JUSTICE, HEALTH & DEMOCRACY
IMPACT INITIATIVE

ETHICS OF DECENTRALIZED SOCIAL TECHNOLOGIES:

Lessons from Web3, the Fediverse, and Beyond

Danielle Allen
Eli Frankel
Woojin Lim
Divya Siddarth
Joshua Simons
E. Glen Weyl

March 2023

Ethics of Decentralized Social Technologies: Lessons from Web3, the Fediverse, and Beyond

AUTHORS

Danielle Allen
James Bryant Conant University Professor
Harvard University

Divya Siddarth
Collective Intelligence Project
Microsoft Research

Eli Frankel
Undergraduate Fellow
Edmond J. Safra Center for Ethics
Harvard University

Joshua Simons
Former Post Doctoral Fellow
Edmond J. Safra Center for Ethics
Harvard University

Woojin Lim
Undergraduate Fellow
Edmond J. Safra Center for Ethics
Harvard University

E. Glen Weyl
RadicalXChange Foundation
Microsoft Research

JUSTICE, HEALTH & DEMOCRACY
IMPACT INITIATIVE

The Justice, Health, and Democracy Imapct Initiative is a multi-institutional collaboration between New America, the Brown University School of Public Health, and the Edmond & Lily Safra Center for Ethics at Harvard University.

CONTENTS

> “To date, little formal work has been done to build ethical frameworks for current and future experimentation with decentralized social technologies. This is a dangerous state of affairs.”

1. Introduction[1]

New eras of technological innovation enable radical experimentation, creating breaks in the social structure with great risks and rewards. In the twentieth century, communication and transportation technologies reshaped social, political, and economic institutions. Already in the twenty-first century, innovations such as mobile telephony and social media have done the same.

We are now living through another such shift. The plethora of experiments with decentralized social technologies (DSTs)—clusters of which are sometimes called “the Web 3.0 ecosystem” or “the Fediverse”—have brought us to a constitutional moment. Here we find it useful to use a single term, DSTs, to capture, hopefully in a fairly neutral way, the full range of recent innovation (also including distributed computation and analysis paradigms, like Federated Learning) that aim to harness technology to address the perceived concentration of power in digital technology.

These technologies enable radical innovations in social, economic, and political institutions and practices, with the potential to support transformative approaches to political economy. They demand governance innovation. There is the potential to overcome persistent injustices, power concentrations, and perversions of capitalism and democracy. In fact, recent advances in artificial intelligence (AI) may make these tools critical to preserving human dignity, agency, and even existence. Yet there are also risks of catastrophe and oppression that eclipse those seen in the twentieth century. Caliber of governance will determine which path we find ourselves upon.

Recent events dramatize these possibilities. On the one hand, the increasing prevalence of highly persuasive machine-generated content (e.g., ChatGPT) makes increasingly urgent the impetus for the kind of cryptographic verification on which the Web3 ecosystems has focused. On the other hand, the recent rapid collapse of FTX underscores the urgency of establishing governance regimes for new technologies in this space. One of the most striking features of the collapse is how those who purported to be introducing new governance protocols and methodologies failed at the most rudimentary elements of firm governance, demonstrated by an absence of records, accounting controls, clear decision-making structures, and the like. Existing understandings of governance need to be merged with clarity about new governance challenges and a readiness to innovate to meet those new challenges.

What’s more, many of the utopian and dystopian promises of decentralized social technologies

1 We thank those who gave feedback on earlier versions of this paper without attributing to them agreement with our arguments: Sarah Hubbard, Kevin Owocki, Ariel Procaccia, Jeffrey Saviano, Shlomit Wagman, and Ethan Zuckerman.

are hypothetical—they are projections about possible futures that the underlying technologies might support. What matters is the choices we make about how these technologies are designed and used over the coming months, years, and decades. That's why we must be intentional about the ethical and governance frameworks we develop to structure social experimentation with these technologies. How and whether we apply these frameworks will shape whether choices about the design and use of decentralized social technologies will simply entrench the challenges humanity is already facing—failures in democratic governance, financialization and massive inequality, the neglect of the provision of public goods and infrastructure, and limited collective responses to pressing crises—or whether these technologies will help build a plural but connected and democratic future in which every person has the opportunity and power to flourish.

To date, little formal work has been done to build ethical frameworks for current and future experimentation with decentralized social technologies. This is a dangerous state of affairs. This paper seeks to develop a framework for making ethical choices in this space that help to both grasp positive opportunities for transformation and avoid the potentially problematic consequences. Most of our specific examples and concerns come from the blockchain/Web3 universe, as this has received the greatest investment, attention, and adoption to date. However, we aim to offer a framework for governance decision-making in conditions of uncertainty that applies more broadly to other DSTs.

We propose a pragmatic, democratic, and pluralist approach to navigating bold experimentation with social practices and political economy enabled by these technologies. Our overarching goal is to provide a framework open to transformative improvement and constrained by guardrails and guiding values supportive of democracy, freedom, and pluralism. We take a relatively strong position, rather than simply laying out ethical issues and potential approaches. We seek to be provocative in order to spur further work, and hope this paper will serve as a first bridge between academic philosophy and the DST community, which have hardly interacted to date.

2. The Perils and Promise of Constitutional Moments

2.1 Context

The question of whether the innovations in the DST ecosystem, and of blockchain specifically, rise to the level of a constitutional moment is fraught with controversy. To point to the potential for transformation inherent in these technologies seems to credit them with a seriousness that the clear presence of fraud and speculation would seem to caution against.[2] Yet even in a context where an innovation is susceptible to abuse or misuse, the potential for significant social impact may remain.[3] To see the stakes, we need to clarify what we mean by the idea of a "constitutional moment," a phrase that we would apply equally to the rise of both social media and DSTs.

2.2 Defining a Constitutional Moment

When the U.S. Constitution was debated and written, it rested on a social platform structured by the economic, demographic, and technological facts of the late eighteenth century. That social platform has since been repeatedly transformed—by the Industrial Revolution, air travel, the biomedical revolution, broadcast media, and the Internet. These shifts in social platforms have prompted, and sometimes demanded, successive efforts to redesign the political institutions

2 According to the Consumer Financial Protection Bureau, between October 2018 and September 2022, 40% of consumer complaints related to crypto-assets appeared to be "fraud and scams." The Federal Trade Commission reports that over 46,000 people lost more than $1 billion on crypto trading to scams and fraud between January 1, 2021, and March 31, 2022. Academic literature on cryptocurrency fraud has focused on the growing frequency of Ponzi schemes and high yield investment programs (HYIPs), scams involving initial coin offerings, phishing scams, pump-and-dump schemes and market manipulation, exchange scams and scam wallet services, smart contract honeypots and attacks, mining malware, securities fraud and identify theft. See Consumer Financial Protection Bureau, "Complaint Bulletin: An analysis of consumer complaints related to crypto-assets"; Fletcher, "Reports show scammers cashing in on crypto craze"; Trozze, Kamps, Akartuna, et al. "Cryptocurrencies and Future Financial Crime."

3 Seyedsayamdost and Vanderwal, "From Good Governance to Governance for Good: Blockchain for Social Impact."

that have been built on them. The Industrial Revolution helped to dispatch enslavement, thereby changing the very definition of the U.S. citizenry. The invention of global governance bodies in the wake of World War I and World War II was unthinkable absent air travel. Around the world, much longer life expectancies and more powerful health resources have transformed expectations about the social goods governments must deliver.

Legal theorist Bruce Ackerman popularized the phrase "constitutional moment" to refer to points in U.S. legal history where "higher lawmaking" with constitutional significance occurs.[4] But the phrase has also been used by other scholars to describe how changes in the interaction between law and underlying social structure can bring about constitutive reorganizations.[5] Internet-driven changes in the social platform on which political institutions rest have been profound, bringing constitutional implications even if legal structures themselves have not yet shown the signs of constitutional-level changes.

