9

The Siren Song: Algorithmic Governance by Blockchain

Kevin Werbach

> A mysterious new technology emerges . . . its effects become profound; and later, many people wonder why its powerful promise wasn't more obvious from the start. What technology am I talking about? Personal computers in 1975, the Internet in 1993, and – I believe – Bitcoin in 2014.[1]

> [L]et them bind thee in the swift ship, hand and foot, upright in the mast-stead, and from the mast let rope-ends be tied, that with delight thou mayest hear the voice of the Sirens. And if thou shalt beseech thy company and bid them to loose thee, then let them bind thee with yet more bonds.[2]

A central theme in internet history since the 1990s is the rise of algorithmic power, enabled through the self-restraint of human governments.[3] Digital platforms were born weak and clumsy. Governments could have stamped them out to enforce traditional territorial boundaries and regulatory categories. They chose not to.[4] Once the digital tornado was unleashed, however, its path was not easily directed. Fledgling innovators in need of protection developed into dominant platforms that transformed many aspects of the world for the better, but also created serious harms through pervasive data collection and automated decision-making. The threats arose from the very attributes that made these digital systems so appealing.

The cycle is repeating itself. Another broad-based technological shift promises huge gains in both efficiency and freedom by replacing established points of control with open decentralized mechanisms. Startups spin visions of overwhelming established industries and surmounting government-established controls. And once again, a great challenge is how to restrain their own penchant for algorithmic overreach.

[1] Marc Andreessen, Why Bitcoin Matters, *N.Y. Times DealBook* (Jan. 21, 2014, 11:54 AM), https://dealbook.nytimes.com/2014/01/21/why-bitcoin-matters/.

[2] Homer, The Odyssey, Book XII (Trans. S.H. Butcher and L. Lang, 1937).

[3] See, e.g., John Danaher, The Threat of Algocracy: Reality, Resistance and Accommodation, 29 *Phil. & Tech.* 245 (2016). Most of the contributions to this volume concern the challenge of algorithmic power in one form or another.

[4] See Kevin Werbach, The Federal Computer Commission, 84 *N.C. L. Rev.* 1 (2005).

This time, the candidate technology is blockchain, and the broader phenomenon of "distributed ledger" systems.[5] Blockchain technology is still relatively immature. There is significant uncertainty about how it will develop in the future, and whether it will achieve anything like its promised level of impact. Already, however, blockchain and its related phenomenon, cryptocurrencies, have captured the imagination of technologists, entrepreneurs, business executives, and governments around the world. The driver for this activity is the belief that blockchain can foster an "internet of value"[6] – a new internet that overcomes the intermediation and centralized control that are increasingly prominent in the current digital environment.[7]

THE NEXT WAVE?

Like the internet, blockchain and cryptocurrencies are stimulating dramatic levels of investment, startup activity, and media attention, as well as creating massive disruption of industries and passionate visions of societal transformation.[8] As with the internet, this excitement often gets ahead of reality. The internet economy recovered from the dotcom crash of the early 2000s to realize its potential through the growth of social media, cloud computing, and mobile connectivity. The crypto economy seems likely to experience a similar trajectory over time. To succeed at scale, however, blockchain-based networks and services will need to address the problem of governance. Immutability, the mechanism that allows these systems to generate trust without central authorities, also creates inherent weaknesses that sometimes turn into catastrophic failures.

The Blockchain Phenomenon

For centuries, ledgers have been the foundation for the accounting and record-keeping around which societies are organized.[9] However, they have always been centralized: controlled by one or more entities with power over the recording and

[5] See Kevin Werbach, *The Blockchain and the New Architecture of Trust* 14 (2018) [hereinafter Werbach, New Architecture]. I use blockchain here as a generic term for the collection of cryptocurrency, blockchain, and distributed ledger technologies. Not all blockchain networks have an integral cryptocurrency, and not all cryptocurrencies use a data structure involving chains of transaction blocks. What they share are common properties such as decentralization (no one entity can control the status of the ledger) and immutability (transactions once made are, ideally, impossible to alter).

[6] Although now widely used, the "internet of value" phrase is most widely associated with the blockchain payments firm Ripple. See, e.g., Shanna Leonard, The Internet of Value: What It Means and How It Benefits Everyone, *Ripple Insights*, June 21, 2017, https://ripple.com/insights/the-internet-of-value-what-it-means-and-how-it-benefits-everyone/.

[7] See Steven Johnson, Beyond the Bitcoin Bubble, N.Y. *Times Mag.*, Jan. 16, 2018, a*vailable at* https://www.nytimes.com/2018/01/16/magazine/beyond-the-bitcoin-bubble.html.

[8] See, e.g., Don Tapscott and Alex Tapscott, *Blockchain Revolution* (2016).

[9] See Quinn DuPont and Bill Maurer, *Ledgers and Law in the Blockchain*, *King's Rev.*, June 23, 2016; Douglas Allen, *The Institutional Revolution: Measurement and the Economic Emergence of the Modern World* (2011).

approval of transactions. Even when there are multiple copies of information, one must either be designated as the master or there must a reconciliation process to redress any inconsistencies. Blockchain offers a decentralized alternative. Each party to a transaction can control its own information, while still trusting the information it sees from others.

Someone, or a group of people, using the pseudonym Satoshi Nakamoto kicked off the blockchain phenomenon on October 31, 2008 with the distribution on an internet mailing list of a short whitepaper titled *Bitcoin: A Peer-to-Peer Electronic Cash System*.[10] As extraordinary a breakthrough as it represented, there were virtually no technical advances in the paper. Instead, Nakamoto cleverly combined concepts from several streams of academic research and hobbyist tinkering, and then applied them to create the first workable form of private digital cash.[11] The Bitcoin network, based on voluntary participation and open-source software, launched in January 2009. Other cryptocurrencies followed. Many added additional functionality and expanded the technology beyond financial applications. A blockchain ledger can reliably record anything. Even more exciting, the ledger can function as a global distributed computer, which operates reliably without anyone in charge. Blockchain technology thus promises to eliminate inefficient intermediaries and overcome interorganizational trust gaps in an extraordinary range of contexts, from supply chain management to digital collectibles to the internet of things to property transfers.[12]

Although designed for functions such as payments and decentralized software applications, cryptocurrencies have so far found their most active use in speculative trading as a financial asset class. The price of bitcoin fluctuated for several years and then skyrocketed during 2017. At its peak in December 2017, the aggregate value of bitcoin in circulation exceeded $200 billion, and the overall cryptocurrency market was more than triple that.[13] Thousands of startups around the world began developing blockchain-based technologies, many of them issuing digital "tokens" to fund their networks. Most of the world's major financial services and industrial firms began to explore potential applications, and virtually all of the leading enterprise information technology services vendors developed substantial blockchain practices.[14]

[10] Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, https://bitcoin.org/bitcoin.pdf.

[11] See Arvind Narayanan and Jeremy Clark, Bitcoin's Academic Pedigree, 60 *Comms. ACM* 36 (2017).

[12] See Werbach, New Architecture, supra note 5, at 82–3.

[13] See Stan Higgins, *$600 Billion: Cryptocurrency Market Cap Sets New Record*, *Coindesk* (Dec. 18, 2017, 6:50 PM UTC), https://www.coindesk.com/600-billion-cryptocurrency-market-cap-sets-new-record/; Historical Snapshot- December 31, 2017, *CoinMarketCap* (2017) (valuing the market cap of Bitcoin at approximately $221 Billion). These numbers somewhat overstate the actual value of available bitcoin, as it ignores the substantial amount of the cryptocurrency that has been stolen or for which the cryptographic keys have been lost.

[14] See Werbach, New Architecture, supra note 5, at 84.

