian regimes that centralize their power by controlling communication channels and discouraging social groups (families, civil society)—making participants more uniform in information, with less differences in beliefs, and therefore easier to control in desires. For a discussion of the relationship between control systems and information theory, see Norbert Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine* (Cambridge, MA: The MIT Press, 1948). In related work, one of these authors has suggested moving beyond the binary of atomistic control and information to offer a richer, networked notion of key control (and recovery) that leans on the partial information of uncorrelated participants (informational diversity) as a security strength. See Puja Ohlhaver, Eric Glen Weyl, and Vitalik Buterin, “Decentralized Society: Finding Web3's Soul,” (2022), available at *SSRN*: <https://ssrn.com/abstract=4105763>. See also n. 50.

³⁵ Attention to information is not a random walk, but influenced by *social ties*. People gain information by participating in different social groups' communication channels (e.g., home, work, church). See David Easley and Jon Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (Cambridge University Press, 2010) (offering an intersection of graph theory and game theory). Depending on a group's governance (e.g., autocratic, oligarchic, or democratic), participants may transform this information edge into a control edge relative to other participants, influencing group decisions and governance. A notable historical example is Stalin who rose to power, among other reasons, by mastering bureaucratic politics and infighting. A tactic to rise to power is to simply attend more meetings than peers, however minute, trivial and painfully administrative to gain an asymmetric information edge. Similarly, authoritarian regimes notably seek to consolidate their power by controlling communication channels and discouraging social groups and cohesion (families, civil society)—narrowing information asymmetries about groups (reducing their differences and what is known about them) while widening any groups' asymmetries about them (limiting the information people and groups have about the regime), thereby further centralizing their power. It's not uncommon for autocrats seeking to solidify their own power to pressure disclosure of information on the basis of “information asymmetry,” without revealing information to the participant. See Martin K. Dimitrov,

Whether composed of duped puppets, informed cooperators, or a mix—pools redistributed money and votes unequally, undermining Idena’s egalitarian ambition for *one-account, one-vote, one-reward*. Contrary to the aspirations of a transparent, open network of solo accounts (or “sovereign individuals”) **the protocol instead devolved into *hidden* groups and subnetworks (autocratic, corporate, democratic) competing for control over a fixed economic pie.**

C. *Waking Up Puppets?*

The community began to contemplate strategies to “awaken” puppets. But the breadth and depth of pools could not be quantified easily. Each pool ran a different server for each node with their own IP address, with the optics of individual accounts running their own nodes. Without expensive chain-analysis, the protocol could not readily identify the automated flow of funds. And even if such wallets were identified, off-chain bargains and payments (or lack thereof) couldn’t be inferred from on-chain data. Moreover, conversations with participants were strategic; for example, a participant asking for improvements to the mobile client interface in September 2021 later revealed themselves to be an Indonesian pool operator with over 1,400 paid accounts.³⁶

In December 2020, the protocol began to wrap the user experience with watermarks on the website URL during FLIP tests, both to alert participants and to also prevent FLIP harvesting by pool operators.³⁷ But these measures were generally ineffective.³⁸ One-way transaction patterns continued, and the watermarks program terminated in August 2022. Flooding information can’t awaken participants if they can’t read, understand the watermarks, aren’t motivated to pay *attention*, or more cynically, are easily replaced.³⁹

“Dictatorship and Information: Authoritarian Regime Resilience in Communist Europe and China,” *Oxford: Oxford University Press*, (2023); Isaac Deutscher, “Stalin: A Political Biography,” *New York: Oxford University Press*, (1967); Stephen Kotkin, “Stalin: Paradoxes of Power, 1878-1928,” *New York: Penguin Books*, Illustrated edition, (2015).

³⁶ For the Indonesian pool address, see Idena Blockchain Explorer, “Pool: 0x96d11da40FDe82D81ebE0EAE61bFe6a47F43d1a6 - Size History,” accessed December 3, 2023, <https://scan.idena.io/pool/0x96d11da40FDe82D81ebE0EAE61bFe6a47F43d1a6#sizeHistory>. The Indonesian operator claimed pools were essential to growth, drawing in users who wanted to be “served” without complication, increasing adoption, and therefore making Idena more attractive to investors seeking monopoly. See Appendix B for further discussion.

³⁷ Watermarks were intended to awaken puppets and also prevent pool operators from collecting flips and then submitting them in later ceremonies to reduce the cognitive cost of operating pools. However, watermarks were not effective as human farms continued to grow. In August 2022, watermarks were disabled when anti-AI noise was added to images to make AI-based image recognition difficult. See “Idena Chronicles 0090,” *Medium*, August 15, 2022, <https://medium.com/idena/idena-chronicles-0090-5f3efec5c3f>; see also “Idena Repository,” GitHub, accessed December 3, 2023, <https://github.com/search?q=repo%3Aidena-network%2Fidena-desktop+watermark&type=pullrequests>.

³⁸ Cases of “waking up” were rare. More common were complaints from known pool operators, claiming *their* balances had been stolen. After delegation was instituted in March 2021, a pool operator claimed their balance was stolen in the community telegram, when on-chain activity seemed to suggest a participant with a validated account had simply acquired their private keys and exited their stake along with pooled rewards to their account. For account transactions, see “Idena Blockchain Address 0xF04F3cB6f02c57926eA968F08D55ABb94364F4DF,” Idena Blockchain Scanner, accessed December 3, 2023, <https://scan.idena.io/address/0xF04F3cB6f02c57926eA968F08D55ABb94364F4DF>.

³⁹ Many protocols will attempt to solve puppeteering with information disclosures. However, there is no guarantee that participants care to pay attention, or have an incentive. In economic games with increasing returns and network effects, these information revelations may come too late, after a monopoly power has been established. A better solution to attention curation is to draw on the partial information of adversarial groups to surface what is relevant and worthy of attention with “bridging bonuses” (or correlation discounts). See Section “*From Sybil-Resistance to Collusion-Resistance*.”



Figure 7: Frame of the video from Egyptian pool operator
(Idena Discord Discussion October 2022)

III. PIVOT TOWARDS DELEGATION (March 2021— May 2022)

As pools became common knowledge, the community debated remedies. Hard forking out known pooled accounts wouldn't eliminate hidden pools, but encourage them to re-emerge under new guises. Rather than fight pool coordination—whether cooperative or puppeteered—the community settled on drawing them out of the unquantified shadows with “delegation” in March 2021.⁴⁰ Delegation enabled accounts to band (or be banded) together as groups under a single pool account and node, thereby saving pool operators the hassle (attention, time and money) of running a node for every account while making an account's membership to the pool visible.⁴¹ Importantly, for participants, **delegation enabled operators to handle operational hassles *without* needing to know participants' private keys.** The bargain was three-fold:

- **pool operators could withdraw an account's identity stake** (20% of all earned rewards) and “terminate” a pooled account (reducing the account's status from “human” or “verified” to “killed”), all without a participant's secret keys.
- **pool operators distributed 80% of earned rewards (transferable rewards) at their discretion**, which streamed into the pool operator's wallet.
- **pooled accounts gave up their voting power**, consistent with *one-node, one-vote*. Notably, if an account undelegated from pool, their voting power as a new node would re-instantiate only after 3 epochs (the same time for a participant to validate a new account to “verified” status and gain voting power).

⁴⁰ See “Idena Hard Fork Announcement: Mining Delegation and Oracle Voting,” *Medium*, March 10, 2021, <https://medium.com/idena/idena-hard-fork-announcement-mining-delegation-and-oracle-voting-8a5f9ddd9797>.

⁴¹ The list of pools and their accounts are available on the blockchain explorer. To see an example of a pool, see “Idena Blockchain Pool Address 0x17b851A11f7d37054928BEf47F0F22166d433917 - Delegators,” Idena Blockchain Scanner, accessed December 3, 2023, <https://scan.idena.io/pool/0x17b851A11f7d37054928BEf47F0F22166d433917#delegators>.

With delegation, participants retained control over their wallets holding their transferable rewards, but otherwise pool operators had significant (almost totalistic) control: the power to *distribute* transferable rewards, terminate accounts, and seize identity stake. Thus, informed participants considering delegation had strong incentives to delegate to *trusted* operators: a friend, family member, or operator with reputation and off-chain accountability. Notably, delegation did not change the relationship of puppets under the thumb of a puppeteer; a participant who had *already* unwittingly ceded their private keys to a puppeteer could not reclaim them (or control over their transferable wallet). To reclaim control, a participant would have to validate a new account and race to transfer their funds before the puppeteer.

A. Proliferation of Pools

Delegation succeeded in making pools transparent, but also strengthened incentives to form them with economies of scale. Specifically, pools could now earn the same rewards from validation ceremonies (c) and mining (m) for every pooled account (a) only incurring the time cost of validation ceremonies (t) without the hassle of running an account-specific node continuously (n). Operational costs decreased from an to n , increasing the pool's profit (P).

Pre-delegation: $P_{pool} = a_{pool}(c + m - t - n)$

Post-delegation: $P_{pool} = a_{pool}(c + m - t) - n$

Pools were now cheaper to operate with greater economies of scale than “solo” accounts running their own node. As the network grew—peaking to 15,778 accounts in mid-April 2022—account growth was notably among large pools (>15 accounts). Solo accounts flat-lined (hovering between ~4100 and ~5400), constituting a smaller proportion of the network over time.

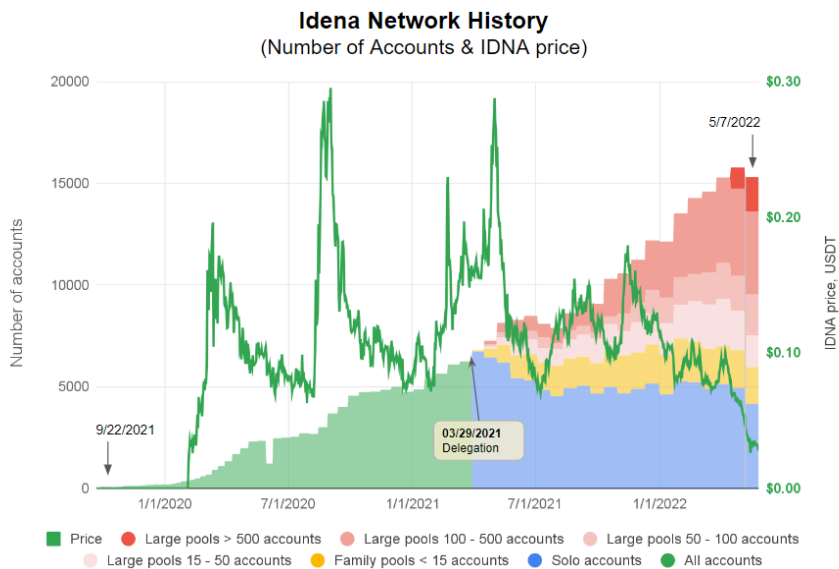


Figure 8: Idena Network History⁴²

From June 2021 to May 2022, solo accounts shrunk from 62% to 27% of the network. Large pools (> 15 accounts) had the reverse trend, ballooning from 22% to 61% of network accounts. Family pools (<15 accounts) hovered consistently between ~12% to 15% of network accounts.

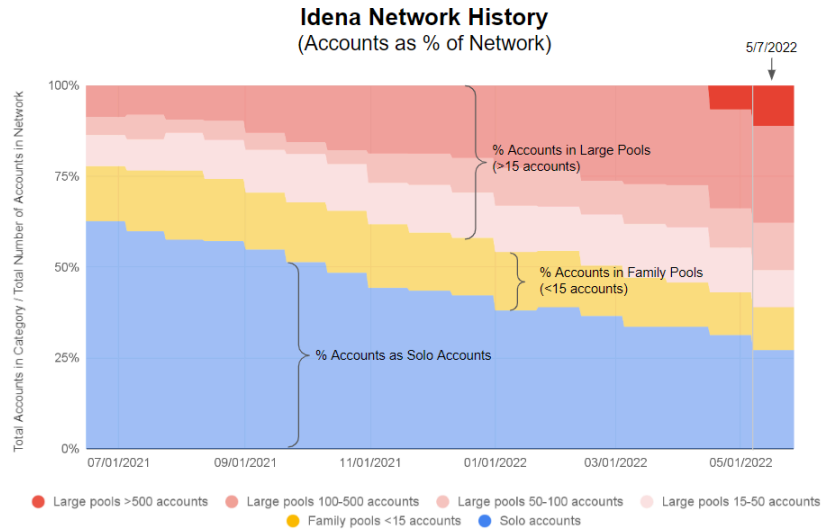


Figure 9: Idena Network Breakdown (post-delegation)

Given *one-account, one-reward*, as large pools ballooned into a larger share of the accounts, they also ate a larger share of rewards from the fixed rewards economic pie, thereby squeezing rewards from solo accounts, which had now become a minority. And because large pools consistently ran their nodes to earn mining rewards, they earned proportionally greater rewards than solo accounts. By mid-May 2022 (May 7), solo accounts constituted 27% of the network accounts but captured 18% of rewards, while large pools (>15 accounts) constituted 61% of network accounts and captured 70% of rewards.

⁴² “Accounts” include any account that has passed at least 1 validation ceremony with “newbie,” “verified,” or “human” status. “Suspended,” “zombie,” and “terminated” accounts are excluded.

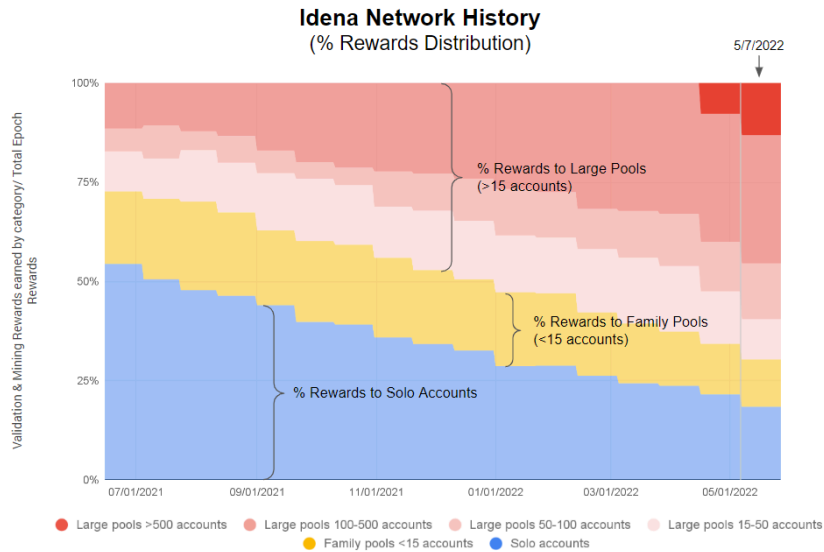


Figure 10: Idena Rewards Breakdown (post-delegation)

As single-node pools bloomed to constitute a larger proportion of the network, **nodes as a percentage of the network in turn precipitously dropped, leading to a loss in throughput and security** (See Fig. 11).⁴³ In the early days of delegation (June 2021), there were 395 pools and nodes constituted 40% of the network. Roughly a year later, by mid-April 2022, when the network peaked at 15,778 accounts and pools climbed to 554, nodes dropped to just 9% of the network. Unless solo accounts with their own nodes captured a larger share of the network, nodes as a percentage of the network would continue to decline.