A constitutional moment, in other words, involves transformations in: (1) political institutions, (2) the structure of civil society, (3) economic organization, (4) common understandings of political institutions/civil society/economic organization, and/or (5) the apparatus of, and social relationships embedded within, the state. Sometimes these transformations also produce changes in written constitutions themselves (e.g., in the U.S. in the period of the Civil War and the Progressive era), but sometimes they don't.

For instance, social media has made it possible for people with extreme views to find each other and coordinate political action across great distances, despite geographic dispersal. Yet the U.S. founders argued that geographic dispersal would help moderate ideological extremes, by forcing people to channel their opinions through elected representatives to get them into the public sphere. The capacity social media provides for enabling those with extreme views to coalesce and coordinate across great distances has undermined the institution of representation itself, removing geography as a forcing factor that activates representatives as mediating, synthesizing, and often moderating forces. The emergence of social media thus makes imperative the development of new institutions that can serve these roles. This fundamental impact flowing from social media, for instance, rises to the level of a "constitutional moment."

DSTs promise similarly fundamental impacts. As Wright and De Filippi argue, Web3 and blockchain culture specifically center around the building of a *lex cryptographica* on top of and around existing legal infrastructures.[6] The potential of accelerated AI development to undermine many traditional elements of legal

4 Ackerman, *We the People*: Foundations, 3, 5.

5 Allen, *Talking to Strangers*, 5–8, 193.

6 De Filippi and Wright, *Blockchain and the Law: The Rule of Code*.

and governance systems (e.g., making it easy to create perfectly undetectable false IDs) makes the relevance of such a regime increasingly clear. Even if we do not see a flurry of formal higher law-making in years ahead, in practice this is already occurring.

Because there is so much skepticism specifically that blockchain is a fundamentally transformative technological innovation, we focus now on why we think these technologies, specifically, are likely to have fundamental impacts.

2.3 Fundamental Impacts

From the nineteenth to the twenty-first century, humans have transitioned from horse and buggy days to the age of the electric vehicle. Many other technical, economic, and social changes are connected to and captured by the example of that transition. To understand the impact of Web3 and other DSTs, it can be useful to itemize other "horse and buggy" social technologies—social technologies that have been operative for generations with little change and that are about to undergo transformations that will have important downstream consequences.

1. ***Titling/ Validation.*** Anyone who has bought a car or house knows that their ownership is reflected in the form of a document called a "title" that is recorded with a public recorder of deeds. This practice of "titling" objects to define ownership undergirds our property system. It is a practice that, in its current mode, dates back to the medieval period and the formation of early modern English law. Blockchain validation methods constitute an alternative set of practices and methods for determining title, potentially divorcing the role of 'protection of private property' from public offices. Where and how those new methods will come to intersect with and/or replace the "horse-and-buggy" methods that have been operative for centuries is unclear, though the increasing ease of using AI to perfectly replicate titles suggests that more secure and transparent approaches to validation will have increasing relevance and any new method that gains significant adoption for doing this foundational work is bound to be transformative.

2. ***Notaries public.*** Notaries public anchored the legitimacy of traditional titling institutions. They have also been used to anchor and secure the use of government issued identification systems. In the Web3 universe, "miners" and "validators" have replaced "notaries public," with various computation, digital assets, or trust frameworks replacing notary certifications. The question of who the miners are, what legitimates their acting like notaries, and the degree of coordination among them, as well as their accountability to legal systems, will have a significant impact on whether, how, and to what degree an alternative digital-native approach to validating transactions and identities comes to replace pre-digital variants of this practice. These questions are actively in play, for example, in the competition between "proof of work," "proof of stake," and other more sophisticated consensus protocols, such as Stellar.[7]

3. ***Identification systems.*** Government identification systems are still, for the most part, grounded in bureaucratic processing of a thin collection of social signals, such as a birth certificate signed by a doctor and parents and marriage certificates signed by religious authorities. Such signals are increasingly becoming forgeable by advanced AI systems.[8] Digital identity systems have thus moved far beyond this, drawing on a wide range of social, transactional, and relational signals that were once available only to security agencies at great cost. They have, however, done so in close proprietary settings and at great cost to norms of autonomy and privacy. As DSTs open a public and increasingly sophisticated set of cryptographic tools that enable a range of stakeholders to issue and verify relationships and credentials, it seems likely that these will supplement or even replace extant government identification regimes and, in the process, greatly unsettle capabilities and expectations around the role of identification.

7 Mazières, "The Stellar Consensus Protocol: A Federated Model for Internet-level Consensus."

8 King et al., "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions."

4. *Transferring Funds.* Global transfers of funds have historically been complex and required significant investment of effort or institutional resources. The universe of cryptocurrencies has changed that dynamic; blockchains today process more transactions than credit cards, with little regard to national borders. A fundamental reorganization in the functioning of liquidity is bound to have significant social and economic consequences.[9]

5. *Media and social media.* Centralized social media platforms have already disrupted the landscape of traditional media, publishing, communication, and association, with society-wide impacts. The rise of decentralized and federated alternatives will present additional challenges. For example, the already difficult questions raised by centralized content moderation become more complex in decentralized or federated contexts, in which oversight and auditing of information flows becomes difficult or impossible. These issues may exponentiate in a future of low-cost, multimodal-generated content produced by models such as GPT-4, indistinguishable from human-generated content.[10]

> "As DSTs open a public and increasingly sophisticated set of cryptographic tools that enable a range of stakeholders to issue and verify relationships and credentials, it seems likely that these will supplement or even replace extant government identification regimes and, in the process, greatly unsettle capabilities and expectations around the role of identification."

6. *Organizational formation, financial management, and governance.* Legal frameworks from one country to the next currently provide a variety of different models for the formation and incorporation of collective enterprises. Existing models for organizational formation depend to some extent on the reality of technological possibilities for validating identities and carrying out specific transactions, such as check signing; the need for efficiency, for instance, has limited the number of signatories it makes sense to have on organizational accounts and complicated the ownership of collective assets, driving the dominance of a small set of models of co-ownership (usually corporate or specific forms of non-profits). Blockchain enabled technologies for validating identification, completing transactions, and collective asset ownership change underlying efficiency dynamics and therefore unlock new organizational possibilities, the most prominent name for which is (misleadingly) "distributed autonomous organizations" (DAOs). Again, where these possibilities will lead is unclear.

7. *Governance and voting.* A core part of organizational form is governance. Democratic participation and voting have, as we noted above, long been rooted in locality and government-issued identifications. Yet with new organizational forms and identification technologies, it is increasingly possible to realize John Dewey's 1927 vision of emergent, geographically dispersed polities emerging to govern topics of common interest and concern, responding in near-real time to patterns of commerce and sociality created by technology.[11] Building such networked polities, and exploring their potential impacts on

9 Investing.com, "Ethereum Processes 4.5x More Transactions than Visa by CoinEdition."

10 Metz and Weise, "A Tech Race Begins as Microsoft Adds A.I. to Its Search Engine."

11 Dewey, *The Public and Its Problems.*

nation-state structures, has become a central goal of many. Furthermore, advances in cryptography and secure digital voting necessary to empower such experiments are already affecting the practice of democracy in traditional nation-states and allowing new mechanisms of participation, involving sophisticated mathematics or natural language processing software, that would have been impossible with paper and pencil. The hard infrastructure of past eras—and existing jurisdictional boundaries—may or may not interact productively with the new and fluid boundaries of emergent polities.