For those who lived through the dotcom bubble of the late 1990s, the parallels are striking. Projects with little more than a whitepaper raised tens of millions of dollars from investors around the world. Companies saw their value skyrocket overnight, without any real customer adoption. Experts talked of a new economy in which old metrics were no longer useful, and established industry leaders were soon swept away. And, as with the dotcom bubble of 1998–99, the 2017 cryptocurrency bubble was quickly followed by a brutal "crypto winter," in which prices plummeted and many projects were abandoned.[15]

Despite overexuberant claims and widespread illicit activity, however, blockchain technology itself, like the internet, is no fraud. It represents an immature but foundational development whose impacts will unfold over time. Where the internet lowered costs of transferring information, blockchain lowers costs of transferring value.[16] The impacts of this shift will be broad. Secure value exchange is not just a property of banking and payments; it is a basic building block of markets and society. Standing behind the money and security is a deeper property of trust.[17]

Blockchain as a Trust-Based Technology

Blockchain is fundamentally a trust-based technology.[18] Although Bitcoin relies on blockchain architecture as its foundation for digital currency, blockchain technology itself has been applied to a broad range of other applications. The unifying attribute of these applications is that they require a network of participants to preserve the integrity of shared information. If digital assets on the network cannot be trusted, for example, they are of little value. The distinctive attribute of the blockchain approach is that it expands trust in the system as a whole by minimizing trust in specific authorities or intermediaries that may prove fallible.[19] Investor and LinkedIn co-founder Reid Hoffman cleverly calls this, "trustless trust."[20] The key technical arrangement is known as consensus: All participants must converge on, and receive verifiable assurances of, the exact state of the network, without any enforceable formal agreements.

[15] See Paul Vigna, Bitcoin Is in the Dumps, Spreading Gloom Over Crypto World, *Wall St. J.*, March 19, 2019, https://www.wsj.com/articles/bitcoin-is-in-the-dumps-spreading-gloom-over-crypto-world-11552927208.

[16] See Christian Catalini and Joshua Gans, *Some Simple Economics of the Blockchain*, Rotman School of Mgmt. Working Paper No. 2874598 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2874598; Marco Iansiti and Karim R. Lakhani, *The Truth about Blockchain*, 95 *Harv. Bus. Rev.* 118 (Jan./Feb. 2017).

[17] See Werbach, New Architecture, supra note 5, at 84–6.

[18] See id.; Sinclair Davidson et al., Blockchains and the Economic Institutions of Capitalism, 14 *J. Institutional Econ.* 639 (2018).

[19] See Kevin Werbach, Trust, but Verify: Why the Blockchain Needs the Law, 32 *Berkeley Tech. L.J.* 489 (2018) [hereinafter Werbach, Trust but Verify]; Werbach, New Architecture, supra note 5, at 28–30.

[20] Werbach, Trust but Verify, supra note 19, at 79–81.

Bitcoin, for example, uses a system called proof of work to avoid the need to trust a bank or intermediary to verify payments. It establishes a competition every ten minutes to validate chunks of transactions (referred to as blocks) and earn a reward (in bitcoin). The winner is effectively selected at random, however the amount of computer processing power each Bitcoin validator, known as "miner," brings to bear will increase their likelihood of winning. Bitcoin miners will therefore spend tens of millions of dollars per day in hardware and electricity to increase the likelihood of winning. The purpose of the proof of work system is twofold: To incentivize participation (on the part of the miners) and to constrain behavior (on the part of anyone who might undermine the integrity of the system). It also enhances the security of the system as a whole: An attacker must compete against the computational power as the rest of the network combined.

Thus, even if any participant in Bitcoin's proof of work system is selfishly motivated to steal from the network, none has the power to do so. Moreover, the network is "censorship-resistant," meaning any transaction cannot easily be altered or removed. There is no master control point that everything depends on. Anyone around the world can become a Bitcoin node by running some open-source software, and the network functions as long as there is enough mining activity to guarantee security.

Bitcoin's proof of work system is the most well-established blockchain consensus mechanism. Since the network launched in 2009, no one has successfully undermined it to alter the transaction ledger or spend the same coin twice.[21] However, it is not the only possible approach. Bitcoin's success sparked an explosion of research and experimentation with approaches making different fundamental tradeoffs among scalability, security, and decentralization. Other prominent blockchain networks include Ethereum, Ripple, EOS, Dash, Monero, and ZCash. There is also ongoing work to address the inherent scalability and functionality limitations in Bitcoin's design. And in recent years, enterprises and governments have begun to implement permissioned blockchain networks that, unlike Bitcoin, are limited to authorized participants.[22]

[21] This does not mean no bitcoin has been stolen. Cryptocurrencies are bearer instruments. Whoever controls a private cryptographic key effectively owns the currency associated with it. The wallets and exchanges where most users store their cryptocurrency are separate from the decentralized consensus ledger itself. Those centralized systems can be hacked, or keys can be stolen through other means. An estimated $1.2 billion of cryptocurrency was stolen between the beginning of 2017 and May 2018. See Gertrude Chavez-Dreyfuss, About $1.2 Billion in Cryptocurrency Stolen Since 2017: Cybercrime Group, *Reuters* (May 24, 2018, 10:59 AM), https://www.reuters.com/article/us-crypto-currency-crime/about-1-2-billion-in-cryptocurrency-stolen-since-2017-cybercrime-group-idUSKCN1IP2LU.

[22] See UK Government Chief Scientific Advisor, Distributed Ledger Technology: Beyond Block Chain (2015), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/492972/gs-16-1-distributed-ledger-technology.pdf; Tim Swanson, Consensus-as-a-Service: A Brief Report on the Emergence of Permissioned, Distributed Ledger Systems (Apr. 6, 2015), http://www.ofnumbers.com/wp-content/uploads/2015/04/Permissioned-distributed-ledgers.pdf.

The other important innovation of blockchain systems is the smart contract.[23] Smart contracts are securely self-executing software code that run on a blockchain network. Essentially, smart contracts allow a blockchain application to function as a parallel distributed computer, in which every machine running the application provably does so in exactly the same way. Smart contracts are the foundation of the functionality of blockchain technology. Smart contracts are broader than legal contracts, in that they can – within limits of performance scalability – encode anything that can be written into a computer program. From a conceptual and doctrinal perspective, however, they are simply contracts.[24] They allocate rights and responsibilities among parties who voluntarily bind themselves into enforceable commitments. Contracts are a powerful means of generating trust because they backstop voluntary human commitments with formalized legal enforcement embodying the power of the state. Smart contracts are designed to offer a similar kind of confidence backed by the integrity of the blockchain ledger. Which is to say, blockchain is a legal or regulatory technology.[25] It is a method of governance.[26]

However, to the extent blockchain is a governance technology, it is immature, without the flexibility or capacity to correct for errors or unforeseen situations. In order to garner broader trust and move past its current limited applications, blockchain governance must become more robust.

LASHED TO THE MAST: THE TWO SIDES OF IMMUTABILITY

In Homer's *The Odyssey*, the hero Odysseus encounters sirens, mermaids who lure sailors to their deaths with their enchanting song.[27] Odysseus is curious about the content of their songs, but he knows that if he hears them, he will not be able to resist plunging into the ocean. So he orders his men to lash him to the mast of his ship. He further orders them to fill their ears with wax, so that if he later urges them to untie him, they will not hear his pleas. Odysseus thus empowers himself to hear the music

[23] See Kevin Werbach and Nicolas Cornell, Contracts Ex Machina, 67 *Duke L.J.* 314 (2017); Nick Szabo, Formalizing and Securing Relationships on Public Networks, 2 *First Monday* (1997), http://ojphi.org/ojs/index.php/fm/article/view/548.

[24] See Werbach and Cornell, supra note 23.

[25] See Lawrence Lessig, *Code and Other Laws of Cyberspace* (2nd revised ed. 2006); Werbach, New Architecture, supra note 5, at 189; Primavera de Filippi and Aaron Wright, *Blockchain and Law: The Rule of Code* (2018).

[26] See Rachel O'Dwyer, Code != Law: Explorations of the Blockchain as a Mode of Algorithmic Governance (2018), https://www.academia.edu/34734732/Code_Law_Explorations_of_the_Blockchain_as_a_Mode_of_Algorithmic_Governance; Sinclair Davidson, Primavera De Filippi, and Jason Potts, Economics of Blockchain (March 8, 2016), http://dx.doi.org/10.2139/ssrn.2744751; Marcella Atzori, Blockchain Technology and Decentralized Governance: Is the State Still Necessary? (2015), http://nzz-files-prod.s3-website-eu-west-1.amazonaws.com/files/9/3/1/blockchain+Is+the+State+Still+Necessary_1.18689931.pdf.

[27] See The Odyssey, supra note 2.

that no mortal man can survive. He does so, ironically, by radically disempowering himself and his sailors at the critical moment.

The same strategy lies at the heart of the blockchain's capability to decentralize trust. In the blockchain context, this strategy is known as immutability. Immutability is a great strength of blockchain-based governance systems, but also potentially a catastrophic weakness.