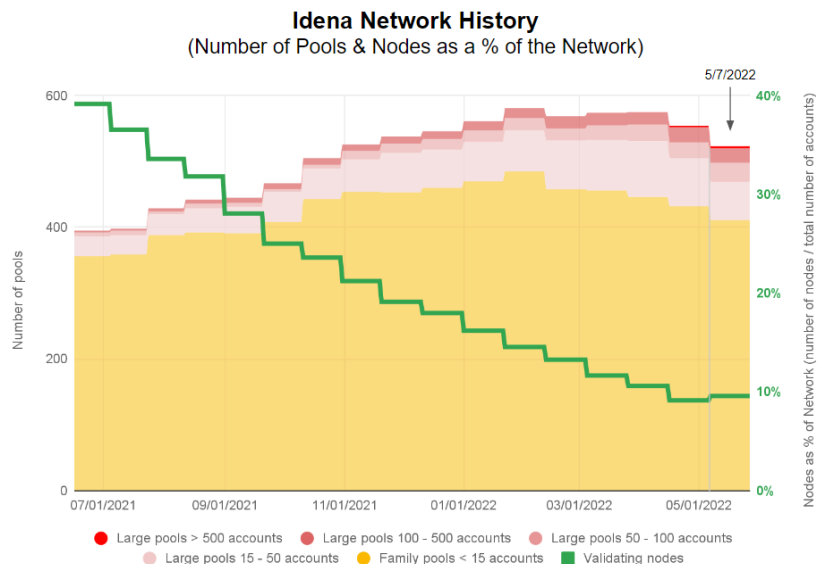


Figure 11: Idena Pool & Nodes History (post-delegation)

⁴³ There were 4 shards on 10/11/2021, 2 shards on 9/30/2022 and 1 shard from 11/24/2022 to the present date.

Given *one-node, one-vote*, although nodes as a percentage of the network declined precipitously, **the voting power of solo accounts increased**. By May 2022, large pools (>15 accounts) constituted a majority of accounts (~61%), but only controlled a minority (~2.4%) of votes. Meanwhile, solo accounts were a minority (~27%) of the network accounts, but controlled a supermajority (~89%) of votes.

B. 3rd Party Key Access in Top Pools

The purpose of delegation was to make coordinated pools visible and enable operators to handle operational hassles for participants (e.g., running a node) without needing to know their private keys. But oddly, large pools continued to show signs of 3rd party key access.⁴⁴ A tell-tale sign was the unlikely coincidence of *simultaneous* or *sequential* transactions from different accounts in the same pool (e.g., “account delegations” or “account terminations”). Funneling transactions were also corroborative:

- pool **operators withholding (rather than distributing) transferable rewards**, often then funneling them to an exchange or a hive wallet
- delegated **accounts funneling all transferable rewards earned *before* delegation** to the pool operator, before the pool operator provided any service
- delegated **accounts funneling all identity stake *after* the account was terminated** to the pool operator⁴⁵

Looking at these factors, we examined all 31 “top pools” that had ever been delegated more than 100 accounts in the protocol’s history, including pools that had been delegated more than 100 accounts *after* May 2022. **All top 31 pools exhibited all signs of 3rd party key access**—both simultaneous/sequential transactions along with funneling—with minor exceptions. One pool showed funds flowing back to some accounts, which suggested the operator might be a service provider. But upon closer examination, this pool belonged to an Egyptian pool operator, with photographic evidence of child puppets (see Fig 7. & Appendix B).⁴⁶ Another pool appeared to have accounts that didn’t funnel rewards earned *before* delegation, but showed all other signs. This pool was later revealed to be part of a larger network of puppeteering pools making inter-pool transfers in a complex web of transactions (see Appendix B).

C. An Emergent Puppeteering Oligopoly in Top Pools

⁴⁴ At the start of the protocol, since there were no penalties for getting “caught,” there was no obvious incentive to obfuscate 3rd party key access. In the future, we expect puppeteers to muddy or hide transaction patterns (e.g., Dark DAOs, see n. 90). In other words, there will be a cost to getting “caught” as a puppeteer, and the resulting strategies will incur obfuscation costs that are less than the expected value of avoiding detection.

⁴⁵ Similarly, pooled accounts simultaneously or sequentially transferring their transferable rewards from their member wallets to the pool operator wallet upon account termination was another sign, although this was less observed as most pool accounts lacked any rewards to transfer because puppeteers rarely distributed rewards to member accounts.

⁴⁶ The Egyptian pool with child labor photos was treated as a puppeteering pool, despite signs of being a possibly mixed pool of exploited and some (likely a minority of) consensual participants. On May 7, they held 91 accounts, or approximately 0.6% of the network. Given the small percentage of the network, it does not substantially change our analysis.

All top 31 pools showed evidence of 3rd party key access, raising another question: were these pools distinct and independent, or did they share the *same* operator? The Russian operator who first admitted to running a human pool in 2020, for example, revealed controlling multiple pools in 2021(see Appendix B). Cursory chain-analysis showed financial transfers between pools, confirming a shared operator. Extending the analysis to the remaining pools, **if pools with financial ties were treated as the same entity (or subnetwork), the 31 top pools were in fact 23 entities.** Moreover, by May 2022 almost half of all accounts in the top 31 pools were distributed to just 3 entities, or sub-networks: 24% belonged to the Russian enterprise (yellow), 10% to an Indonesian entity (blue), and 13% to an entity with unknown social origins (red) (see Appendix B). There was an emergent oligopoly of 3 entities among the top 31 pools.

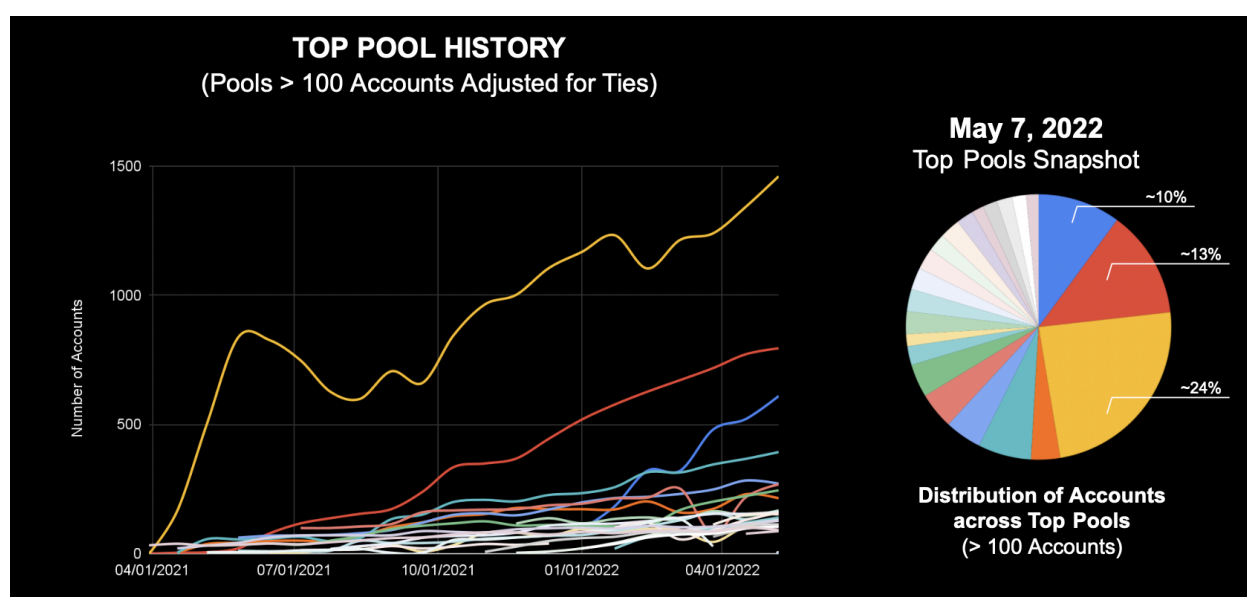


Figure 12: History of top pools accounting for ties among top pools (>100 accounts) & May 2022 Snapshot

Zooming out to the broader network, since all 23 entities (and their constituent 31 pools) had strong evidence of 3rd party key access, and 3rd party key access confers pool operators totalistic control over accounts, we could reasonably infer that by May 2022, **23 entities constituting less than 0.6% of the network's distinct entities controlled at least ~40% of accounts and the distribution of almost half (~48%) the network rewards, and 3 entities (or sub-networks) controlled ~19% accounts and ~24% rewards.**

May 7, 2022 Snapshot of Idena Network

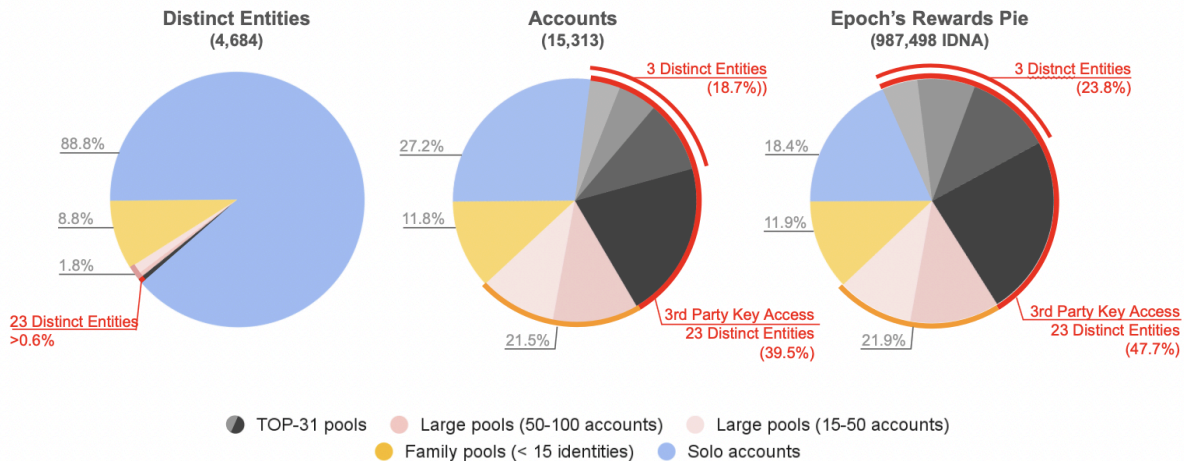


Figure 13: Idena Network May 2022 Snapshot

D. From 3rd Party Key Access to Puppeteering

Philosophically, a “puppet” connotes a participant who lacks *information* and *control* to be considered a meaningful participant acting with knowledge and consent. In the context of a blockchain protocol, secret (or “private”) keys are the locus of account control, logically implying two kinds of puppeteering:

- **Strong Puppets:** “low-information” participants who are unaware of their account’s secret keys, which are controlled exclusively by a 3rd party (exclusive 3rd party control)
- **Semi-strong Puppets:** “low-information” participants who “know” their secret keys but are unaware of their significance and share access with a 3rd party (non-exclusive 3rd party control)⁴⁷

To some—particularly those who subscribe to “not your keys, not your coins”—3rd party key access is *prima facie* evidence (or a sufficient condition) of a “low-information” participant, or puppet.⁴⁸ Under this logic, 3rd party key access confers totalistic control to an operator who can in turn loot wallets with impunity—a risk *only* a “low-information” participant either wouldn’t know about or unwittingly assume (especially after participants *could* delegate operational hassles to an operator *without* ceding their private keys). Given the strong evidence of 3rd party key access in the top 31 pools, the conclusion is that their delegated

⁴⁷ In semi-strong puppeteering, there is an opportunity puppets may “awaken” and presuming they race to their account before the “puppeteer,” they may exit their funds and send them to a new account under their exclusive control.

⁴⁸ Stated formally, the syllogism is: 1) a participant is low information if and only if the participant is a puppet 2) if a participant’s account has 3rd party key access (exclusive or non-exclusive), then the participant is “low-information.” 3) the participant account has 3rd party key access (exclusive or non-exclusive), therefore the participant is a puppet (strong or semi-strong, respectively).

accounts—at least ~40% of Idena’s network—were puppeteered, where “low-information” participants were unaware of their account’s private keys (“strong puppets”), or knew their private keys but were unaware of their significance (“semi-strong puppets”).

From another perspective, however, 3rd party key access doesn’t sufficiently imply puppeteering and *could* be a voluntary choice by a “high-information” participant. Although delegation eliminates operational hassles for participants *without* an operator needing to know their private keys, the hassle of managing private keys remains, akin to managing a password without reset or recovery options.⁴⁹ Under this view, Idena is one system, embedded among others, where participants have varying opportunity costs of time and attention. 3rd party key access *could* be a voluntary “custody-as-a-service” choice by a “high-information” participant, presuming they could hold operators accountable *off-chain*, either informally (social pressure, reputation) or formally with legal recourse. With robust accountability, the relationship implied by 3rd party key access is not puppeteering *per se*, but could also be a principal-agent relationship where optimistically, agents disclose relevant information to their principals and have sufficient feedback to represent their principals’ interests, for a fee.

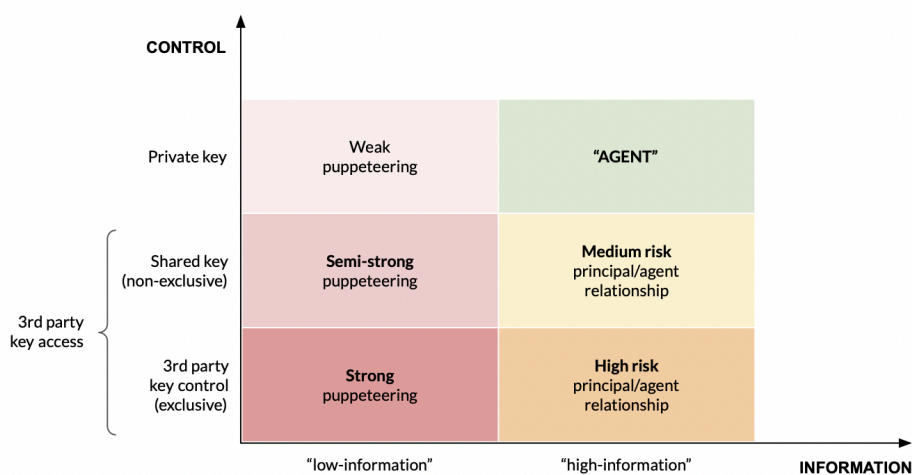


Figure 14: Simplified Representation of a “Agency” in a Single Game⁵⁰

⁴⁹ One might argue another possible benefit to ceding key control was to empower the pool operator to unilaterally un-delegate at any moment from *one-pool, one-node* to *one-account, one-node* for more voting power in protocol hard-forks. But trusting the operator to strategically undelegate both naively presumed that the operator would be voting changes to the participants’ benefit (not their own) and risked being perceived by the rest of the network as an attempt to game outsized influence, either to enact protocol changes to a pool operator’s benefit (capture) or worse, a 51% attack—an attempt to gain control of the majority of network nodes with the power to censor blocks, capture mining rewards, and selectively validate identities. With a 3-epoch waiting period for un-delegation, the network would also have ample time to defend itself and hard-fork out undelegated identities in a new chain. Any participant who valued strategic undelegation was more likely to be an uninformed puppet than a knowledgeable participant. Notably, strategic un-delegation was unobserved, as puppeteers weighted the economic gains from consolidating under one node more than strategic un-delegation.

⁵⁰ We offer a stylized and (misleading) representation of a participant in one game. Contrary to this representation, a participant who has delegated control to an accountable agent with better information may in fact be a *weaker* “puppet” than a participant who controls their keys exclusively but has stale information (e.g., is unaware of participants coordinating in side-channels outside the protocol). This counterexample underscores this representation’s flaw: mainly, it presumes an “agent” is *homo economicus* having full information and control to not be a puppet, but socially isolated, lacking social ties to collude (or be colluded against), obscuring asymmetries that arise from social reality. Yet, this representation underpins many flawed experiments in identity and blockchain protocols. To the contrary, social reality is that humans are members of recombining social groups with *partial* information, *partial* control, coordinating for more

But delegation has risks, particularly when the incentives between principals and agents are misaligned.⁵¹ Without accountability mechanisms to surface *relevant* information from both parties for continued alignment and ensure the agent prioritizes the principal’s interests above their own, a principal-agent relationship (yellow zone) risks devolving to puppeteering (red zone), where “high-information” principals become “low-information” puppets.⁵² **From this standpoint, the combination of third-party key access *and* lack of accountability constitute sufficient conditions for puppeteering.**⁵³ Specifically, an absence of accountability explains *why* a participant becomes a “low-information” participant:

- **Strong Puppets:** “low-information” participants who are unaware of their account’s secret keys, which are exclusively controlled by an *unaccountable* 3rd party.
- **Semi-strong Puppets:** “low-information” participants who know their account’s secret keys but are unaware of their significance and share access with an *unaccountable* 3rd party.