8. Rewarding digital labor and clarifying data rights. The Internet era has brought an unusual economic model where corporations have been able to reap great financial rewards from the uncompensated participation of users of their platforms. Blockchain technologies bring new possibilities for rewarding digital labor, and an opportunity for platform users to change the balance of economic power defining the operations of digitally native firms. Similar impacts may be expected around the social and economic rights to data.

9. Property Relationships. The changes described above touch fundamental elements of the property system that has operated in Western economies for several centuries. As the concrete possibilities of these technologies become clearer, we should expect shifts in some basic features of the structure of property relationships. This will have significant implications for political economy and social structure. For example, De Soto famously highlighted the mismatch between formal titling practices and the structure of property regimes in many developing countries.[12] Yet reforms instituted in response to his work have often forced the round peg of traditional practices into the square hole of property systems imported from industrial era Western Europe. DSTs tools offer the possibility of encoding and making legible, computable, and thus suitable for a modern and dynamic society a much richer range of property regimes, allowing both greater accommodation of traditional practices and a range of innovation.

10. Dispute Resolution Procedures. The relationships already coming into existence via blockchain technologies have generated and will continue to generate disputes about control of significant resources. Such dispute resolution has historically been the province of sovereign governments, typically through judicial systems. New dispute resolution mechanisms are evolving and will evolve to resolve disputes arising from blockchain ecosystems and the nature and site of these mechanisms will help determine the future of sovereignty.

The kinds of changes that blockchains and other DSTs have introduced and will introduce in the future may reshape the social platform on which many of our current political and economic institutions rest, prompting similarly radical shifts in how human societies operate. This may fuel the next phase in the constitutional moment that digital technologies have introduced. This does not mean that particular societies need new written constitutions, but it does mean that the choices we make about technical and social design will be far-reaching. As the undergirding social-technical platform shifts, we need a framework for navigating this constitutional moment.

2.4 Peril and Promise

Taken together, the shifts identified above could produce a wholesale restructuring of the political economic order. A range of outcomes are possible, from the dystopian to the desirable.

First, to the dystopian. One can imagine a world of anti-democratic, atomized market supremacy, in which social and political relationships are increasingly mediated through financial transactions. This classic cypherpunk dystopia was most famously sketched in Stephenson's 1992 science fiction novel *Snow Crash* and then advocated as inevitable in the 1999 book by Davidson and Rees-Mogg, *The Sovereign Individual*.

Such a world could arise as follows. First, a new class of private, for-profit organization begins to emerge outside of existing institutions, governments, and community structures. They begin

12 de Soto, *The Mystery of Capital*.

to perform useful tasks for large swathes of humankind, such as providing identity, payment systems, data transfer pipes, and organizational structuring. They then move into privatized infrastructure and areas like education and housing. These centralized but private intermediaries become a necessary entry point to daily life.

Internal governance of those entities through one-coin-one-vote protocols gives rise to plutocratic networks with increasing power, producing a new class of oligarchic intermediaries that erode the public sector. Tax revenue falls precipitously, and race to the bottom dynamics are evident as corporations move fully digital operations to jurisdictions with favorable terms, such as the erosion of labor protections and regulation. Funds for public investment are rerouted towards the protection of private property rather than the provision of public goods; at an extreme, collective security and policing are replaced by subscriptions to the services of mafias. Existing, highly imperfect democratic processes prove impotent to resist and are thus abandoned by increasingly cynical citizens. Illegal and anti-social activity not only flourishes but scales up, as criminals are able to invent and proliferate ever more powerful weapons with impunity. AI-enabled models of automated decision-making, investment, or content generation provide a powerful engine for the increasingly autonomous activities of these entities.

Economic security is replaced by wild speculation and dangerously exploitative work conditions as the main form of income for many. The increased political power of new institutions hastens this trend towards 'exitocracy,' in which the only agency individuals have exists in the form of high-cost transfer between oligarchically-controlled and conflictual "walled gardens." One does not have to look far for this world: so-called Bitcoin maximalists and adjacent constellations of blockchain investors have eagerly advocated hastening the advent of the world imagined by Davidson and Rees-Mogg. In fact, the latter's son is a prominent British politician and the book's greatest promoter, Balaji Srinavasan, has recently published an updated handbook to its implementation, *The Network State*, which has proved a runaway best seller due to his following in the Web3 world.[13]

Even a less dramatic move in this direction could significantly magnify the problems of the status quo: in the digital realm, disinformation, polarization, speculation, and digital serfdom, and in the physical, rampant inequality, stagnant growth, democratic failure, and climate crisis. In an era characterized by the rise of techno-authoritarianism and the failures of democratic capitalism, the injection of new mechanisms for coordination and communication are likely to magnify existing trends, rather than support new possibilities for social, political, and economic organization. The direction that expanded technological capacity drives tends to be shaped by existing political-economic tendencies and pathologies.

But these are not the only possibilities. One can also imagine radical, positive, and perhaps even existentially necessary transformations enabled by the above shifts. In his classic 1927 treatise, *The Public and its Problems*, John Dewey offered a response to the potential for technological advance to reshape patterns of human interactions in ways that capitalist organization can neither optimize nor hinder from producing catastrophically negative effects.[14] The dangers are clear: Industrialism turns out to have set in motion potentially catastrophic climate change that, left to the devices of capitalism, will only accelerate. Advances in communications and transportation have created affinity groups scattered around the globe but united by shared values and infrastructure. In the current era, the lossy optimization incentives of capitalism may be further magnified by the introduction of carelessly deployed, reward-driven AI algorithms with far greater catastrophic effects.

But Dewey argued that the benefits of technological progress could be harnessed and its harms avoided if new polities could emerge to manage these new patterns of interaction democratically. His argument was that consistent accountability to "the people" would itself serve as a necessary regulator on decision-making, guiding technology in beneficent directions and away from catastrophic tendencies. His argument is similar to that of Amartya Sen who makes the case that one of the greatest values of democracy is its constructive capacity to

13 Srinavasan, *The Network State: How to Start a New Country*.

14 Dewey, *The Public and Its Problems*.

deliver corrective self-regulation for a dynamic human system that pulls the system always back in the direction of the provision of public goods.[15] Danielle Allen has recently amplified and extended both of these arguments for the twenty-first century in *Justice by Means of Democracy*. The central question mooted across these texts is where and how to enable patterns of democratic governance, such that the whole of a social system is democratically steered. The goal is human freedom, yes, but also the epistemic resources of inclusive, effective social choice, elevated as a decision-making mechanism because of its potential to steer societies toward human flourishing.[16] The transformative potential of DSTs is to support the discovery of innovative governance mechanisms and innovative ways of situating democratic governance within decision-making procedures to improve the steering of socio-technical systems toward public goods.

If that is the political utopia we can imagine, there is also a social utopia awaiting. Just as technology brings dangers for governance and the economy, so too there is a dark side to growing social complexity. The founder of social network theory, Georg Simmel, in his 1908 *Soziologie*, described how modernity and urbanity, by proliferating and diversifying social relations, fragmented and individuated identities. Where once the social circles in which citizens worshiped, worked, socialized and so forth heavily overlapped, in modern urban societies these aspects of life differentiate, creating "individuals" as the unique intersection of their social groups.[17] As Simmel anticipated in the 1908 text, such intersectional identity offers opportunities for freedom and growth, but also threatens isolation of individuals who lose track of the social spaces they navigate. As Kimberlé Crenshaw highlighted in her 1994 article, "Demarginalizing the Intersection of Race and Sex," the intersectional nature of identity can also easily entrench the marginalization of those who suffer compounding social disadvantage and oppression by undermining their capacity to organize and be recognized collectively.[18] But here again, DSTs have the potential to overcome some of these harmful tendencies, making it easier to track and convene emergent affinity groups across physical and social distances and to allow individuals to harness technology

> "Responsible technology seeks transformative improvement toward human flourishing constrained by guardrails and guiding values supportive of democracy, freedom, and pluralism."

to manage and navigate the complexity of their sociality. In both political and social spaces, it is possible to imagine a transition from descriptive to emergent representation. On the former, social protocols are designed to try to "fairly" or "accurately" capture social difference that is believed by technocrats to exist. On the latter, social protocols enable the self-governing emergence into visibility of diverse and fluid social groupings. Notions of "value alignment" core to the AI imaginary are only coherent in the context of value pluralism; perhaps Web3 technologies can provide necessary technological primitives for this pluralism. Algorithms and mechanisms could be systematically designed to encourage non-domination in the presence of and cooperation across these increasingly dizzying social differences, enabling a far richer nation of egalitarianism than we've even had the capacity to describe in the past.