Blockchain Immutability

Immutability on a blockchain means that once a transaction has been incorporated into a validated block and added to the ledger, it cannot be altered.[28] This kind of guarantee is quite difficult to achieve in digital systems, whose records are naturally ephemeral and encoded in the universal language of binary ones and zeros. In the words of computer scientist and smart contracts pioneer Nick Szabo: "Typical computers are computational etch-a-sketch, while blockchains are computational amber."[29] Blockchain systems enforce immutability by making every piece of information reflect the consensus agreement of a network of computers. Changing even the smallest fact means convincing a large percentage of the network to reconsider its settled transaction history. The algorithms and cryptography of the consensus system are designed to make that exceedingly difficult.

From an internet policy perspective, immutability seems to put things backwards. The internet regulation debate is fundamentally about freedom. Decentralized global networks make it easier for people to engage in conduct that some would like to prevent, whether that involves dissidents challenging authoritarian regimes or consumers accessing media they didn't pay for. As only became clear over time, those networks also concentrate power in digital platforms whose freedom of action is difficult to shackle under conventional mechanisms of antitrust, contract, or privacy protection. Governments responded to the first concern through a variety of mechanisms; their ability to put the platform power and surveillance capitalism genies back in the bottle is yet to be seen.

Like the internet, blockchain systems are often described as technologies of freedom, but in their core functioning they are just the opposite. What makes a blockchain trustworthy is precisely that it restricts freedom to diverge from the consensus state of the ledger. This characteristic is important for security. Transactions involving scarce or valuable assets would not be trustworthy if someone could easily alter the ledger. Beyond

[28] See Nick Szabo, *Money, Blockchains, and Social Scalability*, *Unenumerated* (Feb. 9, 2017), http://unenumerated.blogspot.com/2017/02/money-blockchains-and-social-scalability.html ("To say that data is post-unforgeable or immutable means that it can't be undetectably altered after being committed to the blockchain."); Angela Walch, The Path of the Blockchain Lexicon (and the Law), 36 *Rev. Banking & Fin. L.* 713 (2016–2017); Marc Pilkington, Blockchain Technology: Principles & Applications, in *Research Handbook on Digital Transformations* 15 (F. Xavier Olleros & Majlinda Zhegu eds., 2016).

[29] Szabo, supra note 28.

that, however, immutability is blockchain's most significant contribution to governance. It is also the property that creates the most significant risks of catastrophic failure.[30]

Immutability poses a novel set of legal and regulatory challenges. For the most part, cyberlaw is concerned with the plasticity of digital systems. Software can be coded to arbitrage around legal rules. Information can be combined and analyzed to create challenges not present at the outset, such as data aggregation to undermine privacy protections. The challenge has been to tie down actors and systems to particular jurisdictions or classifications. Immutability creates a different problem. The illegitimacy or harm of certain actions may be well-established, but no one may have the ability to do anything about it.

Immutability as a Means of Trust

Immutability is essential to blockchain technology in several ways. It is a proxy for the basic security of the network. If you know that information you see on a blockchain is immutable, you can rely on it. Even more significant, immutability is implicit in blockchain's approach to trust. If any actor had the power to change the ledger retrospectively, everyone else would need to trust that actor not to do so in secret or illegitimate ways. This is true whether the empowered entity is a thief, a validator, an intermediary, or a government. A blockchain network must be immutable to be censorship-resistant, because a censor is a government agent that demands changes to the information recorded. Thus, the decentralized model of blockchain trust depends on immutability.

Satoshi Nakamoto emphasized this point in the original Bitcoin whitepaper. In the centralized financial system, he or she or they pointed out: "[C]ompletely non-reversible transactions are not really possible, since financial institutions cannot avoid mediating disputes."[31] This need for dispute resolution puts power in the hands of governments and intermediaries. And thus, as Nakamoto continued: "With the possibility of reversal, the need for trust [in particular entities] spreads."[32] In order to separate generalized trust in transactions from trust in specific fallible actors, Bitcoin had to ensure that records on the ledger could not be reversed.

Immutability is not a precise concept.[33] In particular, it does not mean changing the ledger is categorically precluded.[34] For Bitcoin and similar blockchain networks,

[30] There are also plenty of ill-intentioned parties that use blockchain technology to facilitate illegal or unethical activity. See, e.g., Kevin Werbach, What the Russia Hack Indictments Reveal about Bitcoin, *N.Y. Times*, July 22, 2018, https://www.nytimes.com/2018/07/22/opinion/russia-hacking-indictments-bitcoin.html. These issues lie beyond the scope of this paper.

[31] Nakamoto, supra note 10.

[32] Id.

[33] See Walch, supra note 28.

[34] See id. at 738–41; Werbach, Trust but Verify, supra note 19; Gideon Greenspan, The Blockchain Immutability Myth, *Coindesk* (May 9, 2017), http://www.coindesk.com/blockchain-immutability-myth/.

immutability is a statistical property. The more time that has passed since a block was validated, the less likely it has been altered.[35] However, the integrity of the network can never be established absolutely; there is always some miniscule possibility that an attacker has successfully altered the chain.[36] Some other blockchain systems provide for "finality," which after a certain time prohibits changes to a validated block.[37] Even then, however, the ledger is not truly immutable.[38] And public blockchains are always potentially vulnerable to "51% attacks" if someone can obtain a majority of the total power in the network.[39] There has never been a successful 51 percent attack against Bitcoin, but there have been several against less-valuable cryptocurrencies.[40]

There are also situations in which changing the status of validated blocks may be desirable. Because blockchain networks are decentralized, every node can independently propose a new block to add to the existing chain. The consensus process is designed to ensure that the network continually converges to a single valid chain. When some percentage of nodes on a blockchain network choose to follow a different path than the rest of the network, it is called a fork.[41] This may occur for mundane reasons. For example, developers may upgrade a network's software with new features that are not backward-compatible with the earlier version. Those nodes running the non-upgraded software will remain on a different blockchain from everyone else, although if all goes well, that chain will quickly die out. Sometimes a fork is necessary to fix problems with the network, as when denial-of-service attacks were grinding the Ethereum network to a halt in late 2016.[42] A successful fork, however, can effectively reverse or alter prior transactions, thus undermining immutability.

The imperfection of blockchain immutability corresponds to the imperfection of trust. Trust is not the same as certainty. No one would say they trusted that $2 + 2 = 4$, or that a heavy object dropped from a height will fall toward the ground. Neither

[35] See Tim Ferriss, *The Quiet Master of Cryptocurrency—Nick Szabo, Tim Ferriss Show* (June 4, 2017), https://tim.blog/2017/06/04/nick-szabo/.

[36] Of course, this is true of centralized financial networks as well.

[37] See Vitalik Buterin, *On Settlement Finality, Ethereum Blog* (May 9, 2016), https://blog.ethereum.org/2016/05/09/on-settlement-finality/.

[38] The consensus algorithms that provide for finality also generally trade off some other property, such as security or decentralization, for the guarantee that prior blocks may not be changed.

[39] For proof of work systems like Bitcoin, this means a majority of computation devoted to mining.

[40] See Daniel Oberhaus, *Cryptocurrency Miners Are Sabotaging Blockchains for Their Personal Gain, Motherboard* (May 25, 2018, 11:00 AM), https://motherboard.vice.com/en_us/article/a3a38e/what-is-a-51-percent-attack-silicon-valley-bitcoin-gold-verge-monacoin-cryptocurrency. The lower the total value of a cryptocurrency, the less resources it makes economic sense to spend on mining, which means the costs of a successful 51 percent attack are also lower.

[41] A soft fork means both paths are compatible. There is still one consensus chain of blocks, but some network nodes have access to different features than the others. A hard fork means there are two incompatible chains after a certain point.