Assuming this stricter definition, we argue that *at least* ~40% of Idena’s network was puppeteered (strong or semi-strong) given the evidence of 3rd party key access paired with an *unlikely silence* from participants: an *improbable* lack of marketing and disputes that otherwise would accompany *off-chain* accountable custody relationships where agents have been conferred totalistic control over principals’ accounts. Given the breadth of accounts with 3rd party key access—at least 40% of the network accounts—at least *a few* complaints of broken bargains would be expected (for example, by participants who voluntarily entered into custody relationships, but

power (*more* information and *more* control), often leading to accidental or unintentional collusion. We are members of many games—many nested—and may be “high-information” in some, and “low information” in others. Given that no participant can have “full information” in a multi-player, multi-game scenario, one of these authors argues in future work that “agency” is a *relational* and *dynamic* property, depending on the *partial* information and *partial* control a participant has *relative* to other overlapping and disparate social groups. Paradoxically, a participant may become more “agentic” and “secure” the less they are a single point of coercion and the more key recovery is spread across conversational partners with uncorrelated “high-bandwidth” channels to the participant. Plural recovery is a third way between the Scylla of self-custody and Charybdis of third-party custody, leaning on the consent of a qualified majority of conversational partners with whom the participant has crossed socially “long bridges” to reach (“security in diversity”). See Ohlhaber, Weyl, and Buterin, “Decentralized Society.” For a discussion of “embeddedness,” see Mark Granovetter, “Economic Action and Social Structure: The Problem of Embeddedness,” *American Journal of Sociology* 91, no. 3 (November 1985): 481-510. For seminal work in nested games, see George Tsebelis, *Nested Games: Rational Choice in Comparative Politics*. University of California Press, 1990.

⁵¹ For an overview of how law mitigates agency costs through mechanisms like fiduciary duties, contractual obligations, and liability rules, see Richard Posner’s seminal work on the economic analysis of law. Richard A. Posner, *Economic Analysis of Law*, 7th ed. (Austin: Aspen Publishers, 2007). For a shorter introduction, see Eric A. Posner, “Agency Models in Law and Economics” (2000), University of Chicago Law School, John M. Olin Law and Economics Working Paper No. 92, <https://ssrn.com/abstract=204872>. See also n. 82.

⁵² We presume a non-dystopian backdrop. In some dystopias, individuals might opt to relinquish their autonomy, akin to the conundrum of willingly entering into servitude. Such equilibria may emerge when the dominant strategy favors subverting coordination systems, especially where interactions manifest as a pervasive prisoner’s dilemma or a “stag hunt” scenario, where mutual distrust can lead to suboptimal outcomes for all parties involved. Brian Skyrms, “The Stag Hunt,” *Proceedings and Addresses of the American Philosophical Association* 75, no. 2 (2001): 31–41, <https://doi.org/10.2307/3218711>.

⁵³ Stated formally, the syllogism is: 1) a participant is low information if and only if the participant is a puppet 2) if a participant’s account has 3rd party key access (exclusive or non-exclusive) and the 3rd party is unaccountable, then the participant is “low-information.” 3) the participant account has 3rd party key access and is unaccountable, therefore the participant is a puppet. Stated informally: if a 3rd party has key access (exclusive or non-exclusive) to a participant’s account *and* the 3rd party is not accountable to the participant, then the participant is a low-information puppet (“strong” or “weak,” respectively). We use the language of accountability, however, the definition can also be expressed in terms of unmitigated agency costs.

were never adequately remunerated off-chain by the operator).⁵⁴ Yet, formally, there was no evidence of legal disputes nor marketing around key custody services. Notably, the jurisdictions of known large pools—Russia, Indonesia, Egypt—combined cheap labor with poor rule of law, making legal recourse a challenge.⁵⁵ Informally, and to our knowledge, the community forums also lacked complaints from participants seeking recourse from pool operators. To the contrary, pool operators were vociferous in how to grow pools or retrieve “stolen” funds from participants (or possibly managers) who had exited with “*their*” funds.⁵⁶ If there were *informal* accountability mechanisms, they would more likely be present in smaller family-sized pools (<15 accounts) with strong social ties and less likely with larger pools (>100 accounts) with weak social ties.⁵⁷ Yet, as new accounts ballooned, solo accounts and family pools flat-lined, while new accounts quickly delegated to large pools—a signal of paid *recruits*. By May 2022, only ~12% of accounts were in family pools, while ~40% of the network’s accounts were in large pools (>100 accounts).

The admissions of the largest pool operators also corroborated puppeteering. Two from Russia and Indonesia—controlling ~14% of the network accounts by May 2022—openly confirmed they paid participants to perform validation ceremonies (see Appendix B). Finally, the 3 largest networks (Russian, Indonesia, and unknown) constituting ~19% of the network by May 2022 had meteoric rises and subsequent falls (see Appendix B), consistent with the hypothesis that they were paid enterprises sensitive to unit economics (strong puppeteering), or vulnerable to participants “waking up” and exiting their accounts (semi-strong puppeteering).

Combined, the fact pattern around the top 31 pools, where 3rd party key access conferred operators totalistic control over ~40% of the network accounts, was more consistent with puppeteering than an accountable off-chain custody relationship:

- *Silence*: absence of advertising around accountable key custody services, formal legal disputes, and informal customer complaints on the community forum.⁵⁸
- *Rule of law*: the known jurisdictions of 3 large pools were weak in rule-of-law (Russia, Egypt, Indonesia).
- *Pool size and growth*: solo accounts and family pools with strong social ties flat-lining, while large pools (> 100 accounts) with weak social ties blooming.
- *Communication*: conversations with the top “human pool” operators controlling ~14% of accounts confirming they pay participants to perform validation ceremonies, including photographs of child participants (see Figs. 3-5, Fig. 7, Appendix B).

⁵⁴ Pool operators who controlled keys also did not distribute transferable rewards to participants but sent them to exchanges. Assuming custody-as-a-service, presumably the operator would distribute funds off-chain to participants after currency conversion and that conversion would generate a few disputes.

⁵⁵ “Rule of Law Index,” *World Justice Project*, accessed January 15, 2024, <https://worldjusticeproject.org/rule-of-law-index/>.

⁵⁶ See n. 38.

⁵⁷ Mark S. Granovetter, “The Strength of Weak Ties,” *American Journal of Sociology* 78, no. 6 (May 1973): 1360-1380 (introducing “weak ties” as bridges for information diffusion across distant groups); Minjae Kim and Roberto M. Fernandez, “What Makes Weak Ties Strong?,” *Annual Review of Sociology* 49 (July 2023): 177-193 (highlighting empirical research supporting strong ties as more effective channels for information diffusion than weak ties when, for example, brokers over weak ties may extract a “control benefit,” information is more complex, or the information is less relevant to beneficiaries).

⁵⁸ While absence of evidence around accountable key custody services is not evidence of unaccountable puppeteering, we find this absence in a global digital protocol improbable, especially when coupled with with positive signs of on-chain 3rd party key access and other indicators (e.g., lack of rule of law, pool growth, pool size, anecdotal communication among the largest pool operators).

- *Meteoric rise and fall*: the top 3 networks (Indonesian, Russian, and an unknown network) controlling ~19% of the network accounts having meteoric rise and falls (see Appendix B).

Our analysis of 3rd party key access in May 2022 was limited to the top 31 pools with more than 100 accounts. We ignore 411 family pools (<15 accounts) and 84 large pools (15<n<100 accounts)—roughly 30% of the network’s accounts. But there is reason to suspect that many (if not most) of the 84 large pools—delegated 21.5% accounts—were also puppeteered. Confirmed puppeteering pools show low average transferable wallet balances relative to family pools, consistent with puppeteers cashing out regularly.⁵⁹ The unanalyzed 84 pools also show this skew in low average wallet balances. Moreover, identity stake—which was locked and accrued at an automatic rate of 20% rewards per epoch—illustrated a distribution of what transferable wallet balances *would* look like if puppeteers could not cash out, with average stakes being within the same order of magnitude.⁶⁰

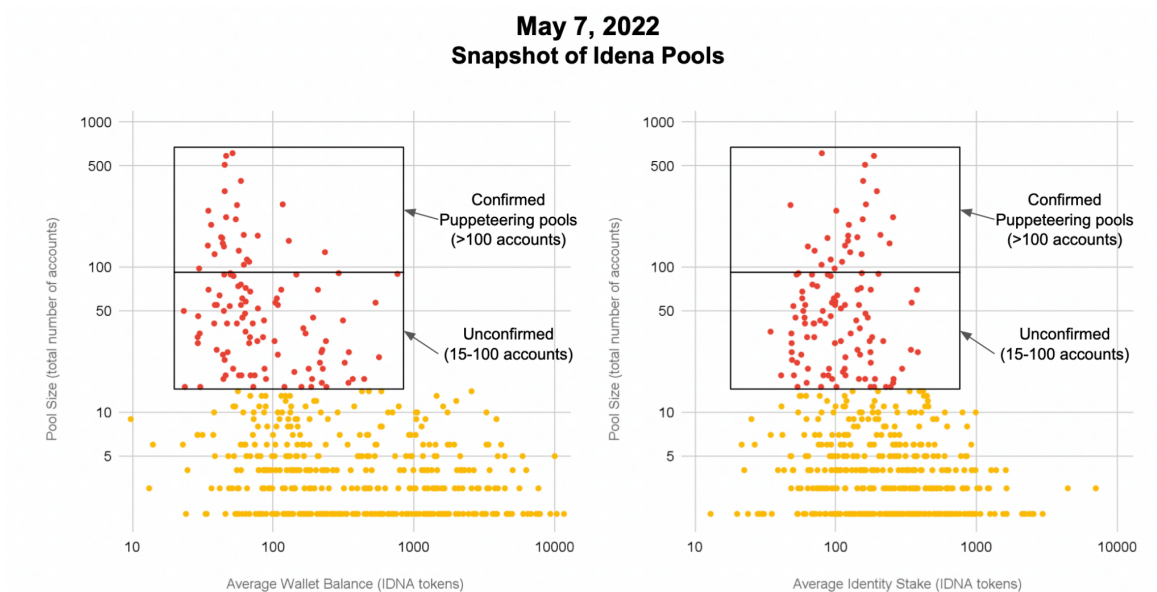


Figure 15: Idena Network Snapshot of Pools on May 7, 2022

We welcome readers skilled in chain analysis to continue this research effort and confirm 3rd-party key access for these unconfirmed 84 large pools. In addition, we welcome further chain-analysis examining the pool funnel to and from solo accounts, pool shopping, pool churn and lifespan (see Appendix B).

⁵⁹ For pools, all account transferable rewards funneled into a single pool operator’s wallet. So to calculate the mean and median wallet balance for pools first required adjusting the wallet balance on a per account basis (pool’s wallet balance/pool’s number of accounts). Single accounts (not displayed) have a single wallet, and therefore the “average wallet balance” is simply the wallet’s balance. For the confirmed 31 puppeteering pools, the average of all pools *average wallet balances* was ~64 (standard deviation 42), while the average for 84 unconfirmed pools was ~129 (standard deviation 132), and 411 family pools was ~1628 (standard deviation 6044).

⁶⁰ If pools did not cash out rewards, mean transferable wallet balance values would be significantly higher, as 80% of rewards flowed to transferable wallets. The identity stake mean value is based on mandatory 20% identity stake lock-up. For the confirmed 31 puppeteering pools, the average of all pools’ *average identity stake* was ~134 (standard deviation 54), while the average for 84 unconfirmed pools was ~123 (standard deviation 78), and 411 family pools was ~383 (standard deviation 557).

E. Panic and Threat of Collapse

The intended model of *one-person, one-vote, one-reward* had not simply collapsed to *one-pool, one-vote, n-accounts rewards* for ~73% of the network, but for at least 40% of the network, it was *one-puppeteering enterprise, one-vote, n-puppet-account rewards*. And yet our analysis was cursory, excluding 84 large pools (100>n>15 accounts) and 411 family pools (<15 accounts)—in total 94.6% of pools. The statistics could only get worse, not better.

As human farms proliferated and captured a larger share of the rewards, large pool operators (puppeteers) were also selling these rewards immediately, increasing the selling pressure of the community IDNA token. At a low enough price, pools would become unprofitable to operate. For the time being, however, pools were growing, despite the dropping price. Presuming participants were being paid, the account rewards earned per epoch fluctuated between roughly \$14 and \$2, which was enough to pay market rates in Russia or Indonesia.

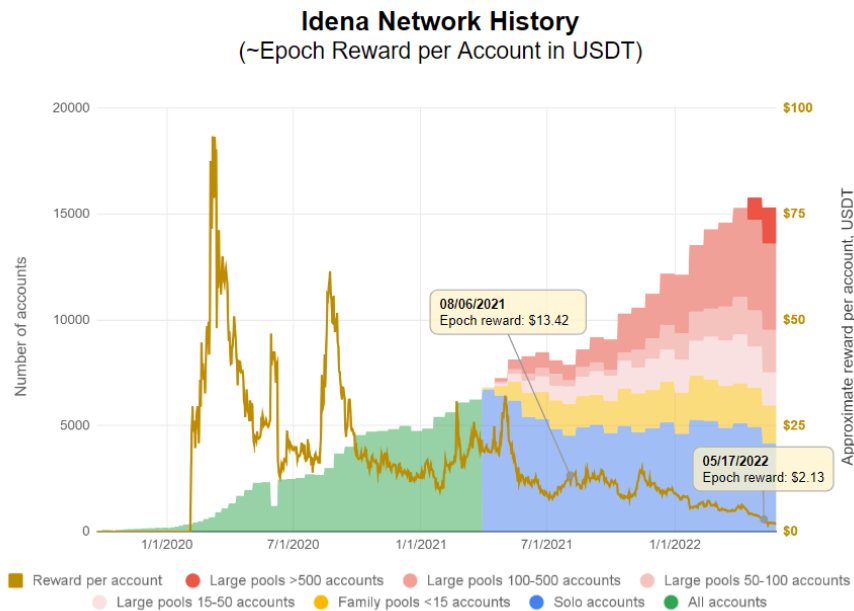


Figure 16: Approximate Epoch Rewards earned per Account in USDT

But unchecked, the trajectory of human farms and large pools risked collapsing the protocol. Morally, solo accounts who joined *because* of the protocol’s egalitarian ambitions were wary to *stay* in a network captured by puppeteers who extracted disproportionate rewards; less than 0.6% entities controlled the distribution of almost half the network’s rewards. Economically, puppeteers weren’t just capturing a larger piece of the economic pie in every epoch, they were also shrinking the economic pie because they aggressively cashed out their rewards (compared to family pools and solo accounts). If the trend continued, solo accounts would continue to drop in absolute and proportional terms, and the number of independent nodes—already at 10% of the network—would shrink more, undermining security. With fewer unique participants in the network, the greater risk of 51% attacks, through collusion among pools or strategic un-delegation, where a pool un-delegates accounts to individual nodes

to increase their voting power.⁶¹ Even if the rapidly shrinking minority of solo accounts hard-forked out attackers in strategic un-delegation and formed a new chain, malicious actors could re-join the new fork to repeat the attack. Without modification, the protocol would collapse.

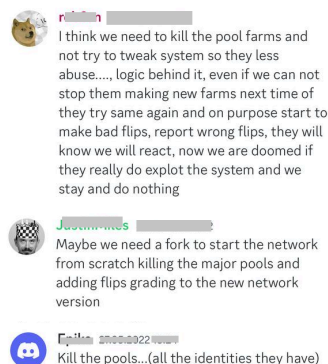


Figure 17: Community discussions on *Idena discord server* (May 2022)

Yet, when it came to voting protocol changes, *one-node, one-vote* offered an important advantage: accounts in large pools were treated as the *same* entity with discounted influence. So although solo accounts were only ~27% of the network accounts on May 7, 2022, they controlled ~89% of the voting power, offsetting the otherwise puppeteering and plutocratic majority. In another essay, we unpack how discounted influence enabled solo accounts to pivot the Idena protocol towards a novel experiment in sublinear identity staking that shifted incentives away from puppeteering, although introducing a different set of challenges.