Visions such as these are not simply a speculation but have been an aspirational pathway for many central technologists who have helped build the Internet. As Mitch Waldrop documents in his 2001 *The Dream Machine*, social theories

15 Sen, "Democracy as a Universal Value.".

16 Farrell and Shalizi, "Pursuing Cognitive Democracy."

17 Simmel, *Conflict* and *The Web of Group-Affiliations*.

18 Crenshaw, "Demarginalizing the Intersection of Race and Sex."

(like those of Dewey and Simmel, and others building on them) that saw networks of social interactions as central to modernity were core inspirations to the technologists, academics, and government bureaucrats who founded the ARPANET that evolved into the Internet.[19] Perhaps the leading figure among these was J. C. R. Licklider, the program officer at the US Department of Defense's Advanced Research Projects Agency (ARPA) who initially funded ARPANET. In 1979, as the TCP/IP protocol began to cohere, Licklider highlighted both the democratizing potential of networking and the danger that the development of open standards would stop with basic communications protocols and allow systems of financial transactions, identification, participation, and the like to be captured by private monopolies.[20] On the utopian picture, Web3 is imagined as a movement to follow through on this Licklider vision of building the technological substrate for the "network society" of Dewey and Simmel.

The utopian vision of technology supporting a plural network society also has its science fictive embodiments, most prominently represented in the Star Trek television universe originally envisioned by Gene Roddenberry. Roddenberry depicts the potential for technology to gradually bridge while also celebrating and proliferating social differences, a vision he puts into the mouths of the Vulcan people as the philosophy of "Infinite Diversity in Infinite Combinations."[21] Especially in later elaborations (such as Deep Space Nine and Picard), the series explores how this state was realized through a series of technologically empowered tools for bridging social divides and overcoming both material scarcities and forms of domination.

Even more modest moves in this direction hold the potential to address important crises of our time. Taiwan's digital minister, Audrey Tang, and the g0v civic technology movement she founded have already harnessed techniques like these to help Taiwan lead the world in its response to the global pandemic, confronting misinformation online, addressing environmental challenges, and much more.[22] It is hard to imagine how we survive, much less thrive, as a species on this planet without the sort of bold improvements to our capacity for coordination across diversity that these tools may empower.

Our task in this paper is not to make predictive arguments about whether the dystopian or utopian trajectory is more likely. Both are live possibilities. These utopia and dystopias are hypothetical and we do not need to take as inexorable most claims about the possibilities/direction of the underlying technologies. They will be shaped by human action and human choices, and those in turn will be shaped by institutions. Probably aspects of both will emerge in the years ahead. Instead, our goal is to describe an ethical framework to guide experiments with DSTs that will, over time, shape which trajectory in fact unfolds. To develop this framework, we draw first on pre-existing examples from the arena of AI ethics and bioethics, looking both at where previous work has gone wrong and where it goes right.

3. From AI Ethics to Ethics for Decentralized Social Technologies (via Bioethics)

The ethical challenges introduced by DSTs—whether via blockchain or the Fediverse or another technology—may be new in the specifics of their substance, but they are not different in kind from ethical challenges produced by other moments of technological revolution. As the field of ethics for decentralized social technologies is in its infancy, with few publications to date, we can glean important insights from both AI ethics and bioethics. Bioethics as a field has been under development for fifty years, and AI ethics for twenty.[23]

Similar concerns motivated the creation of these

19 Waldrop, *The Dream Machine: J. C. R. Licklider and the Revolution That Made Computing Personal*.

20 Licklider, "Topics for Discussion at the Forthcoming Meeting"; Licklider, "Computers and Government."

21 Gregory, *Star Trek: Parallel Narratives*.

22 Lanier and Weyl, "How Civic Technology Can Help Stop a Pandemic."

23 Allen, Wallach, and Smit, "Why Machine Ethics?"; Brundage, "Limitations and Risks of Machine Ethics"; Anderson and Anderson, *Machine Ethics*; Kaspersen and Wallach, "Why Are We Failing at the Ethics of AI?"; Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines"; Wallach and Kaspersen, "Creative Reflections on the History & Role of AI Ethics"; Simons, *Algorithms for the People: Democracy in the Age of AI*.

fields of applied ethics, both of which aim to structure the trajectory of technological development towards human flourishing. Each field recognizes the potential positive outcomes to be unlocked by technological progress—from vaccine development to carbon capture to efficient resource allocation—while remaining clear-eyed about the dystopian possibilities. Crucially, these fields have highlighted the centrality of social determination to technological trajectories. The expansion of technological capabilities does not determine social outcomes. Co-constituted sociotechnical systems, however, have significant effects on long-term social and political economic structure. Responsible technology seeks transformative improvement toward human flourishing constrained by guardrails and guiding values supportive of democracy, freedom, and pluralism.

3.1 Responsible Technology

3.1.1 *AI ethics*

As far back as 2011, scholars began exploring what they called "machine ethics," ethical questions related to intelligent systems indistinguishable from humans in communication and superior in performing well-defined tasks, such as playing chess. A few years later, computer scientists and lawyers began wrestling with the patterns of inequality and discrimination that data-driven tools could reproduce. AI ethics morphed into a literature on fairness and discrimination in AI.[24] Among the most well-publicized findings of this literature is that when patterns of social injustice are reflected in datasets, predictive tools trained on those datasets tend to replicate and compound those patterns. This forces computer scientists and engineers to choose which values and interests to prioritize in the design of these tools.[25]

More recent work has been somewhat critical of how the field of AI ethics has developed. Some authors have focused on the field's centralization in a narrow set of journals and institutions that are funded by large technology companies. Others have centered criticism on the focus and framing of AI ethics, arguing that "artificial intelligence" is a misleading focal point for ethical reasoning and that instead, our evaluation should focus on how data should be used to generate predictions and how those predictions should be used by specific institutions in particular contexts to make decisions. What matters is the institutional context in which predictive tools are designed and used to make decisions, rather than the statistical features of those tools considered abstracted from social, economic, and political context.[26] Still others have raised concerns with the short-term focus of the field of AI ethics, pointing to long-term consequences, catastrophic risks, and path dependence in the AI developed today.

These criticisms prompted a shift in focus in the AI ethics field towards both governance and democracy. If there are inevitably value-laden choices to be made in the design and deployment of predictive tools with far-reaching and potentially catastrophic consequences, then what's required is not just arguments about which choices should be made in individual cases, but a framework for establishing structures of governance over ongoing processes of experimentation with building and using data-driven systems. This is a vital lesson for the evolving ecosystem of DST. The focus of ethics for decentralized social tech should be on a framework that describes how we should evaluate and oversee ongoing experiments supported by blockchain and other pertinent technologies in multiple domains, not just on evaluating individual use cases for blockchain technologies. In other words, the focus should be on governance and evaluation in a process of experimentation, not simply on the rights and wrongs of abstracted individual cases.