[42] See Thomas Jay Rush, *Defeating the Ethereum DDOS Attacks, Medium* (Feb. 12, 2017), https://medium.com/@tjayrush/defeating-the-ethereum-ddos-attacks-d3d773a9a063.

unshakable confidence in an outcome nor a rational calculus that drives reliance is equivalent to trust. Trust is a human quality, and as such, it requires some modicum of vulnerability.[43] It is the willingness to commit even though there is some residual risk in doing so. What makes trust valuable is that it goes beyond certainty. A trustworthy counterparty allows one to dispense with cumbersome verification or self-help enforcement, greatly enhancing the scope and efficiency of transactions.[44]

Trustworthy systems must therefore balance the necessary confidence to inspire action with the acknowledgement of imperfection. A thick bank vault may nonetheless be cracked, just as blockchain immutability may in some circumstances be undermined. Moreover, trust expands with experience or relationships. A system that is foolproof on paper may not be in practice. The design of Bitcoin published in 2008 convinced a small number of early adopters, but it was only after years of secure operation that more mainstream users were willing to trust their money to the seemingly strange decentralized system. Just as validated blocks on a public blockchain become more trustworthy over time, the entire blockchain becomes more trustworthy with successful experience.

Immutability as Commitment Device

Another way to think of blockchain immutability is as a kind of commitment device. Economists define a commitment device as "an arrangement entered into by an agent who restricts his or her future choice set by making certain choices more expensive, perhaps infinitely expensive."[45] Commitment devices bridge between our present and future selves. Odysseus rationally knew ahead of time that he should not heed the call of the sirens. However, he also realized that, in that nonrational moment, he would be powerless to resist. So he prospectively deprived himself not only of the capability to act, but of the capability to belay his earlier order to his crew.

The need for commitment devices is not limited to mythical mermaids. It comes up virtually any time we envision our future selves. Many of us have an easier time resisting the prospect of ice cream tomorrow than the Ben & Jerrys in front of us right now. In addition, behavioral economists have identified several cognitive biases that make actual behavior in the future diverge from rational expectations in the present.[46] Most notably, people tend to discount benefits hyperbolically. According to textbooks, the net present value of a future benefit declines linearly

[43] See Werbach, New Architecture, supra note 5, at 77.

[44] See Francis Fukuyama, *Trust: The Social Virtues and the Creation of Prosperity* (1995); Niklas Luhmann, *Trust and Power* (1979).

[45] Gharad Bryan et al., Commitment Devices, 2 *Ann. Rev. Econ.* 671–98 (2010). The authors include two additional conditions for precision. The agent must value the commitment effect itself, as opposed to, for example, merely paying now for a benefit later. And the commitment must not be for some strategic purpose, such as deterring action that might invoke it or influencing someone else.

[46] See id.

over time based on the relevant discount rate. In practice, most people overvalue near-term benefits and strongly undervalue those arriving far in the future.[47] Just as they have a hard time imagining the beneficial results of compound interest, they fail to properly appreciate even large far-off gains.

A commitment device allows us to bind our future selves to our present rational calculus. Yale University economists Gharad Bryan, Dean Karlan, and Scott Nelson give a simple example of a runner about to embark on a ten-mile training session.[48] She wants to run the whole way, but she knows that at some point she will become tired and likely slow to a walk. So she signs a contract agreeing to pay a friend $1,000 if she fails to run the whole way. The committed payment makes the walking option considerably less desirable.

Those who commit transactions to a blockchain do so with the knowledge that they are not easily reversible. As Satoshi Nakamoto explained in the Bitcoin white-paper, they are choosing nonreversibility ahead of time to avoid the trust-inducing processes of mediation and dispute resolution that will seem appealing in the future. Their commitment is necessary if the blockchain itself is to be trusted.

Credible commitments are essential to any bargaining relationship.[49] If I tell you I won't pay more than $100, your willingness to agree to my terms depends on your assessment of my credibility. In particular, contractual arrangements depend on the ability of the parties to convince one another that their commitments are credible. If you do not believe I will deliver the products you are paying for, you will not enter into such a contract with me. As elaborated by economist Oliver Williamson, the game theorist Thomas Schelling analogized credible commitments to hostage-taking in primitive societies.[50] The hostages were part of the agreement process. Each side would be confident in the performance of the other, because otherwise it would kill its hostages. Such gruesome mechanisms seemed necessary in the absence of legal dispute resolution mechanisms. As Williamson explains, we no longer require human hostages because we assume "efficacious rules of law regarding contract disputes are in place and that these are applied by the courts in an informed, sophisticated, and low-cost way."[51]

The philosopher John Elster, in his essay *Ulysses and the Sirens*, points out that commitment devices turn the rational actor model of neoclassical economics against itself.[52] Credible commitments are necessary for the contractual process of market exchange; otherwise, counterparties would breach agreements at will. However, carrying out those threats often requires behaving in a way that would

47 See Richard Thaler, Some Empirical Evidence on Dynamic Inconsistency, 8 *Econ. Lett.* 201 (1981).
48 See id. at 674.
49 See Thomas Schelling, An Essay on Bargaining, 46 *Am. Econ. Rev.* 281–306 (1956).
50 See Oliver Williamson, Credible Commitments: Using Hostages to Support Exchange, 73 *Am. Econ. Rev.* 519, 519 (1983).
51 Id. at 520.
52 See John Elster, Ulysses and the Sirens: A Theory of Imperfect Rationality, 16 *Soc. Sci. Info.* 469 (1977). Elster uses the Roman name of Ulysses instead of the Greek Odysseus.

otherwise be irrational. At the later moment when parties have already made relationship-specific investments in a contract, for example, consenting to an unjustified reduction in price would be preferable to walking away entirely. In an extreme case, Thomas Schelling famously applied game theory to the doctrine of mutually assured destruction in the Cold War. The United States and the Soviet Union avoided nuclear war by committing themselves to retaliation that would end life on Earth in the event of an attack. Humans, Elster concludes, are "imperfectly rational creatures able to deal strategically with their own myopia."[53]

THINGS GO WRONG

Serious problems emerge when the imperfect rationality implicit in credible commitments is implemented through the perfectly rational vehicle of computers executing smart contracts on a blockchain. The dark side to immutability is that even invalid or illegitimate transactions cannot easily be reversed. Immutability creates the potential for catastrophic failures with no clear means of remediation.

Three examples illustrate the problems with blockchain immutability: The DAO hack, the Parity wallet bug, and the abortive Segwit2x fork.

The DAO Hack (2016)

In June 2016, approximately $50 million in Ether cryptocurrency was extracted from the DAO, a decentralized crowdfunding application.[54] The DAO was a set of smart contracts on the Ethereum network that allowed individuals who purchased tokens to vote "yes" or "no" on financing a given project.[55] The more money a user put into the DAO, the more votes the user would receive, and subsequently the greater share the user would receive of income from successful projects. The fund raised 11.5 million Ether through its initial crowd sale, worth approximately $150 million at the time and representing nearly 15 percent of Ether in circulation.[56] Before it ever began funding projects, however, the DAO was undermined by a catastrophic hack.

[53] Id. at 502.

[54] See Klint Finley, A $50 Million Hack Just Showed that the DAO Was All Too Human, *Wired* (June 18, 2016), https://www.wired.com/2016/06/50-million-hack-just-showed-dao-human/.

[55] DAO is an abbreviation for decentralized autonomous organization, a concept of a corporation governed through self-executing smart contracts. See Vitalik Buterin, DAOs, DACs, DAs and More: An Incomplete Terminology Guide, *Ethereum Blog* (May 6, 2014), https://blog.ethereum.org/2014/05/06/daos-dacs-das-and-more-an-incomplete-terminology-guide/. The DAO was styled as the first real-world implementation of the concept, which cryptocurrency advocates such as Ethereum founder Vitalik Buterin had developed a few years earlier.

[56] See Toshendra Kumar Sharma, *Details of the DAO Hacking in Ethereum in* 2016, *Blockchain Council* (Aug. 20, 2017), https://www.blockchain-council.org/blockchain/details-of-the-dao-hacking-in-ethereum-in-2016/.