IV. DISCUSSION

A. *A Failure in Egalitarianism (one-person, one-reward, one-vote)*

Human identity is as diverse and dynamic as the unique combination of associations that individuate a person. Yet, Proof of Personhood is reductive, compressing identity into a standardized binary (“verified” or “not verified”) and overlooking the social and economic ties from talking and trading that differentiate people. When power is at stake—money, votes—such global identity systems with uniform rules for qualifying as “human” pave the way for those who *already* have power—those with resources, knowledge, or status—to find loopholes, align interests, and collude to exploit the system’s simplicity to their advantage. Idena offers a cautionary tale. Despite *technically* proving biological uniqueness, the protocol’s initial experiment with Proof of Personhood had a range of unintended *social* consequences that departed from egalitarian aspirations of *one-person, one-vote, one-reward*, fracturing instead into groups and subnetworks—some cooperative, most puppeteered—competing over an economic pie to the detriment of solo accounts. Notably, a few dozen enterprises rapidly rose to the top, controlling *at least* ~40% of the network’s accounts and the distribution of almost half (~48%) the network rewards. Just 3

⁶¹ Strategic un-delegation can be a kind of 51% attack when a pool operator undelegates member accounts to increase the numbers of nodes, and by extension voting power. See n. 49 for a discussion of strategic un-delegation.

controlled ~19% accounts and ~24% rewards. **An off-chain oligopolistic system subsumed an on-chain egalitarian system, giving rise to a shadow power structure that operated opaquely and silently.** On-chain, *asocial* personhood collapsed into off-chain *social* arrangements that obfuscated power at best, or reinforced it at worst.

Proof of Personhood protocols such as Worldcoin⁶² and Proof of Humanity⁶³ have also been riddled with exploits, underscoring that Proof of Personhood's challenges are not accidental, but systemic. Notably, WorldCoin WorldID has had to grapple with allegations of account trading—one-time sales of private keys which allow buyers to control accounts and accrue UBI rewards.⁶⁴ While a one-time account sale is different from an ongoing relationship of puppeteering, it does not imply immunity to puppeteering. To the contrary, in Idena, puppeteering came *after* protective mechanisms that stymied account trading, *not before*; these mechanisms included identity staking, identity slashing, periodic re-validation (or re-authentication),⁶⁵ and simple account revocation and re-registration.⁶⁶ After successfully discouraging account trading, the next best strategy to game disproportionate

⁶² The largest Proof of Personhood network is Worldcoin WorldID (<https://worldcoin.org/>). Worldcoin's World ID verifies the uniqueness of participants through biometrics (storing "iris codes" converted from hashed iris scans) and rewards participants with WDC tokens. See n. 12. For a general discussion, see Edd Gent, "Is Worldcoin a Crypto-Currency for the Masses or Your Digital ID? The project aims to scan all the world's eyeballs," *IEEE Spectrum*, December 22, 2022, <https://spectrum.ieee.org/worldcoin>; Frank Hersey, "Worldcoin Says SDK Lets You Prove You're a Human Online. Coins Not Included," *Biometric Update*, March 17, 2023, <https://www.biometricupdate.com/202303/worldcoin-says-sdk-lets-you-prove-youre-a-human-online-coins-not-included>.

⁶³ See "Sock Puppeteer," *IPFS* (evidence submitted to Kleros court case concerning Proof of Humanity profile 0xe825e609d15dd004d4b35dd858a55fd094db7f11 engaging in sock puppeteering), accessed December 3, 2023, <https://ipfs.kleros.io/ipfs/QmNQxQff3UN4KfHqjxvNca8pv96dUSKvvd7fiQCfz3AV/sock-puppeteer.pdf>. In Proof of Humanity (<https://proofofhumanity.id/>), participants submit a video proving their human status along with a deposit that is refunded if verification is successful. The submission must be vouched for by at least one of the already verified participants. After vouching, the submission enters a "pending" status that lasts for several days, where any verified participant can challenge the submission and withdraw the deposit if the online court (Kleros) confirms a violation of the protocol's rules. If submission is not challenged or challenges are not successful, then the participant gets "verified" status. Apart from an allocation of tokens reserved for investors and teams, participants are otherwise rewarded with UBI tokens and voting power at its governing DAO, which has evolved away from 1p1v. Humanode (<https://humanode.io/>) is another Proof of Personhood protocol worthy of mention, which uses facial recognition AIs to verify the uniqueness of participants and rewards them with HMND tokens and voting power in its DAO.

⁶⁴ According to social media posts, sellers were offering KYC verifications for the World App on Chinese social media and ecommerce sites, with credentials funneled from developing countries, such as Cambodia and Kenya. See Eliza Gkritsi and Lingling Xiang, "Black Market for Worldcoin Credentials Pops Up in China," *Coindesk*, May 24, 2023, <https://www.coindesk.com/policy/2023/05/24/black-market-for-worldcoin-credentials-pops-up-in-china>. There have also been allegations that the market price for a WorldCoin ID was at one point \$30. See Andrew M. Bailey and Nick Almond, "Worldcoin Isn't as Bad as It Sounds: It's Worse," *Block Works*, July 26, 2023, <https://blockworks.co/news/worldcoin-privacy-concerns>.

⁶⁵ For re-authentication in WorldCoin, among the proposed methods are returning to an orb for an iris scan or facial recognition with a user's phone using zero-knowledge machine learning—a cryptographic method that allows one party to prove to another that they know a certain piece of information without revealing that information; "facial recognition, performed locally on the user's device in a fashion similar to Face ID, could be used to authenticate users, thereby ensuring that only the person to whom the World ID was originally issued can use it for authentication purposes." However, the White Paper also acknowledges trust assumptions; "given that the user's device is not intrinsically trusted, there is no absolute assurance that the appropriate code is being executed nor that the camera input can be trusted." See "WorldID: Implementing PoP at Scale," *Worldcoin Whitepaper*, accessed December 3, 2023, <https://whitepaper.worldcoin.org/proof-of-personhood#recovery-2>. For a general discussion of security and trust assumptions, see Matthew Green, "Some Rough Impressions of Worldcoin," *Matthew Green's Blog*, August 21, 2023, <https://blog.cryptographyengineering.com/2023/08/21/some-rough-impressions-of-worldcoin/>.

⁶⁶ Revocation and re-issuance (or re-registration) of accounts are key mechanisms to subvert coercion. See Vitalik Buterin, "What Do I Think about Biometric Proof of Personhood?" *Vitalik Buterin's Blog*, July 24, 2023, <https://vitalik.eth.limo/general/2023/07/24/biometric.html>. According to the WorldCoin Whitepaper, participants "can get their World ID re-issued by returning to the Orb..but neither

rewards in Idena became buying cheap labor (participants' time and attention) to control accounts. Thus, illicit account trading in protocols should not be treated as evidence of advanced mechanisms or protections; to the contrary, *illicit trading may signal a lack of them* and be a *precursor* to puppeteering.⁶⁷ Given global economic disparities, the rewards need not be significant to incent strategic behavior: just \$2 for 30 minutes of work every few weeks, as the Idena experiment proved. Even charitably assuming operators paid participants, they could extract up to the delta between labor market price and the account's epoch rewards, which in at least two confirmed locations (Russia and Indonesia) ranged anywhere between 2x to 55x of a participant's market wages, depending on the fluctuating value of the IDNA token. WorldCoin WorldID's ambition for an AI-funded UBI anticipates significantly greater rewards (and greater deltas) to motivate off-chain strategic behavior, whether account trading *or* puppeteering.

B. Democratic Governance in a Network

Idena offers a cautionary tale about naive attempts at democratic governance with *one-person, one-vote*. A condition of democracy is that participants can express their will and intent without coercion.⁶⁸ In Idena, some participants preferred joining a pool, and delegating voting and partial economic control. But many gave up totalistic control—ceding their private keys—without understanding the significance of their participation, or that they *could* express an intent. In the optimistic case, participants traded their time for a paycheck. Yet, this trade became akin to ***vote-buying, where the well-resourced could buy more time and therefore more votes, transforming a system that was intended to be one-person, one-vote into one-token, one vote where plutocrats and puppeteers gain outsized influence.*** Were it not for Idena's pivot to delegation which discounted the voting power of pools—*one-pool, one-vote*—plutocrats and puppeteers would have continued to wield outsized influence, where outcomes disproportionately reflected *their* interests, not the underlying

other credentials held by the user's wallet nor the wallet itself can be recovered..." See "WorldID: Implementing PoP at Scale," *Worldcoin Whitepaper*, accessed December 3, 2023, <https://whitepaper.worldcoin.org/proof-of-personhood#recovery-2>. As of December 2023 (and to our knowledge), this re-issuance method has not been yet implemented, though it is a stated priority for the protocol in 2024. See WorldCoin, "Introducing World ID 2.0," *WorldCoin Blog*, December 13, 2023, <https://worldcoin.org/blog/announcements/introducing-world-id-2.0>.

⁶⁷ When account trading becomes harder (or more costly), puppeteering becomes the next best alternative for the unscrupulous. A similar non-intuitive escalation is when bribery becomes harder (more costly), physical violence may become a preferred alternative among criminals, further underscoring the importance of systems thinking. See Massimo Pulejo and Pablo Querubin, "Plata Y Plomo: How Higher Wages Expose Politicians to Criminal Violence," *NBER Working Paper No. w31586 National Bureau of Economic Research*, (2023); DOI 10.3386/w31586 (examining the significant increase in criminal violence against Italian municipal cabinet members after an increase in their wages).

⁶⁸ See Danielle Allen, "Justice by Means of Democracy," *University of Chicago Press*, (2023). See also Stevens Le Blond, Alejandro Cuevas, Juan Ramón Troncoso-Pastoriza, Philipp Jovanovic, Bryan Ford, and Jean-Pierre Hubaux, "On Enforcing the Digital Immunity of a Large Humanitarian Organization," *École Polytechnique Fédérale de Lausanne*, accessed December 3, 2023, <https://bford.info/pub/dec/immunity.pdf> (offering a qualitative analysis of the security and privacy challenges, including coercion, that humanitarian organizations face when collecting, processing, transferring, and sharing data to enable humanitarian action). See also Bryan Ford, "Identity and Personhood in Digital Democracy: Evaluating Inclusion, Equality, Security, and Privacy in Pseudonym Parties and Other Proofs of Personhood," *Swiss Federal Institute of Technology in Lausanne (EPFL)*, (2020): <https://arxiv.org/pdf/2011.02412.pdf> ("the central missing foundation that digital democracy needs is digital personhood: an enforceable assurance that every real, natural human person may participate freely in digital democracy, expressing their true and uncoerced preferences in online governance, while exercising one and only one vote in online agenda-setting, deliberation, and decision-making.")

participants, undermining legitimacy, compromising protocol security and consensus, and eventually risking protocol collapse.

Idena was a small experiment in Proof of Personhood, peaking at 15,778 participants. Protocols like WorldCoin have onboarded more than 3,000,000 participants,⁶⁹ with a greater ambition to serve as an identity substrate for global democratic processes, including governance over AI and a distribution channel for AI-funded UBI.⁷⁰ WorldCoin's network effect arises from verifying unique individuals; with every new participant, the greater likelihood Worldcoin emerges as *the* frontrunner platform for global AI governance and UBI "windfall"⁷¹ distribution, if materialized.⁷² However, this network effect also undermines democratic ambitions, particularly when *one-person, one-vote* partially corrupts to *one-token, one-vote* through account trading or puppeteering. Because the incentive to join the network increases super-linearly as the network grows, new participants may lack alternatives but to join a partially corrupted network in order to have any influence over governance or to earn UBI. As non-participation becomes synonymous with socio-economic exclusion, participation risks becoming quasi-mandatory. Thus, new participants might find themselves ensnared in a dilemma: entering a system not out of genuine belief in the system's legitimacy but to counterbalance coordinated, plutocratic factions that will

⁶⁹ WorldCoin, "Introducing World ID 2.0;" Camille Bello, "Worldcoin: The Crypto Project Looking to Take on the World with Its Iris-Based ID Tech," *Euronews*, August 11, 2023, <https://www.euronews.com/next/2023/08/11/worldcoin-the-crypto-project-looking-to-take-on-the-world-with-its-iris-based-id-tech>.

⁷⁰ See Alex Blania and Sam Altman, "Introducing Worldcoin: A Letter from Alex Blania and Sam Altman," *Worldcoin Blog*, accessed December 3, 2023, <https://worldcoin.org/blog/worldcoin/introducing-worldcoin-alex-blania-sam-altman> ("If successful, we believe Worldcoin could drastically increase economic opportunity, scale a reliable solution for distinguishing humans from AI online while preserving privacy, enable global democratic processes and eventually show a potential path to AI-funded UBI.") See also "Humanness in the Age of AI," *Worldcoin Blog*, March 31, 2023, <https://worldcoin.org/blog/engineering/humanness-in-the-age-of-ai> ("[a]nother particularly important application is AI. To ensure that the benefits of AI are shared among all people, rather than being restricted to a privileged few, enabling inclusive participation in its governance is essential.") See also Sam Altman (@sama), *Twitter*, July 24, 2023, <https://twitter.com/sama/status/1683380242491260928> ("the goal is simple: a global financial and identity network based on proof of personhood. this feels especially important in the AI era. i'm hopeful worldcoin can contribute to conversations about how we share access, benefits, and governance of future AI systems.") Notably, it's unclear how Worldcoin intends to reconcile their token distribution (allocating 25% to insiders) with egalitarian ambitions. See Eliza Gkritsi, Oliver Knight, "Worldcoin's Tokenomics Shrouded in Mystery as Website is Reportedly Geo-Blocked Worldwide," *Coindesk*, July 24, 2023, <https://www.coindesk.com/business/2023/07/24/worldcoin-release-tokenomics-report-geofenced-for-some-countries>.

⁷¹ The "Windfall Clause" is a proposal for an *ex ante* commitment by AI firms to donate windfall profits, authored by several researchers some of whom work at leading AI labs. Cullen O'Keefe et al., "The Windfall Clause: Distributing the Benefits of AI for the Common Good," *Center for AI Governance*, January 30, 2020, <https://www.governance.ai/research-paper/the-windfall-clause-distributing-the-benefits-of-ai-for-the-common-good>. See also Dylan Matthews, "How 'windfall profits' from AI companies could fund a universal basic income," *Vox*, July 28, 2023, <https://www.vox.com/future-perfect/23810027/openai-artificial-intelligence-google-deepmind-anthropic-ai-universal-basic-income-meta>. For criticism, see Sam Shead, "Silicon Valley leaders think A.I. will one day fund free cash handouts. But experts aren't convinced," *CNBC*, March 30, 2021, <https://www.cnn.com/2021/03/30/silicon-valley-leaders-think-ai-will-fund-free-cash-handouts-experts-doubt-it.html>.