This recent focus of AI ethics on governance is further underscored by taking the standpoint of democracy. Among the most fundamental questions in AI ethics are not just *what* choices should be made in the design and deployment of data-driven systems but *who* should make those decisions, on what timescale, and against what criteria. Here, there has been cutting-edge work

24 Barocas and Selbst, "Big Data's Disparate Impact"; Knight, "Biased Algorithms Are Everywhere, and No One Seems to Care."

25 Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments"; Chouldechova and Roth, "The Frontiers of Fairness in Machine Learning"; Birhane et al., "The Values Encoded in Machine Learning Research."

26 Simons, *Algorithms for the People*, chaps. 2–3; Vaughan and Wallach, "A Human-Centered Agenda for Intelligible Machine Learning."

to explore what, in practice, it means to structure democratic control over the socio-technical processes involved in the widespread integration of data-driven systems into many of our fundamental social, economic, and political institutions. Whereas it took almost a decade for the field of AI ethics to recognize the importance of democratic determination, we argue that ethics frameworks for DSTs should build in that principle and address how to structure that determination from the start. Our paper makes first steps in that direction.

3.1.2 *Bioethics*

The field of bioethics provides further guidance for ethics for DSTs, given the centrality of experimental ethics and adjudication of value trade-offs. In particular, recent work on synthetic biology has grappled with the problem of how to structure openness to experimentation. The 2010 U.S.-based Presidential Commission on Bioethics Report, *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*, is a useful guide. This report is focused on both the importance of experimentation and the need for steering in the direction of democratic values. The report's authors write:

> Democracies depend on *intellectual freedom* coupled with the responsibility of individuals and institutions to use their creative potential in morally accountable ways. Sustained and dedicated creative intellectual exploration begets much of our scientific and technological progress. While many emerging technologies raise "dual use" concerns—when new technologies intended for good may be used to cause harm—these risks alone are generally insufficient to justify limits on intellectual freedom. As a corollary to the principle of intellectual freedom and responsibility, the Commission endorses a principle of *regulatory parsimony*, recommending only as much oversight as is truly necessary to ensure justice, fairness, security, and safety while pursuing the public good. This is particularly important in emerging technologies, which by their very definition are still

> in formation and are not well suited for sharply specified limitations. …
>
> Because synthetic biology poses some unusual potential risks, as "amateur" or "do-it-yourself" (DIY) scientists and others outside of traditional research environments explore the field, these risks must be identified and anticipated, as they are for other emerging technologies, with systems and policies to assess and respond to them [including reliable containment and control mechanisms] while supporting work toward potential benefits. …
>
> Responsible conduct of synthetic biology research, like all areas of biological research, rests heavily on the behavior of individual scientists. Creating a culture of responsibility in the synthetic biology community could do more to promote responsible stewardship in synthetic biology than any other single strategy.[27]

The culture of responsible stewardship that they recommend involves finding a middle ground between the traditional "precautionary principle" and a "proactionary principle." On the precautionary principle, experimentation is "halt[ed] or substantially slow[ed] until risks can be identified and mitigated." A typical formulation is that "if an action or policy has the potential to cause harm but uncertainty exists regarding the likelihood or severity of harm, the responsibility for demonstrating the safety of the approach belongs to those advocating for the policy or action."[28] In contrast, the "proactionary principle" assumes that technology "should be considered 'safe, economically desirable and intrinsically good unless and until shown to be otherwise, which means that the burden of proof is on those who want to slow down a given line of research.'"[29]

The report's authors define that middle ground as an effort to achieve "prudent vigilance," where processes are put in place to assess both benefits and risks not only before but also after projects are undertaken, and in particular to track benefits and risks as technologies "diffuse into public and private sectors" (Presidential Commission). There are roles for both technologists and government actors in carrying out "prudent vigilance." Technologists can look for opportunities to design their technologies in risk-mitigating and opportunity-opening ways. Here, we underscore the importance for technologists of intentionally structuring experimentation. This requires: (1) being explicit about goals and offering justifications; (2) having governance structures to oversee the design and evaluation of experiments; (3) publicly reporting what happens, what worked, what didn't, or where unintended consequences are beginning to emerge; and (4) having clear mechanisms for democratic oversight, developed in collaboration with public democratic authorities.

Governments, to carry out their part in achieving prudent vigilance, need to be fully informed of ongoing developments, and both their risks and opportunities, and need to ensure that relevant existing oversight authorities are folding new sociotechnical practices into their domain as they emerge. The failure to identify algorithmic-selection of content on social media platforms as a form of editorial curation and publishing, for instance, would be an example of a governmental failure at prudent vigilance.

3.2 An Ethics of Experimentalism

Above, we briefly sketched the components of an ethical framework for structuring experimentation. We spelled out four elements required to achieve prudent vigilance. We now will say more about these elements. Importantly, as we proceed to flesh out this ethics of experimentation, we should note that this approach constitutes an alternative to the most common ways that both utilitarian cost-benefit reasoning and deontological, rule-based reasoning are applied. Both of those philosophical frameworks are usually applied with a presumption of a static landscape where distant consequences can be reasonably predicted in advance of

27 Presidential Commission for the Study of Bioethical Issues, *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*, 5, 8, 11.

28 Presidential Commission for the Study of Bioethical Issues, *New Directions*, 123.

29 Presidential Commission for the Study of Bioethical Issues, New Directions, 127; Parens, Johnston, and Moses, *Ethical Issues in Synthetic Biology: An Overview of the Debates.*

"In conditions of clear, present, and radical uncertainty, an ethics of experimentation is the only kind of framework that can gain traction on the realities of the moment."

action. Utilitarianism falls into this trap because (as executed by economists and the like) it relies on precise quantitative and usually statistically grounded calculations. Deontology does so because (as applied mostly by analytic philosophers) it presumes the clear discernibility and universal applicability of detailed guides to action. Of course, we have some sympathy for objections both that such a static decision-making landscape never pertains and that there may be ways to pursue utilitarianism and deontology without such static assumptions. Nonetheless, it should be evident that neither of these well-developed standard approaches pertains in the context of constitutional moments, where decisions are in fact bringing new horizons of possibility into existence. In conditions of clear, present, and radical uncertainty, an ethics of experimentation is the only kind of framework that can gain traction on the realities of the moment.

In a 2022 essay, economist Dani Rodrik and political scientist Charles Sabel sketched out in some detail the design principles needed "for building dynamic governance arrangements… under conditions of uncertainty and learning, through ongoing review and revision of objectives, instruments, and benchmarks."[30] They make the point that under conditions of great uncertainty, neither contracting nor regulation can proceed with full specification at the outset about how risks and rewards will be allocated, or about the obligations, precisely because none of these are fully visible from the outset, nor have norms stabilized sufficiently to fill in the "gaps and ambiguities in formal agreements." The beginning of a project, therefore, should not be a fully specified scope or contract but rather the establishment of "broad goals and a regime for evaluating achievement of them." They write:

> As observed in domains as diverse as biotechnology, information technology (IT), and advanced manufacturing, this regime establishes regular, joint reviews of progress toward interim targets or milestones, procedures for deciding whether and with what exact aim to proceed or not, and mechanisms for resolving disagreements. … regular review and deliberate consideration of the interim results thus create the conditions in which informal norms and self-interested calculations bind the parties to continue promising collaboration in good faith.[31]

They find a similar way of structuring regulation viable for conditions of uncertainty:

> Under uncertainty, neither the regulator nor the regulated parties have reliable information on the possibilities and costs of adjustment in the medium term, and only conjectures regarding the possibilities that will open—or not—upon further investigation. Again, the response—seen in food safety, civil aviation, and pharmaceuticals, among many other industries—is the creation of an information- exchange regime that ties ongoing specification of goals… to continuing exploration of new solutions.[32]

In the context of DSTs, then, both technologists and regulators should be looking to develop structured approaches to experimentation that secure intentional goal statements from technologists as well as processes for ongoing review and adjustment. These might be private-public collaborative review processes, or processes created by technologists, building a culture of responsible stewardship. Either way, the goal statements should be "ambitious

30 Rodrik and Sabel, "Building a Good Jobs Economy," p. 63.

31 Rodrik and Sabel, "Building a Good Jobs Economy," 70.

32 Rodrik and Sabel, "Building a Good Jobs Economy," 70–71.

and open-ended" and, we would argue, should be developed in reference to human flourishing and the multiple values that constitute it, as we will explore in more depth below. The method requires working in the first instance to seek solutions that maximize alignment among these several values, with debates over trade-offs following only after alignment has been maximized.