Someone took advantage of a vulnerability in the smart contract code that governed a narrow component of the fund's payment structure.[57] By repeatedly executing the same request function, the hacker was able to drain roughly one-third of the pool of committed investment currency into a private "child DAO." Thankfully, the system included a failsafe that prohibited fund withdrawals from the system for thirty days. During that period, the Ethereum Foundation produced a software upgrade that forked the entire blockchain to a new state in which the stolen funds were returned to their rightful owners.[58] However, the fork was controversial. It essentially broke the immutability of the blockchain in order to reverse the theft. Most members of the community considered this a worthwhile tradeoff. The price of Ether recovered from the uncertainty the DAO hack generated, and then climbed dramatically the following year. Many viewed the Ethereum Foundation's willingness to act a comforting example of effective governance.[59]

Others were not convinced. Immutability, they argued, was the essence of blockchain decentralization. If the Ethereum Foundation could convince most network nodes to roll back $50 million of transactions once, it could do so again. Perhaps the next time would be less clearly a case of theft. Perhaps it would be a controversial move that advantaged the Foundation's leadership over the rest of the community. And given the disruption involved in implementing a hard fork, it made no sense to take this tack every time someone exploited a smart contract bug. Where was the line to determine when immutability should be broken? While these opponents were a minority and couldn't prevent the hard fork, they could do something else. They started mining the other side of the fork, the chain in which the DAO funds were still in possession of the hacker.[60] This fork, labeled Ethereum Classic (ETC), continues today to exist in parallel to the main Ethereum (ETH) blockchain.[61]

The ETC objection to the DAO fork centered around credible commitments. Why trust the blockchain if it can be forked whenever something goes wrong? A noncredible commitment is worth nothing, or worse. When financial institutions in the 2000s realized they were "too big to fail" and would be bailed out for the government if their bets failed to pay off, their appetite for risk grew to the unsustainable levels that precipitated the global financial crisis of 2008. When several Central and Eastern European governments experienced hyperinflation in the years after World War I, in spite of increasingly vigorous monetary policy initiatives, they

[57] See id.

[58] See E. J. Spode, The Great Cryptocurrency Heist, *Aeon* (Feb. 14, 2017), https://aeon.co/essays/trust-the-inside-story-of-the-rise-and-fall-of-ethereum.

[59] See Nathaniel Popper, Move Over, Bitcoin. Ether Is the Digital Currency of the Moment, *N.Y. Times DealBook*, June 19, 2017, https://www.nytimes.com/2017/06/19/business/dealbook/ethereum-bitcoin-digital-currency.html.

[60] See David Morris, The Bizarre Fallout of Ethereum's Epic Fail, *Fortune* (Sept. 4, 2016), http://fortune.com/2016/09/04/ethereum-fall-out.

[61] The two chains are identical up to the moment of the fork, then diverge. All holders of Ether at the time saw their holdings double, although Ethereum Classic coins are worth substantially less than those of the more popular Ethereum.

mandated convertibility of their currencies into gold.[62] The gold standard made it impossible to debase the currency too far. By the 1970s, when countries were more stable and central banks more sophisticated, the gold standard and its limiting tether to physical assets were no longer needed.

The ideal credible commitment is strong enough to promote the desired behavior, but weak enough to be overcome through appropriate mechanisms when absolutely necessary. The ad hoc nature of the response to the DAO hack, and the fact that most of those connected with the DAO were also associated with the Ethereum Foundation, created skepticism about the need to break immutability.

The Parity Wallet Bug (2017)

In November 2017, Parity Technologies, an Ethereum-based blockchain developer, suffered a critical security vulnerability that affected certain users of the company's wallet software for storing cryptocurrency.[63] An update caused a bug that could have allowed a malicious user to control a large number of Parity's "multsignature" wallets. A user found the flaw and, allegedly to prevent theft, deleted the smart contract involved.[64] Unfortunately, this made it impossible for anyone to access the relevant wallets. As a result of this hack more than $280 million of Ether was frozen.[65] While the Ether was still immutably recorded on the Ethereum blockchain, it was simply inaccessible. Like the DAO, Parity had close ties to the Ethereum Foundation. Gavin Wood, its CEO, was the co-founder and chief technologist of Ethereum, and a large component of the frozen Ether was associated with Parity's own token offering for a blockchain interoperability project called Polkadot. A hard fork to restore the trapped Ether would seem like a bailout for insiders. Other solutions met with similar skepticism.[66] As of summer 2018, the funds remained trapped.

Unlike the DAO hack, the Parity wallet bug had no villain.[67] The cryptocurrency was apparently rendered inaccessible by accident. Yet the impact was similar.

[62] See Thomas J. Sargent, The Ends of Four Big Inflations, in *Inflation: Causes and Effects* (R.E. Hall, ed., 1982).

[63] See Parity Technologies, Security Alert (November 8, 2017), https://paritytech.io/security-alert-2/.

[64] See Jordan Pearson, Ethereum Wallet Company Knew about Critical Flaw That Let a User Lock Up Millions, *Motherboard* (Nov. 15, 2017, 2:21pm), https://motherboard.vice.com/en_us/article/d3djwj/ethereum-wallet-parity-knew-about-critical-flaw-that-let-user-devops199-lock-up-millions.

[65] See Ryan Browne, "Accidental" Bug May Have Frozen $280 Million Worth of Digital Coin Ether in a Cryptocurrency Wallet, *CNBC* (Nov. 8, 2017, 6:42 AM), https://www.cnbc.com/2017/11/08/accidental-bug-may-have-frozen-280-worth-of-ether-on-parity-wallet.html.

[66] See Rachel Rose O'Leary, The New Last-Ditch Effort to Unfreeze a $260 Million Ethereum Fortune, *Coindesk* (Apr. 18, 2018, 4:00 UTC), https://www.coindesk.com/new-last-ditch-effort-unfreeze-260-million-ethereum-fortune/.

[67] There is some question whether the user, Devops199, offered a truthful account, or was actually behaving maliciously. Even if deletion of the smart contract was malicious, however, it was not theft, because no one obtained access to the trapped funds.

Legitimate users who relied on the immutability of the blockchain lost their money as a consequence of that very immutability function. There was no mechanism to alter undesirable transactions after the fact, even when a transaction – locking every user permanently out of their wallets – produced benefits for no one.

Parity wallet users had good reason to trust the firm's software with their cryptocurrency. Parity's leaders were highly respected technologists who were intimately involved in the creation of Ethereum. Gavin Wood, in fact, was the primary creator of the Solidity programming language used for Ethereum smart contracts. One would not expect his company to make a relatively elementary Solidity coding flaw. And one would certainly not expect it to leave the flaw in place for months after being told about it.[68] Yet the reality is that individual and companies are fallible. Trusting Parity was as reasonable as trusting the banks that imploded during the 2008 financial crisis. The difference was that, thanks to a combination of government-mandated insurance and operational mechanisms, no one would ever find their money "permanently stuck" in a bank's savings account with no recourse.

Trust is a double-edged sword. Users trust Parity because its software operates on an immutable blockchain. However, they don't necessarily trust Parity enough to implement a hard fork to restore its frozen Ether. The second requires trust in specific human organizations, which is exactly what the blockchain's immutability was designed to overcome.

The SegWit 2x Battle (2017)

For a number of years, there has been a contentious technical debate among leading Bitcoin developers about how to scale the network. Bitcoin can process a theoretical maximum of seven transactions per second, which is thousands of times fewer than centralized payment-processing systems. As the price of Bitcoin rose and transaction activity increased, the network began to slow down even further. Some developers believed the solution was to change the protocol to increase the amount of data processed in each block. However, that would require a hard fork. It would represent the first substantial step away from the basic architecture that Satoshi Nakamoto outlined in 2008, which is the basis for the Bitcoin network's remarkable run of secure, uninterrupted operation. Other developers felt that different mechanisms could address the scalability challenge without changing the core protocol, or that rock-solid security was simply more important than handling more transactions.

In spring 2017, a compromise was brokered among major Bitcoin-related companies to implement two competing scalability proposals.[69] The first, SegWit, could go

[68] See O'Leary, supra note 64.

[69] See Laura Shin, Will This Battle for the Soul of Bitcoin Destroy It?, *Forbes* (Oct. 23, 2017, 1:35 pm), https://www.forbes.com/sites/laurashin/2017/10/23/will-this-battle-for-the-soul-of-bitcoin-destroy-it/#42adb77f3d3c.

into effect prior to a hard fork.[70] It provided foundation for scaling Bitcoin without disturbing the core protocol. The second component was a doubling of the block size referred to as 2x, which was to be implemented in a hard fork later in the year. The SegWit implementation proceeded smoothly. As the date for the 2x hard fork approached, however, controversy reemerged. Critics labeled the compromise, known as the New York Agreement, an illegitimate back-room deal and a corporate takeover of Bitcoin.[71] And it began to seem likely that, as with Ethereum Classic, some network nodes would continue mining the original, small block-size chain even after the fork. That led to speculation about which chain deserved to carry forward the "Bitcoin" name and its BTC ticker symbol on exchanges.[72] The hard fork was ultimately abandoned.[73]

The Segwit 2x battle, unlike the prior two examples, didn't deprive anyone of their cryptocurrency. It involved neither theft nor buggy code. Yet it provoked a similar sense of existential crisis over the essence of Bitcoin. Does immutability mean it must be next to impossible to change the basic properties of a blockchain network, in addition to the transaction records it stores? Removing human intervention from every commitment by means of a software-implemented commitment device seems well and good, but software is created by humans too. They can't ever fully anticipate the needs of the future. At some point, there will be a need to evolve the system if it is to remain trustworthy. Yet the upgrade process itself opens the Pandora's Box that immutability was supposed to seal shut.