⁷² Although WorldCoin has a distributed token (\$WLD) with a traded market price, WorldCoin's WhitePaper also states, "the Worldcoin Protocol is not intended to generate profits to distribute UBI, and instead, it requires a separate funding source (e.g., a share of the profits generated by an AI Lab) to distribute global UBI." See "Limitations," *Worldcoin Whitepaper*. Notably, the founder of WorldCoin, Sam Altman, is also the founder of OpenAI, an AI company, and has independently advocated for UBI in the form of an "American Equity Fund," thereby generating cross-market effects and bolstering speculation that WorldCoin is in a position to credibly funnel "windfall" profits. Sam Altman, "Moore's Law for Everything," *Sam Altman's Blog*, March 16, 2021, <https://moores.samaltman.com/>.

otherwise dominate them. From an atomic perspective, individual participation may appear consensual,⁷³ but from a broader, network perspective, participation is more compelled as the network grows—the paradox of a “free but forced” Hobson’s choice.⁷⁴

In a winner-take-all race for global UBI and voting infrastructure, network effects mean that a *one-person, one-vote* network that corrupts to *one-token, one-vote* nonetheless has staying power. The dilemma in network legitimacy from partial corruption prompts us to consider broader democratic principles beyond *one-person, one-vote*. Wary of majoritarianism and faction, James Madison advocated for a bundle of mechanisms: separation of powers, checks and balances, federalism, bicameralism, and representation—to name a few. Combined, these checks would temper the tyrannous majorities and factions that might otherwise capture power to advance their narrow interest—or biases, transforming public goods into private goods.⁷⁵ The problem is Proof of Personhood only seeks to differentiate humans from *bots*, not their *biases*. And as Idena demonstrated, when given incentives to differentiate themselves from bots, humans also have incentives to align, control and puppeteer other humans *like* bots to amplify *their* biases. **As humans further integrate with information technology—even biologically with neural interfaces—the distinction between filtering humans from bots (*de jure* sybil resistance) and filtering humans *acting like* programmable bots (*de facto* sybil-resistance) will blur more, if not collapse, revealing a more foundational challenge than establishing biological uniqueness: establishing the *informational uniqueness of participants—or the extent to which they cluster with the same interests and biases*. This is the old problem of faction under new computational guises, but no longer constrained by geography—as in Madison’s days—but instead limited only by the breadth and depth of a digital interface.**

However, just as *one-person, one-vote* misses the essence of democratic governance—checking faction—similarly it would be a mistake to naively transpose systems of representation, separation of powers, and checks and balances to a global protocol. Before the 21st century’s digital age, democratic governance could rely on geography to roughly correlate with *biases*, or informational clusters.⁷⁶ Thus, systems of representation roughly

⁷³ For an overview of informed consent issues in biometric Proof of Personhood Protocols, see Eileen Guo and Adi Renaldi. “Deception, exploited workers, and cash handouts: How Worldcoin recruited its first half a million test users,” *MIT Technology Review*, April 6, 2022, <https://www.technologyreview.com/2022/04/06/1048981/worldcoin-cryptocurrency-biometrics-web3/>; see also Edd Gent, “Worldcoin Launched. Then Came the Backlash: The globe-spanning cryptocurrency and biometric identity project has agitated regulators,” *IEEE Spectrum*, August 28, 2023, <https://spectrum.ieee.org/worldcoin-2664361259>.

⁷⁴ See David Singh Grewel, *Network Power: The Social Dynamics of Globalization*, (New Haven: Yale University Press, 2008), p. 112, Kindle (“As a standard gains greater numbers of users, it passes the threshold of visibility and begins to exert network power. Given the existence of multiple powerful networks, the act of choosing one rather than another may indeed be wholly voluntary. But once we see the rise of a single dominant network—particularly if the threshold of inevitability has been passed or is on the collective horizon of expectations—the voluntariness of individual choice-making is increasingly eviscerated until all that remains is the individual’s ability to actively take up the one viable option that she faces. These circumstances of network power may be described as a version of Hobson’s Choice: an individual must either choose to use the dominant standard, or else choose not to conform, suffering social isolation and the loss of access to everyone pursuing the activity in question. As one standard overtakes another, the option to take up the dominant standard gradually becomes an offer that cannot be refused.”)

⁷⁵ “By a faction, I understand a number of citizens, whether amounting to a majority or a minority of the whole, who are united and actuated by some common impulse of passion, or of interest, adverse to the rights of other citizens, or to the permanent and aggregate interests of the community... There are two methods of curing the mischiefs of faction: the one, by removing its causes; the other, by controlling its effects.” See James Madison, “Federalist No. 10,” *The Federalist Papers*, The Avalon Project, Lillian Goldman Law Library, Yale Law School, (1787), https://avalon.law.yale.edu/18th_century/fed10.asp.

⁷⁶ Madison relied on the geographic expanse of the United States as a check on the influence of factions, a restraint that no longer exists in digital worlds; “Hence, it clearly appears, that the same advantage which a republic has over a democracy, in controlling the effects of faction, is enjoyed by a large over a small republic,—is enjoyed by the Union over the States composing it... The influence of factious leaders

surfaced diverse interests that could be reconciled and bridged across multiple perspectives (in Congress) and checked by other branches to yield policies in the “the public good.” But in the digital landscape, communication networks cut across geography, clustering biases and correlating beliefs and desires in unpredictable ways (e.g., attention auctions).⁷⁷ Participants immersed in correlated information channels are prone to aligning around the same biases, forming opaque and tacit majorities among otherwise geographically diverse participants—structurally similar to the problem of tacit monopolies in markets among otherwise seemingly diverse firms.⁷⁸ The challenge in democratic governance is to find new, computational ways to surface overlapping and recombining informational (or bias) clusters consensually *without* succumbing to a surveillance panopticon and *before* transposing analog systems of representation. In future work, we argue social identity systems rich in subsidiarity (e.g., federalism)—both physical and digital—are key to surfacing *bona fide* commitments and bias, challenging the utility of global identity protocols for democratic governance.

C. From Sybil-Resistance to Collusion-Resistance

“Proof of Personhood” seeks to offer a more egalitarian alternative to Proof of Work and Proof of Stake, where influence does not confer to participants who *already* have it by virtue of their wealth. This ambition has motivated sybil-resistance, where each unique human “*controls*” a corresponding unique account 1:1, and each account becomes a vehicle for distributing money (UBI) or votes. Yet, if an unintended consequence is humans

may kindle a flame within their particular States, but will be unable to spread a general conflagration through the other States.” Madison, “Federalist No. 10.” See also Danielle Allen and Justin Pottle, “Democratic Knowledge and the Problem of Faction,” *Knight Foundation White Paper Series*, (2018), https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/152/original/Topos_KF_White-Paper_Allen_V2.pdf (explaining how geography was key to Madison’s framework and ensured “epistemic pluralism.”) See also Jenna Bednar, “Polarization, Diversity, and Democratic Robustness,” edited by Simon Levin, *Proceedings of the National Academy of Sciences* 118, no. 50 (December 6, 2021) (arguing that democracy’s safeguards rely upon diversity, modularity, and redundancy, but Madison took that diversity for granted, which federalism may restore by opening the possibility space for cross-cutting cleavages and bridges essential to compromise in the wake of polarization).

⁷⁷ See Diego A. Martin, Jacob N. Shapiro, and Michelle Nedashkovskaya, “Recent Trends in Online Foreign Influence Efforts,” *Journal of Information Warfare* 18, no. 3 (Winter 2019): 15-48; Andrea Prat and Tommaso Valletti, “Attention Oligopoly,” *American Economic Journal: Microeconomics* 14, no. 3, (2022): 530-57; Jamie Doward, “The big tech backlash,” *The Guardian*, January 28, 2018, <https://www.theguardian.com/technology/2018/jan/28/tech-backlash-facebook-google-fake-news-business-monopoly-regulation>.

⁷⁸ In markets, Posner and Weyl highlight how firms can engage in tacit collusion forming tacit monopolies without direct communication or explicit agreements. Large institutional investors, such as asset managers, hold significant stakes across all rivals in an industry (e.g., every major airline) with the uniform incentive to exert shareholder pressure (*one-share, one-vote*) to cut salaries, cut R&D and discourage price competition, maximizing industry-wide profits at loss of competitiveness, innovation, & growth. This uniform anti-competitive pressure has the *appearance* of competition, but *behavior* of monopoly. Asset managers holding the *same* diversified portfolio across rivaling firms in the same sector effectively act as the *same* entity exerting uniform anti-competitive pressure, given their economic holdings. Eric A. Posner and E. Glen Weyl, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society* (Princeton: Princeton University Press, 2018). Whereas in markets the source of bias (or correlation) is the same economic interests that make seemingly diverse asset managers (and their portfolio companies) effectively act as one entity (tacit monopoly), the source of bias in democracy is broadly the overlapping informational sources that correlate individuals to act effectively as one entity (tacit majority). Both tacit monopoly and tacit majorities are cases of unintentional or accidental forms of collusion resulting from hidden correlation in control (ownership) and information, respectively.

being controlled by groups coordinating for more power and influence, sybil-resistance is *under-specified*.⁷⁹ The optics of a participant controlling their account *on-chain* account—for example through re-validation or re-authentication—ignores the *substance* of control: the sources of information that *influence* a participant to make a decision *off-chain* to, for example, cede their private keys in exchange for a paycheck, or delegate control to a 3rd party who makes choices to their detriment. A lesson from Idena’s experiment with Proof of Personhood is that whenever power is at stake—money or votes—myopically differentiating humans from bots (*de jure* sybil resistance) is to the detriment of overlooking the broader informational problem: differentiating humans acting *like* programmable bots (*de facto* sybil resistance). Just as information and control are two sides of the same coin, so are *de facto* and *de jure* sybil resistance.⁸⁰

Yet, *de facto* sybils (“programmed puppets”) is not a technical problem, but a *social* one. Human motivation is not *de novo*. Instead, beliefs and desires vary in correlation with whom we communicate information through a range of interactions from talking and trading—in short, our social ties.⁸¹ *De facto* sybils (“puppets”) are merely correlated in beliefs and desires to a socially tied 3rd party (“puppeteer”) who wields a corresponding information and control (or power) advantage over *them*, because they control more (e.g., channels, resources, status), have more information, or some combination of the two.⁸² Members of a puppeteering enterprise (“colluders”), on the other hand, are correlated with *each other* in beliefs and desires, wielding an information/control advantage over *others*. Thus, *de facto* sybils (puppets) are the natural objects of “colluders” (puppeteers), in the same way principals are the objects of agents. By extension, *de facto* sybil resistance is a mutually-implicated (or mirror) challenge to “collusion-resistance:” neither can be solved independently but both must be tackled simultaneously.

⁷⁹ Of the seven laws of identity, the first law is “[t]he system must be designed to put the user in control — of what digital identities are used, and what information is released.” Kim Cameron, “The Laws of Identity,” *Identity Blog*, May 2005, <https://www.identityblog.com/?p=352>. Yet the experiment in Idena reveals that user “control” is a nuanced question, as users (in the best case scenario) may delegate control for convenience, or in the worst case, may not *know* they are in control though they *appear* to be.

⁸⁰ See. n. 34. The term “Sybil” traces its origin to the book titled “Sybil,” which narrates the life of Shirley Mason who was reported to have developed sixteen distinct personalities as a result of child trauma. “Sybil” is relevant—and perhaps even more aptly suited—to the problem of *de facto* Sybil resistance (humans acting like bots, and in this case many personalities). Later, Shirley Mason admitted to these personalities being fake and intentional, drawing stronger parallels to *de jure* sybil resistance (deploying multiple personalities like bots), underscoring (ironically) the complexity and interrelationship between both concepts. Lynn Neary, “Real ‘Sybil’ Admits Multiple Personalities Were Fake,” *NPR*, October 20, 2011, <https://www.npr.org/2011/10/20/141514464/real-sybil-admits-multiple-personalities-were-fake>.

⁸¹ “Information is a name for the content of what is exchanged with the outer world as we adjust to it and make our adjustment felt upon it.” Norbert Wiener, *The Human Use of Human Beings* (Da Capo Press, 1988), Kindle edition, 17. For a discussion of the range of human communication and new possibilities unlocked by technology, see E. Glen Weyl, Audrey Tang, and Community, “Chapter 5: Collaborative Technology and Democracy,” in *PLURALITY: The Future of Collaborative Technology and Democracy*, (accessed February 27, 2024), <https://www.plurality.net>.

⁸² Stated alternatively, “[w]hat makes Sybil agents Sybils is that the will of one entity centrally coordinates them. They should be recognized as precisely the same because they all listen to that same entity and that entity alone. To take a rather extreme real-world example, then (and now starting to move rightward down the spectrum), we might think of a Sybil agent as similar to an individual who identifies very strongly with one specific group, and mostly coordinates their actions with the will of that group...Of course, the number of such groups an individual belongs to naturally varies and evolves...So as we move to the other end of the spectrum, individuals in fewer and fewer social groups begin to look more and more like the type of self-interested agents that economists usually put into their models – i.e., the *homo economicus*.” Joel Miller, E. Glen Weyl, and Leon Erichsen, “Beyond Collusion Resistance: Leveraging Social Information for Plural Funding and Voting,” *SSRN* (2022), December 7, 2022, <https://ssrn.com/abstract=4311507>.

Paradoxically, while social groups are the root of “collusion” (and *de facto* sybils) they also are the remedy. A naive conclusion would be to stomp out “collusion” and remove all asymmetries to make participants informationally the same, only to end in the worst power asymmetry: surveillance. Rather than control the causes of collusion, a better approach is to check its effects, or excesses.⁸³ Specifically, if the excesses are capture, then a remedy starts with checking groups from over-influencing (or “programming”)—often innocently, accidentally and weakly—other participants to the detriment of capturing public goods for their narrow interests.⁸⁴ One way is to consensually expand social ties, broadening the set of conversations and by extension beliefs and desires which may motivate a participant. Another is to reconcile diverse, informationally unique perspectives to find broader, shared public goods. In this way, the goal of collusion-resistance (and its mirror problem of *de facto* sybil resistance) dovetails with the goal of democratic governance: checking faction.

How do we temper collusion and awaken *de facto* sybils without global surveillance and gaming? For now, we leave readers with a sketch of “plural attention mechanisms” and explore their tools and conditions (e.g., subsidiarity, privacy as contextual integrity, social identity) in future work. One mechanism already emphasized is *consensus across difference*, which elevates agreed proposals by participants (or informational clusters) who otherwise generally disagree—a signal that a proposal is more likely to be bridged across divergent interests, and therefore in the broader public good. Another is *peer prediction*, which surfaces expertise and elevates truth. Combined, both mechanisms enable participants to direct their attention to ideas and policies grounded in truth, with broad-based support and eschew narrow policies biased towards special interests (mitigating “collusion”).⁸⁵ At the same time, these mechanisms encourage puppets to “awaken” with novel information. Because proposals endorsed by adversarial groups receive more attention and ascend to prominence, participants have an incentive to bridge social distance, or “cross long bridges,” to find new points of consensus among groups with whom they might typically disagree.⁸⁶ Such atypical conversations yield novel information which, in turn, fosters social recombination, the formation of new groups and solidarities, and ultimately increases the *cost of influencing* a participant (mitigating *de facto* sybils).⁸⁷ Thus, bridging (or anti-correlation) mechanisms that check faction coupled with peer prediction that

⁸³ “There are again two methods of removing the causes of faction: the one, by destroying the liberty which is essential to its existence; the other, by giving to every citizen the same opinions, the same passions, and the same interests.” James Madison, “Federalist No. 10.”

⁸⁴ For mathematical approximations of collusion-resistance, see Miller, Weyl, and Erichsen, “Beyond Collusion Resistance.” In the context of Idena, one example of “weak” or accidental collusion is the validation ceremony’s timing which biases network participation along a longitudinal time-zone. This in turn may correlate with another set of biases in governance (e.g., partialities to compliance to a set of norms or laws over others).

⁸⁵ As Bryan Ford acknowledges, “[p]eople need reliable information sources protected from both subversion through “fake news” and polarization through automated overpersonalization” and at the same time, “any approach to information filtering and selection runs into the fundamental problem of accounting (or not) for expertise.” Bryan Ford, “Technologizing Democracy or Democratizing Technology? A Layered-Architecture Perspective on Potentials and Challenges” in *Digital Democracy and Democratic Theory*, ed. Lucy Bernholz, Hélène Landemore, and Rob Reich (Chicago: University of Chicago Press, 2021), 275, 281. In future work, we argue a combination of anti-correlation (bridging bonuses) and peer prediction interleaved with deliberation is a healthy step towards improving the information environment of participants.

⁸⁶ We thank Audrey Tang for articulating “bridging bonuses” as a positive expression to “anti-correlation.”