Finally, the harms currently afflicting development of DSTs might all be addressed through this method. The challenges around harm to environmental sustainability, illegality, privacy violations, speculation, financialization, and exploitation are all examples of violations to the values that are components of human flourishing. A technological project that lays out ambitious goals supportive of core components of human flourishing would, of necessity, also need to propose methods for avoiding the harms that undermine human flourishing. This combination of flourishing-advancing design features and harm-avoiding design features would be what responsible technologists and/or regulators should be monitoring over time. Steering with these parameters will significantly increase the likelihood that decentralized social tech innovations support human well-being. We turn now to an account of human flourishing that can provide guardrails for experimentation.

4. Prudent Vigilance for Web3 Ethics

Above we defined responsible technology as seeking transformative improvement toward human flourishing constrained by guardrails and guiding values supportive of democracy, freedom, and pluralism. The positive potential of the new technologies to deliver public beneficence brings an accompanying duty to pursue experimentation. But the potential for harm requires that experimentation be undertaken with prudent vigilance. The first step in having a framework for Ethics for DSTs is to offer a definition for human flourishing and justifying the link between that flourishing and democracy, freedom, and pluralism. The second step will be to propose some guardrails that support prudent vigilance through a period of experimentation when uncertainty about outcomes and consequences is a necessary part of moving forward.

While we might apply the prudent vigilance framework to any number of experiments in the Web3 ecosystem, including the Fediverse, here we will focus specifically on the application to blockchain-based developments in order to offer one, but only one, very concrete case study. As we shall see, what is required of prudent vigilance will be somewhat different for the different layers of the Web3 stack: Layer 0 = Network Protocols; Layer 1 = Blockchain; Layer 2 = Scaling Solutions; and Layer 3 = Applications.

4.1 Human Flourishing

Decisions about the development of emerging technologies can optimize for a variety of different possible value sets. The orientation to any particular set of values as "good" outcomes for human life constitutes an implicit theory of justice, and technologists make choices about which normative framework will structure their own decision-making. A theory of justice does not seek to describe the rules that have come to be in human society—whether as a result of the emergence of self-organizing systems of human cooperation or as a result of intentional efforts to organize human governance. Instead, a theory of justice seeks to identify the parameters establishing which among possible sets of rules for human interaction yield the best prospects for human flourishing, at both an individual and a collective level. These parameters would then be relevant to political economy in setting directions for and bounds to our experimentalism, as we seek to identify which economic policies count as redesigns that improve, rather than worsen, human prospects.

While philosophy throughout the ages has offered many potential paradigms, we argue for plurality or power-sharing liberalism.[33] Power-sharing liberalism rests on the idea that two kinds of freedom are necessary for human flourishing—negative liberties where we are protected in our person, our property, our conscience, our expression, and our associations; and positive liberties where we are able to govern ourselves in our private lives and share in the governance

33 Allen et al., Political Economy of Justice; Allen, *Justice by Means of Democracy*, 2023.

of our public lives.

Importantly, the defining purposes of specific human lives that can count as examples of flourishing are various; there is no single picture of the flourishing life. What is shared, however, across cases of human flourishing is that human beings are creatures who need to chart their own courses in life. Humans thrive on autonomy, the opportunity for self-creation and self-governance.

That autonomy is made real in our political institutions via the protection of both negative and positive liberties. Negative liberties are those rights of free speech, association, freedom of religion, and so forth, that permit us to chart our own course toward happiness, based on our own definitions of the good. Positive liberties and rights are those opportunities that we have to participate in our political institutions as decision-makers, as voters, as elected officials, as people who contribute to the deliberations of our public bodies. Through our positive liberties, or political equality, we have the chance to shape our collective world together. The autonomy that delivers human flourishing requires shared autonomy through political institutions in order to reach its fullest form. This makes democracy necessary to the achievement of human flourishing and justice.

A plurality or power-sharing liberalism approach to the pursuit of justice seeks proactively to advance a set of values and to avoid the harms that are their opposites. Those values include but are not limited to non-domination; individual and community self-determination; egalitarian pluralism; connective and coordinating capacity (bridging); and collective ownership of the assets needed for shared governance.

Each of those values tracks a "techno-normative concept" in the DST space. A "techno-normative concept" is a proposed functionality or feature that is made possible (but not necessarily fully realized) by technological innovation. These are normative because technologists and developers attribute desirability to them. In order to evaluate both (1) whether these really are desirable new features, and (2) how these features relate to broader social and political goals, we link these "techno-normative concepts" arising from new blockchain innovation, in particular, to normative concepts from power-sharing liberalism. The relationship between the values and the concepts can be understood as follows:

Non-domination	Trustlessness
Individual and community self-determination	User control, Sovereignty
Egalitarian pluralism	Decentralization
Coordination capacity	Coordination, Disintermediation
Collective ownership	Ownership, Regenerative economies

As we sketch out the relation between the philosophical and techno-normative concepts below, we will see limits in the fit. The techno-normative concepts only imperfectly count as operationalizations of the philosophical concepts. In some cases, it is not even entirely clear that the techno-normative concept does in fact capture a normatively desirable phenomenon (e.g., trustlessness). But at the same time, what will also emerge—but which we cannot pursue fully in this paper—is that the techno-normative concepts themselves bring to the surface places where the philosophical concepts may be unstable.

4.1.1 Non-Domination and Trustlessness

To be free from domination, in the argument of Philip Pettit, is to be free from the prospect of arbitrary interference or "reserve control."[34] In his brilliant book *Just Freedom*, Pettit explains freedom from domination with reference to the expression "free rein." If you give a horse free rein, it may be able to go where it wants, but the rider retains "reserve control" and can reassert constraint at any point. We must answer three questions to better understand the principle of non-domination. First, what are the possible

34 Pettit, *Just Freedom: A Moral Compass for a Complex World*.

sources of domination? We need to think of structures—including those not intentionally constructed or maintained by any individual or set of actors—as possible sources of domination. These might be exclusionary patterns or access to/ownership of social infrastructure, hierarchical organizations such as firms within the marketplace, relations of power in markets, or forms of private property relations, to name only a few.[35] Second, what is the relation between different forms of domination? For instance, how should we think about the relationship between monopolistic practices in markets and forms of political domination? Does one inexorably produce the other such that a key principle for pursuing a just economy is an anti-monopoly principle. Simple anti-monopoly cannot be an absolute value once one recognizes that markets are pervaded by power even in the absence of monopoly. Regulated monopolies or public utilities, for example, may be less susceptible to domination or exploitation than "competitive" unregulated markets; indeed, because they are monopolies they can enjoy economies of scale while public political power counteracts their market power.[36] Also, there is a sense in which the ownership of property is intrinsically monopolistic, leading, for instance, to the inefficient allocation and use of property. Third, how high is the bar we are setting up when we seek to build our political economy on non-domination? Difference is an irreducible and even desirable feature of any society.[37] The hard question is when and how difference becomes domination. In particular—a question for all of us to reflect on, related to the prescriptive aspect of our work—is, how do we design institutions that detect when difference becomes domination?[38] And what kinds of institutions can correct for domination?