BE CAREFUL ABOUT YOUR COMMITMENTS

Political theorist Kenneth Shepsle distinguishes two forms of commitment device: Motivational and imperative.[74] The first involves commitments that are incentive compatible. That is to say, at the time the device operates, the person involved rationally desires to comply. The second form of commitment device requires coercion, because otherwise the person involved would not follow through on the commitment. Blockchain systems employ both. Consensus systems like proof of

[70] The term is short for "segregated witness," an allusion to the famous prisoners' dilemma scenario in game theory. Two alleged witnesses to a crime must be put in separate rooms in a prison, unable to communicate, in order to produce the classic result that both will confess.

[71] See Shin, supra note 69.

[72] See id.

[73] See Paul Vigna, Bitcoin Dodges Split that Threatened Tts Surging Price, *Wall St. J.* (Nov. 8, 2017, 3:25 pm), https://www.wsj.com/articles/bitcoin-dodges-split-that-threatened-its-surging-price-1510172701/. The price of Bitcoin spiked to new highs when the hard fork was called off, indicating that uncertainty about its outcome was a substantial overhang for investors. See Evelyn Cheng, Bitcoin Hits Record High after Developers Call Off Plans to Split Digital Currency, *CNBC* (Nov. 8, 2017, 12:40 pm), https://www.cnbc.com/2017/11/08/bitcoin-surges-11-percent-to-record-above-7800-after-developers-call-off-plans-to-split-digital-currency.html.

[74] See Kenneth A. Shepsle, Institutions and the Problem of Government Commitment, in *Social Theory for a Changing Society*, pp. 245, 247 (Pierre Bordieu and James S. Coleman, eds., 1991).

work create economic incentives for accurate validation of the ledger. In cryptocurrency circles, this approach is known as cryptoeconomics.[75] The blockchain is immutable because the costs of breaking it exceed the returns. By the same token, the immutability of smart contracts is imperative. The victims of the DAO hack or the Parity wallet bug were strongly incentivized to overturn the outputs of the smart contracts. They lacked the power to do so.

If, instead of approaching the beautiful sirens, Odysseus saw his boat heading directly for dangerous rocks, his cries to his men to turn the rudder would be futile. His commitment device would be operating beyond the intended scope, leading to disaster. As the three examples described earlier illustrate, the same issue appears in the blockchain context. Smart contracts cannot necessarily distinguish the scenarios for which immutability was designed from those where it causes harm. There are two fundamental reasons. Contracts of any consequence are generally incomplete; that is to say, they do not precisely specify outcomes for every possible scenario.[76] Smart contracts magnify this incompleteness. They can only express their terms in sharp-edged software code, eliminating the interpretive discretion of human judges and juries.[77]

The strong immutability of blockchain systems therefore creates significant opportunities for dramatic failures that undermine trust rather than cementing it. As Shepsle concludes: "[W]e should . . . not be too precipitous in our admiration of commitment and our condemnation of discretion."[78] To avoid causing significant harm, blockchain-based solutions must do more than enforce immutability; they must incorporate regimes of governance to temper its excesses.[79]

Blockchain Governance by Design

Blockchain is a governance technology. Consensus algorithms shape how users of networks behave. Through affirmative incentives and cryptographically enforced limits on certain actions, these systems combat hostile conduct and promote cooperative behavior. They establish and enforce rules for "good order and workable arrangements," which is how the Nobel Prize-winning economist Oliver Williamson defines governance.[80] Governance provides a framework for establishing accountability, roles, and decision-making authority in an organization.

[75] See Werbach, New Architecture, supra note 5, at 47–8.

[76] See Werbach and Cornell, supra note 23; Oliver D. Hart, Incomplete Contracts and the Theory of the Firm, 4 *J. L. Econ. & Org.* 119–39 (1998).

[77] See Werbach and Cornell, supra note 23.

[78] Shepsle, supra note 74, at 249.

[79] Permissioned blockchain and distributed ledger networks do not face quite the same challenges. Because participants are identified and authorized, traditional mechanisms of consortium governance can be applied. While not always effective, such arrangements only require the agreement of a relatively small group of organizations to make governance decisions for the network.

[80] Oliver Williamson, The Economics of Governance, 95 *Amer. Econ. Rev.* 1, 1 (2005).

Digital governance is not a new phenomenon.[81] Software code, as Lawrence Lessig famously declared and many others have elaborated since, can function as a kind of law, with its own affordances and limitations.[82] Software-based systems can serve as alternatives to the state, markets, firms, and relational contracting as means of governing relationships. Facebook's newsfeed algorithms, YouTube's ContentID system for digital right management, and Uber's mobile application are examples of digital systems that constitute and shape communities. However, these communities are centralized. The operators of the network control the algorithms and adapt them to ultimately serve their interests. Blockchain instead maintains the possibility of decentralized digital governance. By disempowering intermediaries and network operators, it promises both greater efficiency and greater fairness. Nick Szabo, one of the original developers of the idea of smart contracts, describes this property as social scalability.[83] A blockchain-based system can, it is claimed, avoid the human biases, imperfections, and inefficiencies that make it difficult for communities to scale without rigid hierarchy.[84]

FROM COMMITMENTS TO INSTITUTIONS

Blockchain governance epitomizes a broader challenge in our increasingly connected and digitized world. There is a growing gap between rule definition and rule execution. The terms of a smart contract must be specified entirely ex ante. A conventional legal contract, by contrast, is subject to relational development, the potential for mutual modification, and ex post judicial dispute resolution.[85] The case for smart contract modification can be analogized to human intervention in artificial intelligence technology. Machine learning systems produce outputs based on statistical analysis that cannot easily be traced back to their inputs, opening the door for hidden biases to creep in.[86] To avoid this issue, there is a growing consensus that humans must remain in the loop to ensure the machines avoid bias and unforeseen outputs.[87] Blockchain-based systems need something similar. The hard problem is how to reincorporate humans without forfeiting the benefits of decentralization and automation that blockchain systems promote.

[81] See, e.g., Lisa Welchman, *Managing Chaos: Digital Governance by Design* (2015).
[82] See Lessig, supra note 25.
[83] See Szabo, supra note 28.
[84] See id.
[85] See Werbach and Cornell, supra note 23.
[86] See Andrew Selbst and Solon Barocas, The Intuitive Appeal of Explainable Machines, 87 *Fordham L. Rev.* 1085 (2018).
[87] See Rachel Courtland, Bias Detectives: The Researchers Striving to Make Algorithms Fair, *Nature* (June 20, 2018), https://www.nature.com/articles/d41586-018-05469-3; Frank Pasquale, toward a Fourth Law of Robotics: Preserving Attribution, Responsibility, and Explainability in an Algorithmic Society, 78 *Ohio St. L.J.* 1243 (2017).

In the wake of the controversies of 2016–17, prominent new blockchain networks such as Tezos, Decred, and Dfinity touted their "on-chain" governance mechanisms.[88] With these systems, proposals, such as an increase in a block size, can be decided by voting of token holders, with one coin in the relevant cryptocurrency equal to one vote. The will of the majority is automatically implemented on the blockchain network.

On-chain governance is a promising area of experimentation, although it raises a host of questions.[89] For example, are those holding a majority of the cryptocurrency always the ones who should decide the fate of the network? Or what happens when, as in real-world elections, a substantial percentage of voters do not participate or lack full understanding of the issues? How might those with a vested interest manipulate the vote? Even if effective, however, on-chain governance systems are at best only one piece of the solution. Just as every possible scenario cannot be coded into smart contracts, every desirable governance action cannot be coded into a self-executing election. On-chain mechanisms cannot completely solve the problem of blockchain governance because they rely on the same immutability that generates it.