⁸⁷ Paradoxically, individuality increases with diverse sociality (or intersectionality). Georg Simmel, *Conflict & The Web of Group-Affiliations* (New York: The Free Press, 1995), 151 (“The individual may add affiliations with new groups to the single affiliation which has hitherto influenced him in a pervasive and one-sided manner. The mere fact that he does so is sufficient, quite apart from the nature of the groups involved, to give him a stronger awareness of individuality in general, and at least to counteract the tendency of taking his initial group’s affiliations for granted.) Applied to the context of Proof of Personhood, if the goal of is to filter “fake” from

elevate truth interweave solidarities across distance. **Conversational webs thicken in a virtuous cycle of social recombination and deepening informational diversification.**⁸⁸ **Instead of reinforcing old divisions, power is ceaselessly cut across new lines, revealing and reconciling adversarial interests, reconfiguring old asymmetries, carving new cleavages, and incrementally building more adaptive, cooperative networks resistant to collapse.**⁸⁹

D. Dark DAOs & Voting Security

Pools—whether cooperative or puppeteered—were an instance of off-chain coordination easily detected through chain-analysis and then later through economic incentives for delegation. However, in the future, coordination could also go undetected, or “dark;”⁹⁰ for example, in a trusted execution environment, a participant running their own node could auction off their vote and prove to a 3rd party (briber or coercer) they have voted in a certain way without knowing how they were deputized to vote.⁹¹ This has spurred innovation towards *ex-post* measures, such as “receipt-freeness,” which thwart a participant from proving to a 3rd party that they voted a certain way, even if they want to. Similarly, and more recently, it has motivated *ex-ante* measures, such as Proofs of Complete Knowledge⁹² that rule out partial or whole key encumbrances before a vote occurs; the rationale is that if

“authentic” accounts, the threshold for “fake” moves from attestations from a verification method (whether biometric, cognitive, or otherwise) to a constellation of uncorrelated attestations from participants who are unlikely to be talking to each other.

⁸⁸ A participant generally isn’t either/or, but both a “puppet” and a “colluder,” depending on context. Social recombination increases this nuance and complexity. For example, within a group, a participant may be “low-information,” but that group might be “high-information” relative to other groups. Whether the group colloquially is “cooperative” (positive) or “collusive” (negative) also depends on whether the group aligns with the observer’s interests, which changes over time. Zooming out, participants might believe the system is largely competing collusive groups, when in reality the system is rife with puppeteering (like in surveillance states); here, participants are “collusive” relative to each other, but uninformed “puppets” in the broader context. As people talk and trade, however, new channels form, groups mix and recombine. Colluders may recruit sybil puppets into their group, or eventually themselves turn into unwitting puppets after ceding control without accountability. Conversation across distance surfaces adversarial information to encourage recombination and reconfigure asymmetries. See n. 50.

⁸⁹ Whereas theories of cooperation have focused on “enlarging the shadow of the future” with more frequent, durable interactions, we also emphasize “crossing long bridges.” Robert Axelrod, *The Evolution of Cooperation: Revised Edition* (Cambridge, MA: Perseus Books, 2006), 52. With social recombination, players interact on multiple games, on many dimensions (requiring us to acknowledge time and social discounts). Multiple games and social recombination is also a check against polarization, which compresses information and hinders compromise necessary for cooperation. See Bednar, “Polarization, Diversity, and Democratic Robustness” (“Polarization does more than divide a population; it simplifies it. When people conform to one another, they let go of their differences...not only are groups becoming more sharply divided, but they do so by becoming more like those in their in-group. At the system level, information is lost. When agents conform, the dimensionality of the issue or identity space is reduced.”)

⁹⁰ Researchers have emphasized how on-chain voting is vulnerable to actors buying on-chain votes (at best) or coercing participants (at worst), and furthermore concealing their coordination by using smart contracts, trusted hardware, off-chain organizations (“Dark DAOs”)—or some combination of all three. Given that trusted hardware is a substrate for more sophisticated, undetectable attack, researchers have argued that “the only defense from this is more trusted hardware: to know a user has access to their own key material (and therefore cannot be coerced or bribed), some assurance is required that the user has seen their key.” Philip Daian, Tyler Kell, Ian Miers, and Ari Juels, “On-Chain Vote Buying and the Rise of Dark DAOs,” *Hacking, Distributed*, July 02, 2018, <http://hackingdistributed.com/2018/07/02/on-chain-vote-buying/>.

⁹¹ A corollary phenomena in traditional finance is “empty voting,” where voting rights are decoupled from economic rights, breaking the assumption of one-share, one-vote. See Henry T. C. Hu and Bernard Black, “Empty Voting and Hidden (Morphable) Ownership: Taxonomy, Implications, and Reforms,” *The Business Lawyer* 61, no. 3 (May 2006): 1011-1070.

⁹² “Complete knowledge by a prover P of a secret sk means, informally, that it has unencumbered access to sk and thus can use it for any desired purpose, e.g., can sign any message of her choice.” For example, “Alice might encumber her voting key sk so that she can only sign

a participant can furnish a proof that they can use their private key in whatever way they want, by extension, they cannot furnish a proof of encumbrance to a 3rd party, thereby undermining their ability to credibly sell their voting right (or vote in a certain way without knowledge).⁹³

While salutary advancements in voting security, Idena's experiment also highlights limitations to exclusively *on-chain*, technical approaches. Specifically, a Proof of Complete Knowledge may establish that *someone* has direct access to a private key, but doesn't guarantee that *the intended* or designated participant does. Notably, at least ~40% of Idena's network accounts had evidence of 3rd party key access by pool operators, where the most plausible explanation was puppeteering:

- Some operators controlled keys wholly from the outset at the registration phase, masking the existence of private keys and reimbursing participant's for their time in periodically performing validation ceremonies (strong puppeteering).
- Other participants "knew" their key material, but did not understand their significance and unwittingly rented them (and their vote) in exchange for a paycheck from operators (semi-strong puppeteering).

Thwarting *on-chain* vote-buying doesn't solve for *off-chain* vote-buying into "meatspace," and may encourage it as a low-cost alternative. Idena undermined account trading (buying and selling accounts) with identity staking—a mechanism similar to voting deposits in MACI.⁹⁴ The next best alternative for off-chain vote-buying became puppeteering (buying participant time). Both cases—account trading and puppeteering—present a security conundrum: participants may sell, rent, or unwittingly share their unencumbered keys *off-chain* with 3rd parties, thereby undermining measures to keep keys unencumbered *on-chain*. In future work, we explore how the social layer (off-chain social encumbrances) may be harnessed to reinforce (rather than undermine) on-chain security and define the scope of *legitimate* encumbrances.

V. CONCLUSION

In the wake of advancing frontier models, experiments in Proof of Personhood to achieve *one-person, one-vote, one-reward* have focused on biological uniqueness to the detriment of overlooking the underlying social reality that people talk and trade. Even if biologically verified, people are not informationally the *same*. As Idena's

a ballot with candidate Bob, the choice of adversary Mallory. Alice can then sell Mallory an enforceable promise that if she votes, she will vote only for Bob...Here, CK would imply that Alice can always cast any desired vote and therefore, cannot sell Mallory an enforceable promise to vote only for Bob." Mahimna Kelkar, Kushal Babel, Philip Daian, James Austgen, Vitalik Buterin, and Ari Juels, "Complete Knowledge: Preventing Encumbrance of Cryptographic Secrets," *Cryptology ePrint Archive* (2023), accessed January 15, 2024, <https://www.cs.cornell.edu/~babel/papers/ck.pdf>.

⁹³ For illustrations, see James Austgen, Kushal Babel, Vitalik Buterin, Phil Daian, Ari Juels, and Mahimna Kelkar, "Complete Knowledge," *IC3 Medium*, January 16, 2023, <https://medium.com/initc3org/complete-knowledge-eecdda172a81>.

⁹⁴ Identity staking is similar to a deposit in MACI (Minimal Anti-Collusion Infrastructure), which requires a participant to put a deposit in order to vote—a deposit which would be stolen by a 3rd party if they also had key access. "Each user is also expected to put down a deposit; if anyone publishes a signature of their own address with the private key, they can steal the deposit and cause the account to be removed from the list (this feature is there to heavily discourage giving any third party access to the key)." Vitalik Buterin, "Minimal Anti-Collusion Infrastructure," *EthResearch Blog*, May 2019, <https://ethresear.ch/t/minimal-anti-collusion-infrastructure/5413>.

experiment showed, when presented with economic incentives to differentiate themselves from bots, people also have incentives to align and control the less informed *like* programmable bots. Groups quickly coalesce and coordinate to their advantage for more control and economic rewards, undermining the egalitarian goals with a trendline towards oligopoly.

Yet, quashing social coordination to make *one-person, one-vote, one-reward* “work” would be a greater folly, ending in greater harm. Both “cooperative” and “puppeteered” groups emerged in Idena’s experiment to take advantage of economies of scale. But the difference was not of kind but *degree*, depending on who captured the benefit of the bargain: the principal or agent, the puppet or puppeteer. Given this spectrum and the complexity of social reality (people delegate control and groups mix, recombine, nest, and dissolve), quashing “bad” puppeteering to keep the “good” kind of cooperation would be naively fraught. Enforcement would require either total surveillance to decide what is “good” or “bad” communication, or simply manipulating communication until all participants are informationally the same, leaving no differences to coordinate around. Both correctives conjure failed experiments in 20th century communism, where the dogged pursuit of equality and “perfect enforcement” progressively centralized power until social and economic collapse. Rather than ignore social ties (and succumb to oligopoly), or try to control or eliminate them (and succumb to totalitarian uniformity), a better approach is to presume them as a starting point.

Within the pursuit of egalitarianism, the choice is two-fold: level the playing field, or *expand* it. Experiments in Proof of Personhood that compress identity to one global game opt for the former, seeking to brute force *one-person, one-vote, one-reward*. A better alternative is to *expand* into a plurality of identity games as broad, deep, and infinite as the combinatorial possibilities of human association that arise from talking and trading—the sources of our informational differences. When identity is expressed as a dynamic constellation of memberships to groups, participants reveal different, *partial* aspects of their identity in communication, depending on how many dimensions they are associated. No two participants share the same perspective and a participant’s identity does not neatly reduce to being a “puppet” or a “colluder;” participants can be both, neither, and either, depending on the partial perspective of the observer. Similarly, nested groups (*individuals within groups, groups within pools, pools within megapools, and protocols within other systems*), invites a more networked conception of “cooperation,” “consent,” “corruption,” “collusion” and “coercion”—depending on *the differentials* in information and control (or power) of underlying groups across different systems. **With many identity games—instead of just one—even if all groups succumb internally to oligopoly, groups may overlap, intersect, and recombine, canceling power out. Thus, whereas a single identity game has inevitable economies of scale towards one global oligopoly, a plurality of games open the possibility space for a normal (Gaussian) distribution of power, achieved through diversity, not brute-forced equality.**

APPENDIX A

Methodology

This study triangulates information from a variety of sources: blockchain transactions, community conversations on Discord and Telegram, and the protocol's website. For blockchain transactions, we sourced data collected by Idena's open-source indexer⁹⁵ which provides transaction data in human-readable format, subsequently storing it in an SQL database, accessible via an open API⁹⁶ and also available through Idena's blockchain explorer.⁹⁷ Pools were discussed on community Telegram and Discord channels and conversations with several pool operators occurred through private Telegram messages with the Idena Team.⁹⁸

This study has two phases: *pre-delegation* starting August 2019, and *post-delegation*, from March 2021 to May 2022. Pre-delegation, Idena's indexer database is limited to account data (e.g., transactions, wallet balance, identity stake, identity status, age, issued invitations, generated flip tests, earned validation and mining rewards).⁹⁹ We did not conduct formal chain-analysis examining transfers between all accounts to gauge the extent of pools. Instead, our window into pools was limited to anecdotal community and private conversations on Telegram and Discord, corroborated by patterns of one-way transactions on Idena's Blockchain Explorer.

Post-delegation, accounts could group under one pool account to piggyback on their node, making pools, their size, and members visible on-chain. To examine 3rd party key access (i.e. shared private keys) of accounts within pools, we formulated a criterion (elaborated in Appendix B) and manually examined 31 pools that had grown to more than 100 accounts at any point in the protocol's history using the Idena Block Explorer. As a second step, we also examined ties between these pools by way of financial transfers. We did not systematically query all addresses or keywords in community channels, nor conduct systemic chain analysis of all pools with less than 100 accounts. We welcome readers skilled in chain analysis to continue this research effort and offer suggestions below in Appendix B.

⁹⁵ Idena, "Idena Indexer GitHub Repository," GitHub, accessed on February 3, 2024, <https://github.com/idena-network/idena-indexer>.

⁹⁶ Idena, "API Endpoint for Idena Blockchain Indexer," accessed on February 3, 2024, <https://api.idena.io>.

⁹⁷ Idena, "Idena Blockchain Explorer," accessed on February 3, 2024, <https://scan.idena.io>.

⁹⁸ Idena, "Community Discord Server," accessed on February 3, 2024, <https://discord.gg/8BusRj7>; Idena, "Idena Network," Telegram, accessed February 3, 2024, <https://t.me/IdenaNetworkPublic>.

⁹⁹ Data is collected every epoch (1-3 weeks). Thus, charts are stepped, with interval width representing the epoch's duration (e.g., Figures 8-11).

APPENDIX B

Pool Analysis

Examining the 31 top pools alone (without acknowledging financial ties), accounts appeared evenly distributed, with the exception of 4 pools which held ~34% of top-31 pooled accounts.

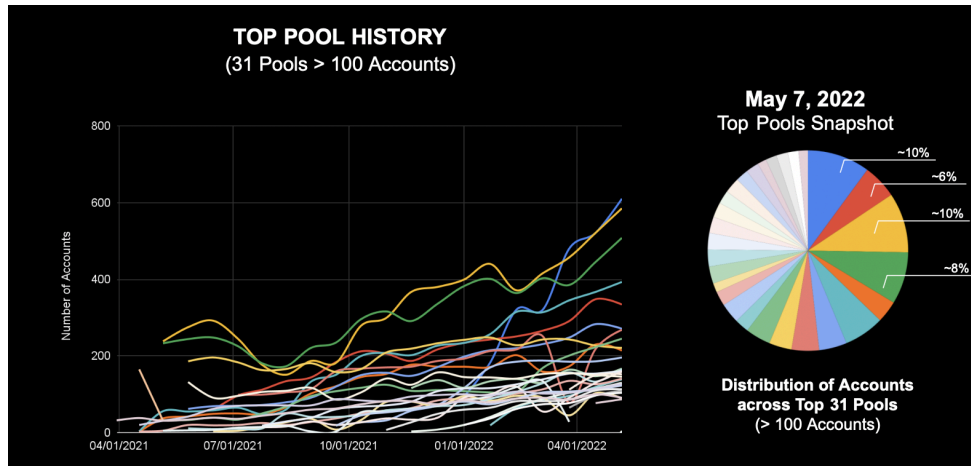


Figure B.1 History of top pools (>100 accounts) & May 2022 snapshot showing the distribution of accounts across these top pools

Acknowledging collusive ties among the top 31 pools, the chart dramatically shifts. The below chart compares the top 31 pools to the 23 distinct entities with an extended timeline, from protocol launch to August 2023.