Technologists seem to reference this concept of non-domination elliptically when they argue that the techno-normative concept known as "trustlessness" ought to be considered a desirable potential feature of blockchain-based technologies. In its absolute form, trustlessness is used to describe a system in which it is not necessary for an individual to make assumptions that anyone else will act either in good faith or as they expect they will to fulfill reasonable desired ends. In the context of blockchain technologies, trustlessness describes a blockchain that stores data and records its history accurately without relying on data maintenance or protocol approval from an individual or small group. Automated smart contracts, governance protocols for DAOs, and other blockchain applications might be used to build networks requiring less trust since they can, in theory, preclude intermediary approval or interference on the part of one or a few actors. In practice, no existing blockchain technology offers perfect trustlessness. Hardline bitcoin fundamentalists and techno-utopian DeFi advocates seem to consider trustlessness as a necessary and noble goal/condition of fully decentralized healthy interaction on a blockchain. They see the immutability of a blockchain's ledger as a valuable and radical difference from prior internet structures, where individuals and groups had to place "trust" in institutions and managers to protect data and network history or to preside over procedures. Ethereum's transition from proof of work to proof of stake was praised as a move towards greater trustlessness in the blockchain, as validators/miners were not rewarded in direct proportion to their processing capacity. In practice, however, no blockchain yet appears perfectly trustless, and it is indeed unclear whether trustlessness is as connected to non-domination as blockchain evangelists argue it is.

Some efforts to motivate the desirability of trustlessness compare the distinction between perfect and imperfect trustlessness to the Rawlsian distinction between perfect and imperfect procedures.[39] This is a bit of a mismatch. Whereas perfect procedures ensure fair and just outcomes regardless of others' foul play, perfect trustlessness ensures an individual's capacity to act on their will insofar as they can on-chain regardless of others' behavior. Perfect procedures might rely on an impartial technocratic adjudication process guided by principles of fairness, whereas a perfectly trustless blockchain

35 Allen and Somanathan, ed., *Difference without Domination: Pursuing Justice in Diverse Democracies.*

36 Rahman, *Democracy Against Domination*.

37 Allen, *Justice by Means of Democracy*.

38 Allen, "A New Theory of Justice: Difference without Domination."

39 Moessinger. "An Overview and a Response to Vitalik Buterin."

might satisfy principles like autonomy without necessarily conferring fair outcomes. For this reason, we might be interested in exploring ways to build protocols that are both trustless and procedurally fair or that balance autonomy with fairness. Thus, while trustlessness is related to the principle of non-domination, since a perfectly trustless society would preclude the possibility of unilateral action to co-opt another's means for undesired ends on the chain, it does not satisfy all the necessary conditions of non-domination. Indeed, even a perfectly trustless blockchain environment cannot guarantee that the social and economic system in which it is embedded precludes the possibility of domination or exploitation by, for example, fraud schemes. As decentralized social technologists experiment with different models for trustless or trust-required interaction in different domains of online, networked life, we ought to think of these experiments in terms of the broader principle of non-domination as a basic social and political principle.

All that said, here are some problems that emerge. It's unclear that we do want to be trying to design away the need for trust in institutions

as sites of authority and legitimacy in a world where, in practice, we rely on trusted institutions to make countless judgments on our behalf. The techno-normative concept seems to go too far. Yet when we try to correct it, for instance, by arguing that one should not have to trust an entity that is also in a position of domination with a power hierarchy, what comes into view is that the concept of non-domination may be better understood as a concept of symmetric power relations. The philosophical concept may be insufficiently precise. And then another challenge, the effort to operationalize symmetric power relations in all aspects of social organization, is devilishly difficult. The conceptual and practical challenges that emerge from trying to link philosophical and techno-normative concepts promises to open up many promising lines of inquiry.

4.1.2 Individual/Community Self-Determination and User Control/Self-Sovereignty

Individual and community self-determination flow from freedom. Three concepts of freedom are relevant: the straightforward political freedom to vote, a necessary condition for any electoral democracy; the freedom from the arbitrary power of the majority, most notably the protection of minorities, a necessary condition of a liberal democracy; and then positive freedom, which is the freedom to participate on equal terms, or "real" liberty. Perhaps unsurprisingly, most discussion focused on what positive freedom was and what the conditions for its effective exercise were.

While positive freedom is another way of describing political equality, the connection of the concept of freedom to markets and property brings an additional question to the fore. What are the conditions for the exercise of effective freedom and how does this bear on political economy? For instance, what is the relationship between the ownership of property and effective freedom? Do those who own property have an unavoidable power over the freedom of those who do not? Another way of putting this might be: Is domination intrinsic to private property? What does a property regime that avoids structures of domination look like? These are questions that careful and intentional experimentation with DSTs can help us to answer.

It seems that technologists and others working to develop and use blockchain technologies invoke a normative principle similar to self-determination when they attribute desirability to a techno-normative term often called "user

control." Whereas centralized technology users have to access the internet through service providers, browsers, search engines, and other platforms, decentralized technologies offer users the capacity to exchange tokens and interact over the network in local wallets not stored or managed through third-party platforms. Because of this, technologists promise to grant users control over their data and experience on the internet. Platform companies currently collect and own non-portable user data, keeping users somewhat trapped in enclosed ecosystems (Google Suite vs. Outlook, Spotify vs. Apple Music). Blockchains, it is said, might allow users to own and tokenize their data, and collectively "own" a disaggregated internet network.[40]

User control through locally stored wallets raises questions related to wallet recovery, how identity should back or be related to wallets, and issues of financializing personal data. There is a risk, it seems, of personal data markets that do not liberate individuals through greater choice and control, but place some in a position where financial insecurity necessitates a sacrifice of privacy while others with more fortune do not have to do the same. This inequality is potentially dangerous both for the users and for any data-based prediction technologies to which one might sell personal data. While blockchain evangelists promise user ownership of the internet as a feature of utopia, the question remains of whether ownership of internet networks should be considered possible at all.

Another techno-normative term that appears frequently in DST discussions is "sovereignty," or "self-sovereign identity" (SSI). Sovereignty or self-sovereign identity is related to user control. One article defines it as "individual control over identity relevant private data," or "the capacity to choose where such private data is stored, and the ability to provide it to those who need to validate it."[41] There are many technologies that aim at SSI functionality and none that fully achieve it, but the general idea is that individuals can have key-accessible private information and can offer other specified users the ability to access keys to some of that information. The result is a "digital" identity of many component data points which can be kept private or shared with whomever one chooses.

Since they elaborate the possibilities for individual participation in online communities and the extent to which individuals can be fully autonomous actors in those communities, these two terms—user control and sovereignty—relate to broader questions of privacy, identity, freedom, and community. As DSTs experiment with different forms of user control and self-sovereign online identity, ethicists might do well to evaluate new innovation towards the techno-normative ideals of user control and sovereignty in terms of freedom, community, and privacy.

4.1.3 Egalitarian Pluralism and Decentralization

With the theme of political equality, we return to and extend on the concept of non-domination. Indeed, political equality is a means for ensuring non-domination, and non-domination is also a necessary condition for political equality. Two ideas of political equality are pertinent. The first is the familiar and straightforward sense of equal participation in the political processes (e.g., voting rights, party political funding).[42] The second is somewhat subtler: equal standing in the political process or sphere. This means more than simply the effective capacity to vote. It means the ability to equally understand and leverage the political process, to have the possibility of shaping the process itself through collective action. With these two different pictures of political equality in view—one relating to formal civil and political rights, the other to a capabilities conception (cf. the work of Amartya Sen and Martha Nussbaum[43])—we argue for cautious experimentation with the aim of democratic mechanisms of decision-making, including for economic decisions, as a way to ensure that a political economy reinforces rather than undermines political equality. Here too the concept of power symmetries and asymmetries lurks just below the surface and points to new lines of both philosophical and technical inquiry.