To address the governance gap, blockchain systems need credible commitments that are not absolute. This is a well-established concept. Structures that marry the security of credible commitments with the flexibility of human governance are known as institutions. The economic historian Douglass North, the great theorist of institutionalism, defined institutions as "humanly devised constraints that structure political, economic, and social interaction."[90] Institutions are voluntarily adopted constraints; that is to say, they are commitment devices.[91] As North described, the development of both public and private institutions was the defining factor in the establishment of the complex global economy. Effective institutions fused the trustworthiness of family and community ties with the social scalability needed for modern society.

Most institutions, however, are centralized. A court system or a stock market can facilitate trustworthy transactions between strangers, but those strangers must accept their authority. Is this level of trust attainable within a decentralized network? The communities around blockchain networks can effectively govern, as when the Ethereum Foundation shepherded support for the hard fork that reverted the theft of funds from the DAO. The process was somewhat chaotic, but many different interests in the community had the opportunity to be heard, several alternatives were thoroughly vetted, and in

[88] See Werbach, New Architecture, supra note 5, at 217.

[89] See Vitalik Buterin, Notes on Blockchain Governance, Vitalik Buterin's Website (Dec. 17, 2017), https://vitalik.ca/general/2017/12/17/voting.html.

[90] Douglass North, Institutions, 5 *J. Econ. Persp.* 97, 97 (1991).

[91] See Douglass North, Institutions and Credible Commitment, 149 *J. Instit. & Theoretical Econ.* 11 (1993).

the end, network nodes voted with their software whether to adopt the proposed hard fork.

However, this leads to a conundrum identified by Oxford economic sociologist Vili Lehdonvirta.[92] The theoretical problem with the blockchain practical success story is that it was a triumph of conventional governance. Respected leaders in the community debated solutions, took input, and converged on a response. As Lehdonvirta points out, this human-centric process contrasted with the vision of a decentralized, machine-centric blockchain. If trusted parties are going to make the rules anyway, who needs a blockchain, he argues. Lehdonvirta effectively rebuts the overheated claims that blockchain represents a "paradigm shift in the very idea of economic organization."[93] As incidents such as the DAO hack, the Parity wallet bug, and the Segwit2x battle illustrate, effective consensus on immutable distributed ledgers does not resolve the hard problems of governance. In some ways, it accentuates them.

Blockchain decentralization enthusiasts strike strikingly similar notes to the cyberlibertarians of the 1990s. As their poet laureate, Electronic Frontier Foundation co-founder John Perry Barlow declared: "We are creating a world where anyone, anywhere may express his or her beliefs, no matter how singular, without fear of being coerced into silence or conformity."[94] In this new world of cyberspace, he continued, governments "have no moral right to rule us nor do [they] possess any methods of enforcement we have true reason to fear."[95] We know how that story turned out. The internet has indeed been a radically empowering force. Yet many are still "coerced into silence or conformity" by governments that have found ways to overcome the internet's decentralization (such as China's Great Firewall) and, surprisingly, by the privately operated platforms such as Facebook and Google that now dominate cyberspace and its communities.

If the blockchain economy is to replicate the successes of the internet while avoiding some of its failings, governance is critical. In fact, the scope of governance must be expanded beyond its traditional domains. Here again, a comparison with internet law and policy proves enlightening.

[92] See Vili Lehdonvirta, The Blockchain Paradox: Why Distributed Ledger Technologies May Do Little to Transform the Economy, Oxford Internet Institute (Nov. 21, 2016), https://www.oii.ox.ac.uk/blog/the-blockchain-paradox-why-distributed-ledger-technologies-may-do-little-to-transform-the-economy/.

[93] Id. The quote, which appears in the introduction to the online transcript of Lehdonvirta's talk, is from Seth Bannon, The Tao of "The DAO" or: How the Autonomous Corporation Is Already Here, *Techcrunch* (May 16, 2016), https://techcrunch.com/2016/05/16/the-tao-of-the-dao-or-how-the-autonomous-corporation-is-already-here/.

[94] John Perry Barlow, A Declaration of the Independence of Cyberspace, *Electronic Frontier Foundation* (Feb. 8, 1996), https://www.eff.org/cyberspace-independence.

[95] Id.

PERVASIVE GOVERNANCE

The internet gave birth to what Shoshana Zuboff calls surveillance capitalism:[96] A global economy built increasingly on the collection, aggregation, analysis, and utilization of data related to the behaviors and intentions of individuals. "Privacy protection" online became an increasingly quaint response to the totalizing nature of information platforms. In response privacy advocates turned increasingly to an approach of totalized privacy, known as privacy by design.[97]

Privacy by design takes the position that privacy protections cannot simply be added on to technical systems. They must be built in from their inception.[98] In other words, privacy by design means more than just raising the bar for protection of personal information. As former Ontario, Canada, Information and Privacy Commissioner Ann Cavoukian explains: "Privacy must be incorporated into networked data systems and technologies, by default. Privacy must become integral to organizational priorities, project objectives, design processes, and planning operations. Privacy must be embedded into every standard, protocol and process that touches our lives."[99] The implementation of this vision in legislation and business practice has left something to be desired, but the premise is sound.

Something similar, call it governance by design, should be incorporated into the development and oversight of blockchain-based systems.[100] Given the structure of blockchains, governance cannot be an afterthought. Neither can it be limited to

[96] See Shoshana Zuboff, Big Other: Surveillance Capitalism and the Prospects of an Information Civilization, 30 *J. Info. Tech.* 75–89 (2015); Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (2019).

[97] It is now widely accepted by policymakers in a variety of contexts, most notably the European General Data Protection Regulation that went into force in 2018. See Lee Bygrave, Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements, 4 *Oslo L. Rev.* 105 (2017).

[98] See *Privacy-Enhancing Technologies: The Path to Anonymity, Vol. 1* (Aug. 1995), http://www.ontla.on.ca/library/repository/mon/10000/184530.pdf; Peter Schaar, Privacy by Design, 3 *Identity in Info. Soc.* 267 (2010).

[99] Ann Cavoukian, Privacy by Design: The 7 Foundational Principles, *IAB.org* (2009), https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf. Cavoukian was the first to develop the concept of privacy by design.

[100] A few other authors employ the same term, but generally in different ways than I do here. Deirdre Mulligan and Kenneth Bamberger use "governance-by-design" to refer to efforts to promote value or implement regulatory mandates through manipulation of technological systems. Deirdre K. Mulligan and Kenneth Bamberger, Saving Governance-by-Design, 106 *Calif. L. Rev.* 697 (2018); Primavera de Filippi uses "governance by the design" for self-governance models in which "rules are embedded directly into the underlying technology of the platforms they use to operate." Rachel O'Dwyer, Commons Governance and Law with Primavera De Filippi, *Commons Transition* (July 31, 2015, 10:55 AM), http://commonstransition.org/commons-centric-law-and-governance-with-primavera-de-filippi/. Embedding governance structures into code, such as through the on-chain governance technology of systems such as Tezos, is a subset of what I propose here. Governance by *design* means systematically embedding governance into all relevant processes, whether implemented in software, in regularized procedures for human discussion, or in decisional structures.

formalized voting on changes to network algorithms. Voting structures insufficiently address the diversity of governance challenges that can arise, as highlighted by the three examples provided earlier.

In the blockchain context, governance by design means recognizing that perfect immutability creates systems with unacceptable fragility. They work well until they don't, and then they have no good means to recover. Advocates of strong immutability see an inherent tradeoff in which flexibility to human decision-making undermines decentralization.[101] However, if we want solutions that can resolve unexpected problems smoothly, we must trust someone to resolve them.

Incorporating governance by design principles, rather than bolt-on governance functionalities, counters this tradeoff. As Cavoukian argues in the analogous context of privacy by design: "Privacy is often positioned in a zero-sum manner as having to compete with other legitimate interests, design objectives, and technical capabilities, in a given domain. Privacy by Design rejects taking such an approach – it embraces legitimate non-privacy objectives and accommodates them, in an innovative positive-sum manner."[102] Governance by design can have a similar effect by incorporating governance as a baseline function at every level, not a "get out of jail free" override.