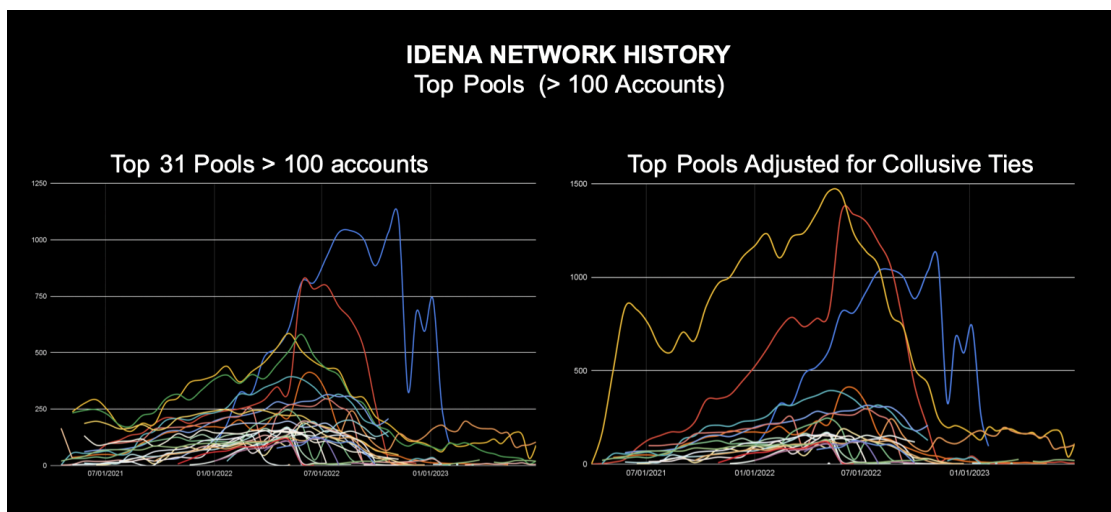


Figure B.2 History of Idena Pools from Launch to August 2023

The 31 pools below were analyzed. The top 3 puppeteering networks (evidenced by transfers) were Russian (yellow), Indonesian (blue), and a network of unknown social origins (red).

	Address
1	0xDDDDaDDB856901ac3e2251b8234EfeaB2188b22A
2	0xDDDDcFdCC512FacD27038BA958742E81e2982cB
3	0x00189F2C6aF07C4f2E6b19b4E818Dc98C9C1Ed0b
4	0xdddddFF74aA78A09ca272054aF01e1bdf22D8A96
5	0x69E19034AA69B2974a3c73F648d335F871c7a9f0F
6	0x96d11da40FDc82D81ebE0EAE61bFe6a47F43d1a6
7	0x17b851A11f7d37054928BEf47F0F22166d433917
8	0x42ABa502E387431daA759E411C6459Fc3b90Bb02
9	0x8D68647a9A32cde4bc7aDDE7E04D5c4a69190668
10	0x21bde8AecC6c99400B8DAA4b893e2434B3D985d0
11	0x83dd8f0d90Fea4C626347c8b00ebF245231fA7F8
12	0x0094B02Dc78a31fF76D9b3a9C1B066299a120d3C
13	0x81877420959D9EE656E4A0Daf7327806be5cB84C
14	0x71182839FD4eA1dA7615D7a499Ba9a3DaFc0FE8a
15	0xce7E25f6e5cD11FbCdB9d7bf4F56409e55Be4B99
16	0x2EbbBfDc6d71D1EebEB61D4FD6367712C7dB575e4
17	0xe5183950ab39223252cE3990118733f80DaB39b6
18	0xDfF5bEB9Cc914372338CbD78d791ae24c94AA238
19	0x4e234e027Ecc83b55286b4d50840A989C809Ac4b
20	0xdd2eb98E1652950dfE1adD1583262E1daD24919b
21	0xB3C3cD8484C7D5d0533f97Ea7Cc0d717760e4E02
22	0x3a39aaf7B7E607B6CBC906AECF6F691376700a02
23	0x63f47c8b614A3b8d1e142755D4019C6411Cbc210
24	0x886Cd1c0E0df1c20a75db7DF3369d5EBE2a53Ae9
25	0x2CdB324821E191561Eab4f0dAfe16a016dfA53db
26	0xee2697e512c90Da18281fA7613Cb102efb72D810
27	0x8dFe01c1B2Ba75d32990E212B28bCd144E7e2015
28	0x648C3EB62b8835bC59ed86802461e5b1691a6C81
29	0xd34b4Bb3039dDA9b4ECB8e96109DC2Df044c3C47
30	0x7f32D1449e83c000a04ff624C2a8A0b006507D1e
31	0x664d30B74a1876D8f846A00E56916129229cd1b4

The most important sign of puppeteering was 3rd party key control, evidenced by the unlikely coincidence of *simultaneous* or *sequential* account transactions in the same pool, including “account delegations” to the same pool or “account terminations” from the same pool. *All pools* showed this transaction pattern. The other tell-tale signs of 3rd party key control were transactions patterns funneling rewards to pool operators, including:

- a pool operator receiving all identity stake *after* the account was terminated (by the account through a “terminate identity” transaction) or by the delegator (through “kill delegator” transaction). *All pools* showed signs of this.
- pooled accounts sending all transferable rewards earned as a solo account *before* delegation to the pool operator (presumably when the pool operator was not providing any value or service). All pools showed this, except for pool 31 (“Egyptian Pharaoh”), which had a mix of accounts and pool 9, which showed aspects of this but was part of a large complex network and composed of many accounts, meriting professional chainanalysis.
- a pool-operator withholding transferable rewards, rather than distributing them back to member wallets, and eventually sending them to a hive wallet, or an exchange. All pools showed this, except for pool 31 (“Egyptian Pharaoh”) and pool 9, where accounts sometimes received rewards only to send them back to the operator or to another account part of a complex network, meriting further chainanalysis.

Russian Pool (yellow)

The largest Puppeteer was a Russian operator, controlling 5 pools. This operator was the most vocal, engaging in discussion with the Idena Team as well as the community on Telegram (see Figs. 3-5 & Fig. 7). Before delegation, the operator revealed themselves to be running the first “human pool.” After delegation, the operator revealed their pools’ wallet address prefixes (“0xddd”) in the community discord. Two pools simultaneously emerged with ~250 accounts at the start of delegation in May 2021, peaking to over 500 accounts each in May 2022. The pooled operator generally sent funds to exchange¹⁰⁰ and a treasury wallet,¹⁰¹ though contributing some identity stake to a handful of accounts.

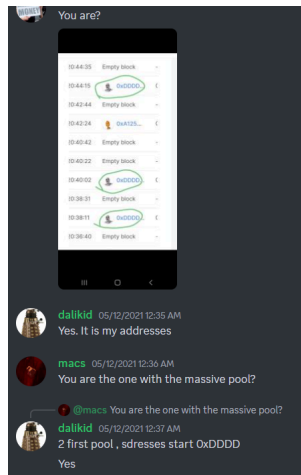


Figure B.3 Message from a Russian puppeteer in Community Discord (May 2021)

Indonesian Pool (blue)

Another pool (color blue) emerged in January 2022, having a meteoric rise and collapse with over 1400 accounts within a year, consistent with the claims of the Indonesian operator who claimed to pay “workers.”¹⁰² This pool also did not distribute back to member accounts but transferred funds to an exchange.¹⁰³ This pool died after IIP-6&7, terminating by sending their identity stake and transferable wallet balances to an exchange wallet.

¹⁰⁰ BSC bridge address to the Pancackswap “Address 0x98D16d7021930b788135dD834983394fF2De9869 - Idena Blockchain Explorer,” Idena Explorer, accessed December 3, 2023, <https://scan.idena.io/address/0x98D16d7021930b788135dD834983394fF2De9869>

¹⁰¹ “Address 0xDDDD06adBF37d5F7997E61e410d567DDC56AE79E - Idena Blockchain Explorer,” Idena Explorer, accessed December 3, 2023, <https://scan.idena.io/address/0xDDDD06adBF37d5F7997E61e410d567DDC56AE79E>

¹⁰² Indonesian pool address: “Pool 0x96d11da40FDe82D81ebE0EAE61bFe6a47F43d1a6 - Idena Blockchain Explorer,” Idena Explorer, accessed December 3, 2023, <https://scan.idena.io/pool/0x96d11da40FDe82D81ebE0EAE61bFe6a47F43d1a6#sizeHistory>. Hive wallet address: “Address 0xC7064Bb35B581E6ed200839303b6394Cb831a99C - Idena Blockchain Explorer,” Idena Explorer, accessed December 3, 2023, <https://scan.idena.io/address/0xC7064Bb35B581E6ed200839303b6394Cb831a99C> .

¹⁰³ Pool operator’s Bitmart deposit address 0xC90366C86bF072e426d2f89968C979558bBFe1b6

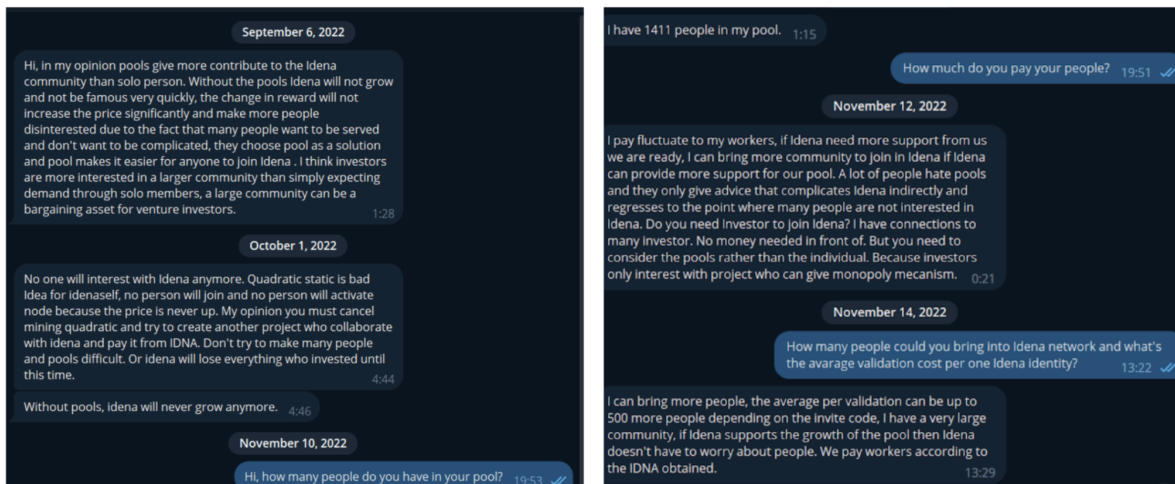


Figure B.4 Message from Indonesian operator to Idena Telegram Administrator when the community was debating quadratic staking (Nov 2022)

Unknown Network (red)

A third network of 5 pools emerged gradually at the start of delegation, peaking to 1280 accounts in June 2022. Whereas other networks had telegram chats, the origins and social ties of these pools were unknown. In addition to signs of puppeteering, the rapid addition of over 500 accounts in May 2022 (epoch 86) was also suggestive. This network requires deeper chain-analysis, to flesh out the flow of funds between pools and possible ties to other pools, including pools with less than 100 accounts. This network eventually died after IIP-5 (experiment in sublinear identity staking).

Egyptian Pool

While this was not a top pool, we discuss it given the photos of child puppeteering (see Figure 7). The maximum number of accounts for this pool was 100 in April 2022, and the pool oddly showed signs of both puppeteering and consensual participation. On the one hand, there were batched delegations (e.g., Feb 2022), suggesting control over private keys. At other times, delegations were at different times, suggesting lack of control. Moreover, at times there were outgoing transactions of the same amount (Feb 2022), suggesting a return of mining rewards to accounts. Yet at other times (e.g., March 2022), different amounts were sent to different wallets, suggesting puppeteering as mining rewards were issued on a per account basis and should have been equally distributed. This pool merits further analysis.¹⁰⁴

¹⁰⁴ See “Pool Address 0x664d30B74a1876D8f846A00E56916129229cd1b4 - Idena Blockchain Explorer,” Idena Explorer, accessed December 3, 2023, <https://scan.idena.io/pool/0x664d30B74a1876D8f846A00E56916129229cd1b4>. For discussion, see <https://discord.com/channels/634481767352369162/634481767843364866/930073156671139901>, Discord, accessed December 3, 2023. For video, see https://cdn.discordapp.com/attachments/634481767843364866/930122162763874371/VID_20210918_111346_196.mp4, Discord, accessed December 3, 2023.

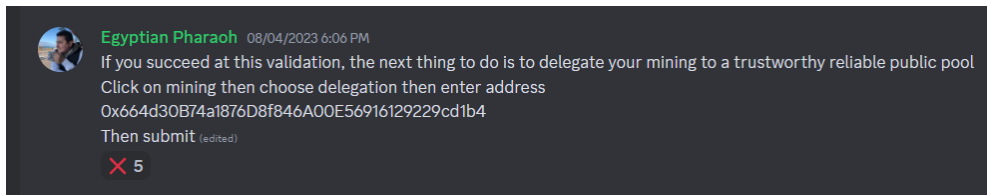


Figure B.5 Message from Egyptian Pool operator in Community Discord (2023)

Future pool analysis

In this paper, we offer a surface overview of the largest pools. There are many open questions on pool dynamics that would require deeper chain analysis, including:

- **Pool funnel** (how often solo accounts became member pool accounts) & reverse-migration (how often pooled accounts became solo accounts)
- **Pool shopping** (how often solo accounts “pool shopped,” jumping between pools)
- **Pool churn & lifespan** (how frequent pools dissolved into solo accounts, or recombined into new pools)
- **Termination rates** as a result of pool operators v. validation ceremony failures
- **Fund travel** (the proportion of funds that left pools for exchanges, cold storage, new accounts, revived accounts, old pools and new pools)
- **Seizure risk** (how often identity stake was seized by pool operators)
- **Nested networks** of pools and cooperation with other pools (deep chainanalysis)

But chainanalysis also has its limits. It is costly in time and money, so protocols are not easily auditable by participants. And it reveals only the changing distributions of rewards and stake across accounts and pools, without revealing substantive off-chain economic arrangements and the kind of participants (puppets, agents, families, friends, puppeteers, companies) behind accounts. Puppeteers, for example, could control and operate a pool of undelegated solo accounts that appear to be independent, especially if they stand to benefit with more voting power. In future work, we propose identity systems which seek to avert these weaknesses and computational costs without a surveillance god's-eye-view.

SELECTED BIBLIOGRAPHY

BOOKS

- Allen, Danielle. *Justice by Means of Democracy*. Chicago: University of Chicago Press, 2023.
- Axelrod, Robert. *The Evolution of Cooperation: Revised Edition*. Cambridge, MA: Perseus Books, 2006.
- Chwe, Michael Suk-Young. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton: Princeton University Press, 2001.
- Deutscher, Isaac. *Stalin: A Political Biography*. New York: Oxford University Press, 1967.
- Dimitrov, Martin K. *Dictatorship and Information: Authoritarian Regime Resilience in Communist Europe and China*. Oxford: Oxford University Press, 2023.
- Easley, David and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge: Cambridge University Press, 2010.
- Grewel, David Singh. *Network Power: The Social Dynamics of Globalization*. New Haven: Yale University Press, 2008.
- Kotkin, Stephen. *Stalin: Paradoxes of Power, 1878-1928*. New York: Penguin Books, Illustrated Edition, 2015.
- Ostrom, Elinor. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990.
- Posner, Eric A., and E. Glen Weyl. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press, 2018. <https://doi.org/10.2307/j.ctvc77c4f>.
- Posner, Richard A. *Economic Analysis of Law*, 7th ed. Austin, Texas: Aspen Publishers, 2007.
- Simmel, Georg. *Conflict & The Web of Group-Affiliations*. New York: The Free Press, 1995.
- Tsebelis, George. *Nested Games: Rational Choice in Comparative Politics*. Berkeley: University of California Press, 1990.
- Weyl, E. Glen, Audrey Tang, and Community. "Collaborative Technology and Democracy." In *PLURALITY: The Future of Collaborative Technology and Democracy*. Accessed February 27, 2024. <https://www.plurality.net>.
- Wiener, Norbert. *Cybernetics: Or Control and Communication in the Animal and the Machine*. Cambridge, MA: The MIT Press, 1948.

Wiener, Norbert. *The Human Use of Human Beings*. Da Capo Press, 1988.

ARTICLES

Ahn, Luis von, Manuel Blum, Nicholas J. Hopper, and John Langford. "CAPTCHA: Using Hard AI Problems for Security." *Advances in Cryptology—EUROCRYPT* (2003): 294—311. https://doi.org/10.1007/3-540-39200-9_18.