In her work on common-pool resource systems, Nobel Prize-winner Elinor Ostrom emphasizes that governance is polycentric and hierarchical.[103] Multiple governments, as well as private mechanisms, may shape the management of a resource or community. Ostrom describes three levels of rule: Operational, collective-choice, and constitutional-choice.[104] Operational governance addresses the day-to-day issues that directly affect a given system. Collective-choice governance determines two things: Who can take certain operational actions and who can change operational rules. Constitutional-choice governance determines who has the authority to change collective-choice rules. A system that works for mundane problems will not necessarily address unusual situations that require extraordinary override. And a system for addressing particular crises will fail to resolve fundamental disagreements about the direction of the community.

A starting point for thinking about governance by design in a blockchain context would be to recognize four hierarchical domains:

1) *Consensus.* Analogous to Ostrom's operational rules, the consensus algorithms of a blockchain network promote honest verification and agreement on status

[101] See Jordan Pearson, The Ethereum Hard Fork Spawned a Shaky Rebellion, *Motherboard* (July 27, 2016, 5:55pm), https://www.vice.com/en_us/article/z43qb4/the-ethereum-hard-fork-spawned-a-shaky-rebellion-ethereum-classic-etc-eth.

[102] Cavoukian, supra note 99.

[103] See Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (1990).

[104] See id. at 52.

of the ledger. In the normal mode of day-to-day operation, the dynamics of the consensus mechanism determine the attributes of the blockchain network. Discussions of blockchain technology as "governance by code" or a new "Lex Cryptographica"[105] generally focus on the consensus layer, which is where transactions are designed to be immutable.

2) *Override*. When immutability produces problematic results, as in the case of the DAO hack, override governance offers a means to reverse immutability by establishing decision-making power at the outset. This is analogous to the first sense of Ostrom's collective-choice rules, in that they define who has decision-making power in such situations. The Ethereum community struggled in responding to the hack because it was not clear who should be part of the decision-making process, and how a consensus of decision-makers should be implemented.[106]
3) *Rule Change*. Bitcoin's Segwit2x fight concerned a general property of the network: The size of blocks. As in Ostrom's constitutional-choice layer, governance here requires a means of determining who sets policy for the network. In the Segwit2x case, groups in the community such as exchanges, miners, users, and core developers had differing views. There was no good mechanism to resolve these views given insufficient structures and norms of governance.
4) *Community Governance*. Ostrom's constitutional-choice layer is about who judges the judges: How the entities empowered to participate in governance and change the rules are constituted. This is often a blind spot in blockchain networks. For example, the launch of Tezos was delayed when the organization developing the software had a conflict with the foundation designed to oversee the network after the project raised over $200 million in a token offering.[107] The irony that a system designed to automate rule-change governance struggled at community governance was not lost.

This high-level framework is just a starting point for blockchain governance by design.[108] There will be many practical decisions to make in any network. While

[105] De Filippi and Wright, supra note 25.

[106] On-chain governance systems are one approach to this challenge. See supra notes 88–9 and associated text. Another is to structure decentralized arbitration mechanisms that leverage the incentive mechanisms of cryptocurrencies themselves. See Werbach, New Architecture, supra note 5, at 215–16 (describing Augur's "computational juries" as an example).

[107] See Paul Vigna, Tezos Raised $232 Million in a Hot Coin Offering, Then a Fight Broke Out, *Wall St. J.* (Oct. 18, 2017, 12:07 AM), https://www.wsj.com/articles/tezos-raised-232-million-in-a-hot-coin-offering-then-a-fight-broke-out-1508354704.

[108] As Mulligan and Bamberger point out, embedding policy prescriptions directly into technological systems can create new problems. Governance by design may fail to address specific cases in a nuanced way, deprecate important human rights values, or reduce transparency of the regulatory process. These considerations should be taken into account in designing blockchain governance mechanisms. They offer a series of recommendations to ameliorate the dangers of governance by design, which are consistent with the layered, polycentric approach proposed here. See Mulligan and Bamberger, supra note 100.

governance and decentralization are not fundamentally in conflict, there is room for different workable tradeoffs dependent on either the goals of the network or the culture of its community. The different ways the Bitcoin and Ethereum communities addressed the Segwit2x hard fork and the DAO hack, respectively, illustrate that both processes and norms play a role in solving for decentralized issues.

The final important factor that Ostrom's polycentric framing emphasizes is that private self-governance and public oversight through sovereign governments are not necessarily in conflict. Her classic study of common-pool resources, *Governing the Commons*, identifies several cases in which the state facilitated private ordering and the creation of community-based institutions.[109] The developers of blockchain networks often begin with a strong resistance to government involvement, just like the pioneers of the internet economy. However, as became clear in the development of the internet, governments can do much more than ban or allow technological innovation.[110] As just one example, the need for strong government-issued network neutrality rules became a rallying cry for advocates of the open internet, as a check on the power of broadband access providers.[111] There are similar calls today for the state to intervene in order to break the stranglehold of large digital platforms such as Google, Amazon, and Facebook.[112] We should not ignore the ways in which government might contribute to the health of the blockchain economy.

CONCLUSION

At this early stage in blockchain development, the adoption path of the technology is quite uncertain. Despite the spike in the price of cryptocurrencies, usage for payments, Bitcoin's original purpose remains limited.[113] Many enterprise blockchain pilots built on specialized cryptocurrency models have failed to see the rapid adoption their boosters predicted.[114] However, blockchain technology itself will continue to see investment and development because it addresses fundamental challenges in organizational recordkeeping and the need for interorganizational trust.[115] Further, there are major applications of the approach such as trading

[109] See Ostrom, supra note 96, at 212.

[110] See Kevin Werbach, The Song Remains the Same: What Cyberlaw Might Teach the Next Internet Economy, 69 *Fla. L. Rev.* 887 (2017).

[111] See id. at 915–16.

[112] See, e.g., Jonathan T. Taplin, *Move Fast and Break Things: How Facebook, Google, and Amazon Cornered Culture and Undermined Democracy* (2017); Lina M. Khan, Amazon's Antitrust Paradox, 126 *Yale L.J.* 710 (2017).

[113] See Paul Vigna, People Love Talking about Bitcoin More Than Using It, *Wall St. J.* (Apr. 12, 2017, 5:30 AM), https://www.wsj.com/articles/people-love-talking-about-bitcoin-more-than-using-it-1491989403.

[114] See Olga Kharif, Blockchain, Once Seen as a Corporate Cure-All, Suffers Slowdown, *Bloomberg* (July 31, 2018), https://www.bloomberg.com/news/articles/2018-07-31/blockchain-once-seen-as-a-corporate-cure-all-suffers-slowdown.

[115] See Werbach, New Architecture, supra note 5, at 226, 236–7.

markets in cryptoassets that seem poised for continued growth even if they do not disrupt traditional markets.[116] Nonetheless, it is far from certain that any blockchain network will achieve the scope and influence of Google, Facebook, Amazon, Tencent, and Alibaba, let alone realize the grand visions of societal disruption that boosters promulgate.

The importance of blockchain governance, however, does not depend on any particular story of blockchain adoption. Blockchain has proved to be a governance technology that seeks to balance on the knife edge of freedom and constraints. That challenge is as old as civilization. In working to overcome this challenge, we can learn from the ways that blockchain networks try – or don't try – to resolve the implicit tensions of immutability. Both theory and practice must play a role. There is no shortcut to designing governance mechanisms, watching how they operate in practice, and iterating based on their shortcomings.

Appropriately, that is also the lesson of Odysseus' encounter. Odysseus has himself tied to the mast so that he, alone, can hear the song of the sirens in safety. What do they sing that is so tempting? The sirens offer a shortcut to knowledge: "For lo, we know all things, all the travail that in wide Troy-land the Argives and Trojans bare by the gods' designs, yea, and we know all that shall hereafter be upon the fruitful earth."[117] The seductive appeal of the sirens is the promise of wisdom without experience, just as the seductive appeal of the blockchain is trust through cryptography and economic incentives without human governance. Believing too strongly in either leads to disaster. Finding the proper balance is the road to valuable insight.

[116] See Kevin Werbach, Blockchain Isn't a Revolution, *Medium* (June 18, 2018),https://medium.com/s/story/blockchain-isnt-a-revolution-it-s-two-big-innovations-and-one-promising-idea-988fca6bofca; Shawn Tully, The NYSE's Owner Wants to Bring Bitcoin to Your 401(k). Are Crypto Credit Cards Next?, *Fortune* (Aug. 14, 2018), http://fortune.com/longform/nyse-owner-bitcoin-exchange-startup/.

[117] The Odyssey, supra note 2 (Book XII).