Allen, Danielle and Justin Pottle. "Democratic Knowledge and the Problem of Faction." *Knight Foundation White Paper Series*, (2018). https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/152/original/Topos_KF_White-Paper_Allen_V2.pdf.

Bednar, Jenna. "Polarization, Diversity, and Democratic Robustness." Edited by Simon Levin. *Proceedings of the National Academy of Sciences* 118, no. 50 (December 6, 2021):e2113843118. <https://doi.org/10.1073/pnas.2113843118>.

Borge, Maria, Eleftherios Kokoris-Kogias, Philipp Jovanovic, Linus Gasser, Nicolas Gailly, and Bryan Ford. "Proof-of-Personhood: Redemocratizing Permissionless Cryptocurrencies." *2017 IEEE European Symposium on Security and Privacy Workshops*, pp. 23-26, (2017). <https://doi.org/10.1109/EuroSPW.2017.46>.

Douceur, J.R. "The Sybil Attack." In *Peer-to-Peer Systems*, edited by P. Druschel, F. Kaashoek, and A. Rowstron, 2429:251–260. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2002. https://doi.org/10.1007/3-540-45748-8_24.

Ford, Bryan. "Technologizing Democracy or Democratizing Technology? A Layered-Architecture Perspective on Potentials and Challenges." In *Digital Democracy and Democratic Theory*, edited by Lucy Bernholz, Hélène Landemore, and Rob Reich, 274-308. Chicago: University of Chicago Press, 2021.

Ford, Bryan. "Identity and Personhood in Digital Democracy: Evaluating Inclusion, Equality, Security, and Privacy in Pseudonym Parties and Other Proofs of Personhood." Swiss Federal Institute of Technology in Lausanne (EPFL), (2020). <https://arxiv.org/pdf/2011.02412.pdf>.

Granovetter, Mark. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91, no. 3 (November 1985): 481-510.

Granovetter, Mark. "The Strength of Weak Ties." *American Journal of Sociology* 78, no. 6 (May 1973): 1360-1380.

Hu, Henry and Bernard Black. "Empty Voting and Hidden (Morphable) Ownership: Taxonomy, Implications, and Reforms." *The Business Lawyer* 61, no. 3 (May 2006): 1011-1070.

Kelkar, Mahimna, Kushal Babel, Philip Daian, James Austgen, Vitalik Buterin, and Ari Juels. "Complete Knowledge: Preventing Encumbrance of Cryptographic Secrets." *Cryptology ePrint Archive*, Paper 2023/044, 2023. <https://eprint.iacr.org/2023/044>.

Kim, Minjae and Roberto M. Fernandez. "What Makes Weak Ties Strong?" *Annual Review of Sociology* 49 (July 2023): 177-193.

Prat, Andrea, and Tommaso Valletti. "Attention Oligopoly." *American Economic Journal: Microeconomics* 14, no. 3, (2022): 530-57.

Le Blond, Stevens, Alejandro Cuevas, Juan Ramón Troncoso-Pastoriza, Philipp Jovanovic, Bryan Ford, and Jean-Pierre Hubaux. "On Enforcing the Digital Immunity of a Large Humanitarian Organization." Paper presented at the 2018 IEEE Symposium on Security and Privacy (SP), May 2018. DOI:10.1109/SP.2018.00019.

Madison, James. "Federalist No. 10." The Avalon Project, Lillian Goldman Law Library, Yale Law School, 1787. https://avalon.law.yale.edu/18th_century/fed10.asp.

Martin, Diego A., Jacob N. Shapiro, and Michelle Nedashkovskaya. "Recent Trends in Online Foreign Influence Efforts." *Journal of Information Warfare* 18, no. 3 (2019): 15-48. <https://www.jstor.org/stable/26894680>.

Mazorra, Bruno, and Nicolás Della Penna. "The Cost of Sybils, Credible Commitments, and False-Name Proof Mechanisms." Preprint, submitted January 2023. <https://doi.org/10.48550/arXiv.2301.12813>

Miller, Joel, Eric Glen Weyl, and Leon Erichsen. "Beyond Collusion Resistance: Leveraging Social Information for Plural Funding and Voting." SSRN, 2022. <https://ssrn.com/abstract=4311507>.

Nakamoto, Satoshi. "Bitcoin: A Peer-to-Peer Electronic Cash System." 2008. <https://bitcoin.org/bitcoin.pdf>.

Ohlhaber, Puja, E. Glen Weyl, and Vitalik Buterin. "Decentralized Society: Finding Web3's Soul." SSRN, May 10, 2022. <http://dx.doi.org/10.2139/ssrn.4105763>.

O'Keefe, Cullen, Peter Cihon, Ben Garfinkel, Carrick Flynn, Jade Leung, and Allan Dafoe. "The Windfall Clause: Distributing the Benefits of AI for the Common Good." Center for AI Governance, January 30, 2020.

Posner, Eric A. "Agency Models in Law and Economics." University of Chicago Law School, John M. Olin Law and Economics Working Paper No. 92, 2000. <https://ssrn.com/abstract=204872>

Pulejo, Massimo, and Pablo Querubin. "Plata Y Plomo: How Higher Wages Expose Politicians to Criminal Violence." NBER Working Paper No. w31586. National Bureau of Economic Research, 2023. <https://ssrn.com/abstract=4546459>

Siddarth, Divya, Sergey Ivliev, Santiago Siri, and Paula Berman. “Who Watches the Watchmen? A Review of Subjective Approaches for Sybil-Resistance in Proof of Personhood Protocols.” *Frontiers in Blockchain vol. 3* (2020). <https://doi.org/10.3389/fbloc.2020.590171>

Skyrms, Brian. “The Stag Hunt.” *Proceedings and Addresses of the American Philosophical Association 75*, no. 2 (2001): 31–41. <https://doi.org/10.2307/3218711>.

Subirà-Nieto, Jordi. “Security of Proof-of-Personhood: Idena.” Supervised by Bryan Ford, Louis-Henri Merino, and Haoqian Zhang. Decentralized Distributed Systems Laboratory—EPFL, July 11, 2021. https://www.epfl.ch/labs/dedis/wp-content/uploads/2021/07/report-2021-1-jordi-iden_a_report.pdf.

Yampolskiy, Roman. “AI-Complete, AI-Hard, or AI-Easy: Classification of Problems in Artificial Intelligence.” Paper presented at the 23rd Midwest Artificial Intelligence and Cognitive Science Conference, Cincinnati, OH, 2012.

NEWS, MAGAZINE ARTICLES, WEBSITE CONTENT & BLOG POSTS

Altman, Sam. “Moore’s Law for Everything.” *Sam Altman’s Blog*, March 16, 2021. <https://moores.samaltman.com/>.

Austgen, James, Kushal Babel, Vitalik Buterin, Phil Daian, Ari Juels, and Mahimna Kelkar. “Complete Knowledge.” *Medium*, January 16, 2023. <https://medium.com/initc3org/complete-knowledge-eecdda172a81>

Bailey, Andrew and Nick Almond. “Worldcoin Isn’t as Bad as It Sounds: It’s Worse.” *Block Works*, July 26, 2023. <https://blockworks.co/news/worldcoin-privacy-concerns>.

Blania, Alex, and Sam Altman. “Introducing Worldcoin: A Letter from Alex Blania and Sam Altman.” *Worldcoin Blog*, July 24, 2023. <https://worldcoin.org/blog/worldcoin/introducing-worldcoin-alex-blania-sam-altman>.

Bello, Camille. “Worldcoin: The Crypto Project Looking to Take on the World with Its Iris-Based ID Tech.” *Euronews*, August 11, 2023. <https://www.euronews.com/next/2023/08/11/worldcoin-the-crypto-project-looking-to-take-on-the-world-with-its-iris-based-id-tech>.

Buterin, Vitalik. “Minimal Anti-Collusion Infrastructure.” *EthResearch*, May 2019. <https://ethresear.ch/t/minimal-anti-collusion-infrastructure/5413>

Buterin, Vitalik. “Progress.” *Vitalik Buterin’s Blog*, November 22, 2019. <https://vitalik.eth.limo/general/2019/11/22/progress.html>.

Buterin, Vitalik. “Sharding.” *Vitalik Buterin’s Blog*, April 7, 2021. <https://vitalik.eth.limo/general/2021/04/07/sharding.html>.

Buterin, Vitalik. "Why Proof of Stake." *Vitalik Buterin's Blog*, November 6, 2022. <https://vitalik.eth.limo/general/2020/11/06/pos2020.html>.

Buterin, Vitalik. "What do I think of Biometric Proof of Personhood." *Vitalik Buterin's Blog*, July 24, 2023. <https://vitalik.eth.limo/general/2023/07/24/biometric.html>.

Cameron, Kim. "The Laws of Identity." *Kim Cameron's Identity Blog*, May 2005. <https://www.identityblog.com/?p=352>.

Daian, Philip, Tyler Kell, Ian Miers, and Ari Juels. "On-Chain Vote Buying and the Rise of Dark DAOs." *Hacking, Distributed*, July 2, 2018. <http://hackingdistributed.com/2018/07/02/on-chain-vote-buying/>.

Doward, Jamie. "The big tech backlash." *The Guardian*. January 28, 2018. <https://www.theguardian.com/technology/2018/jan/28/tech-backlash-facebook-google-fake-news-business-monopoly-regulation>.

Gent, Edd. "Worldcoin Launched. Then Came the Backlash: The Globe-Spanning Cryptocurrency and Biometric Identity Project Has Agitated Regulators." *IEEE Spectrum* August 28, 2023. <https://spectrum.ieee.org/worldcoin-2664361259>.

Gent, Edd. "Is Worldcoin a Crypto-Currency for the Masses or Your Digital ID? The Project Aims to Scan All the World's Eyeballs." *IEEE Spectrum*, December 22, 2022. <https://spectrum.ieee.org/worldcoin>.

Green, Matthew. "Some Rough Impressions of Worldcoin." *Matthew Green's Blog*, August 21, 2023. <https://blog.cryptographyengineering.com/2023/08/21/some-rough-impressions-of-worldcoin/>.

Gkritsi, Eliza, and Lingling Xiang. "Black Market for Worldcoin Credentials Pops Up in China." *Coindesk*, May 24, 2023. <https://www.coindesk.com/policy/2023/05/24/black-market-for-worldcoin-credentials-pops-up-in-china>.

Gkritsi, Eliza and Oliver Knight. "Worldcoin's Tokenomics Shrouded in Mystery as Website is Reportedly Geo-Blocked Worldwide." *Coindesk*, July 24, 2023. <https://www.coindesk.com/business/2023/07/24/worldcoin-release-tokenomics-report-geofenced-for-some-countries>.

Guo, Eileen, and Adi Renaldi. "Deception, Exploited Workers, and Cash Handouts: How Worldcoin Recruited Its First Half a Million Test Users." *MIT Technology Review*, April 6, 2022. <https://www.technologyreview.com/2022/04/06/1048981/worldcoin-cryptocurrency-biometrics-web3/>.

Hersey, Frank. "Worldcoin Says SDK Lets You Prove You're a Human Online. Coins Not Included." *Biometric Update*, March 17, 2023. <https://www.biometricupdate.com/202303/worldcoin-says-sdk-lets-you-prove-youre-a-human-online-coins-not-included>.

Idena. "AI-Resistant CAPTCHAs: Are They Really Possible?" *Medium*, May 8, 2019. <https://medium.com/idenai/ai-resistant-captchas-are-they-really-possible-760ac5065bae>.

Idena. "Idena Chronicles 0090." *Medium*, August 15, 2022. <https://medium.com/idenal/idenal-chronicles-0090-5f3efec5c3f>.

Idena. "Idena Hard Fork Announcement: Mining Delegation and Oracle Voting." *Medium*, March 10, 2021. <https://medium.com/idenal/idenal-hard-fork-announcement-mining-delegation-and-oracle-voting-8a5f9ddd9797>.

Idena. "Docs." Accessed December 2, 2023. <https://docs.idena.io>

Idena. "FAQ." Accessed December 2, 2023. <https://www.idena.io/faq>.

Idena. "Staking." Accessed December 2, 2023. <https://www.idena.io/staking>.

Idena. "White Paper." Accessed December 2, 2023. <https://docs.idena.io/docs/wp/summary>.

Matthews, Dylan. "How 'windfall profits' from AI companies could fund a universal basic income." *Vox*, July 28, 2023. <https://www.vox.com/future-perfect/23810027/openai-artificial-intelligence-google-deepmind-anthropic-ai-universal-basic-income-meta>.

Neary, Lynn. "Real 'Sybil' Admits Multiple Personalities Were Fake." *NPR*, October 20, 2011. <https://www.npr.org/2011/10/20/141514464/real-sybil-admits-multiple-personalities-were-fake>

Schneier, Bruce. "Tigers Use Scent, Birds Use Calls — Biometrics Are Just Animal Instinct." *The Guardian*, January 8, 2009. <https://www.theguardian.com/technology/2009/jan/08/identity-fraud-security-biometrics-schneier-id>.

Shed, Sam. "Silicon Valley leaders think A.I. will one day fund free cash handouts. But experts aren't convinced." *CNBC*, March 30, 2021. <https://www.cnn.com/2021/03/30/openai-ceo-sam-altman-says-ai-could-pay-for-ubi-experts-disagree.html>.

"Sock Puppeteer." IPFS. Evidence submitted to Kleros court case concerning Proof of Humanity profile 0xe825e609d15dd004d4b35dd858a55fd094db7f11 engaging in sock puppeteering. Accessed December 3, 2023. <https://ipfs.kleros.io/ipfs/QmNQxQff3UN4KfHqjjxvNca8pv96dUSKvvd7fiQCfmz3AV/sock-puppeteer.pdf>.

Worldcoin. "Humanness in the Age of AI." *Worldcoin Blog*, March 31, 2023. <https://worldcoin.org/blog/engineering/humanness-in-the-age-of-ai>.

WorldCoin. "Introducing World ID 2.0." *WorldCoin Blog*, December 13, 2023. <https://worldcoin.org/blog/announcements/introducing-world-id-2.0>.

Worldcoin. "Opening the Orb: A look inside Worldcoin's biometric imaging device." *Worldcoin Blog*, January 27, 2023, accessed December 2, 2023. <https://worldcoin.org/blog/engineering/opening-orb-look-inside-worldcoin-biometric-imaging-device>.

WorldCoin. "White Paper." Accessed December 2, 2023. <https://whitepaper.worldcoin.org>.

DATA SOURCES

CEIC Data. "Indonesia Monthly Earnings." Accessed December 3, 2023. <https://www.ceicdata.com/en/indicator/indonesia/monthly-earnings>.

CEIC Data. "Russia Monthly Earnings." Accessed December 3, 2023. <https://www.ceicdata.com/en/indicator/russia/monthly-earnings>.

CoinGecko. "Idena Price." Last modified March 1, 2024. <https://www.coingecko.com/en/coins/idena>.

Idena. "API Endpoint for Idena Blockchain Indexer." Accessed on February 3, 2024. <https://api.idena.io>.

Idena. "Community Discord Server." Discord. Accessed on February 3, 2024. <https://discord.gg/8BusRj7>.

Idena. "Idena Blockchain Explorer." Accessed on February 3, 2024. <https://scan.idena.io>.

Idena. "Idena Public Network." Telegram. Accessed on February 3, 2024. <https://t.me/IdenaNetworkPublic>.

Idena. "Idena Network" GitHub. Accessed on December 2, 2023. <https://github.com/idena-network>.

Idena-Mirror. "Commit History of 'Idena-Mirror' Repository." GitHub. Accessed December 2, 2023. <https://github.com/haritowa/idena-mirror/commits/master>.

WiseVoter. "Median Income by Country: Russia." Accessed December 2, 2023. <https://wisevoter.com/country-rankings/median-income-by-country/#russia>.

WiseVoter. "Median Income by Country: Indonesia." Accessed December 3, 2023. <https://wisevoter.com/country-rankings/median-income-by-country/#indonesia>.

World Justice Project. "Rule of Law Index." Accessed January 15, 2024. <https://worldjusticeproject.org/rule-of-law-index/>.