40 Ali et al., "A Comparative Study: Blockchain Technology Utilization Benefits, Challenges and Functionalities."

41 Ishmaev, "Sovereignty, privacy, and ethics in blockchain-based identity management systems."

42 Estlund, "Debate: Liberalism, Equality, and Fraternity in Cohen's Critique of Rawls."

43 Nussbaum, *Creating Capabilities: The Human Development Approach; Alexander, Capabilities and Social Justice: The Political Philosophy of Amartya Sen and Martha Nussbaum.*

Another normative ideal related to the aim of a just political economy is egalitarian pluralism. The ideal of egalitarian pluralism aims to ensure that people have control over their own lives, relationships, and communities. It recognizes that people value community in different ways and that they have distinct hopes and ambitions for life. Part of the value of egalitarian pluralism is that it makes us think harder about how to pursue a just political economy.[44] That is, it draws our attention to those who are subject to patterns of domination or injustice, to listen to their conception of community and justice rather than impose our own. It does not assume that every well-intentioned proposal for action in pursuit of justice is necessarily justified by its purpose.

This implies, for instance, that we have to reevaluate other values when weighing up proposals for action—such as diversity or efficiency. We have to think about the relation between our values as we seek to enact them in the world: How should we pursue an "infrastructure of inclusion"? How should we build institutions to detect and correct for relations of domination? Egalitarian pluralism reminds us that these are not simply straightforward second-order questions whose answers follow from identifying our first-order values.

The two ideals of political equality and egalitarian pluralism are raised when blockchain enthusiasts discuss the techno-normative value they call "decentralization." Perfect decentralization would in many ways result from a combination (or perhaps a result) of perfectly trustless and permissionless blockchains. Of course, perfect decentralization is not currently achieved by any blockchain technologies—it is merely an idea, a technical potential unlocked by normative aspirations to trustlessness and permissionlessness. Since blockchains store code that runs automatically and allows for user interaction and exchange without gatekeepers holding centralized power, it is said that power is "decentralized" across all the nodes and users supporting a blockchain network. At the highest level, the computational work holding up a chain is spread out through many nodes. At the lowest level, individual transactions do not pass through any central conduit—rather, they are "peer to peer." The extent to which a blockchain is decentralized relies on two things: first, the validation/mining technology and second, the power of intermediary institutions in granting users access to on-chain activities. For example, Ethereum may be more decentralized than Bitcoin because of its use of proof of stake over proof of work, but it is hardly fully decentralized, since many intermediary companies fashioned in the image of financial service firms hold significant power.

Blockchain enthusiasts argue that decentralized networks are less likely to marginalize one community and privilege another, but instead treat all voices and users as co-participants in a market or platform. This projected effect of a decentralized network is a promise of greater political equality and egalitarian pluralism. Decentralization, however, is by no means sufficient for political equality and egalitarian pluralism. Indeed, decentralization may also produce division and infighting, or, as is the case with many of the ideals for on-chain interaction, a decentralized network may not necessarily produce or be embedded within a decentralized society.

4.1.4 Coordination Capacity and Coordination/Disintermediation

Scholars of social capital distinguish among three kinds of social ties: bonding, bridging, and linking. Bonding ties are those (generally strong) connections that bind kin, close friends, and social similars to one another; bridging ties are those (generally weaker) ties that connect people across demographic cleavages (age, race, class, occupation, religion, and the like); finally, linking ties are the vertical connections between people at different levels of a status hierarchy, as in, for instance, the employment context.[45] An associational ecosystem that maximizes bridging ties should minimize the likelihood that social difference articulates with domination.

A "connected society" is one where people can enjoy the bonds of solidarity and community but are equally engaged in the "bridging" work of bringing diverse communities into positive

44 Lippert-Rasmussen, "Pluralist Egalitarianism."

45 Granovetter, "The Strength of Weak Ties"; Szreter and Woolcock, "Health by Association? Social Capital, Social Theory, and the Political Economy of Public Health."

relations. It is one where people also individually desire and succeed at forming personally valuable relationships across boundaries of difference. The goal of acting in this policy space would be the cultivation of an associational ecosystem in which people do have the opportunity to choose their associates in order to realize their personal visions of the good life but also find themselves routinely interacting with those whom they have not, so to speak, chosen, and routinely obliged to share power in a variety of public contexts with these unchosen others.[46]

> "...improving coordination capacity and collective intelligence can contribute to our ability to tackle the major challenges that define not only this century, but the human condition as a whole."

A focus on a normative ideal of connection is one way of operationalizing the goal of inclusion that defines twenty-first–century democratic practice. A challenge that emerges, however, is to achieve efficacy simultaneously with achieving inclusion. The normative ideal of a "connected society" thus requires also coming to grips with capacity for coordination. Connection turned into effective action is coordination.

Coordination requires building common interests among diverse groups, and working together to achieve these interests. In distributed strategies of coordination, this begins from connection. Coordination failures occur when either of these steps does not happen. A range of large-scale problems currently faced by humanity can be chalked up to coordination failures: pandemics, climate change, and the like. These are not coordination failures alone: political conflict, value-based disagreements, and material domination contribute to collective shortcomings. Nonetheless, improving coordination capacity and collective intelligence can contribute to our ability to tackle the major challenges that define not only this century, but the human condition as a whole.

However, coordination among some groups while others are excluded is dangerous and harmful. Imagine a set of large conglomerates coordinating to set prices, hurting consumers (and likely workers), or lobbyists coordinating to sway democratically elected politicians towards outcomes worse for the public interest. The coordination capacity of the powerful against the powerless is common, and our systems are often set up to enable it.

Thus, historically, coordination has *both* been insufficient and unequally dispersed. It has also been often too top-down and hierarchical, rather than distributed, pluralistic, and networked. Decentralized technologies should work on both of these issues. This can happen through:

- Bridging communities, building a connected society **capable of choosing actions in the common interest (as above)**

Building coordination mechanisms that allow us to **act together on those actions, including:**

- Common knowledge and reduced information asymmetries, resulting in greater visibility of shared information both within and between communities;
- Resilient systems built through federation and subsidiarity; and
- Coordination without rent-seeking intermediaries

4.1.5 *Collective Ownership and Ownership/ Regenerative Economies*

Without realizing it, modern, liberal, capitalist states have actually facilitated an important form of collective ownership, namely "co-ownership of political institutions." This involves recognizing that all the machinery and value of a democracy—all of the assemblies, congresses, and judicial offices at federal, state, county, and

46 Allen, "Toward a Connected Society."

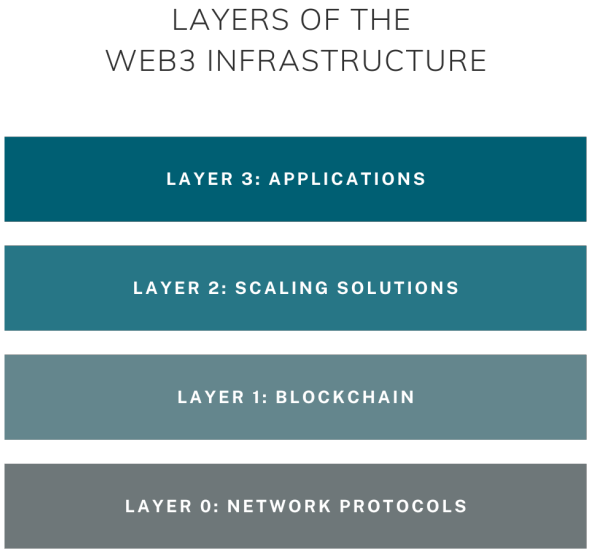
municipal levels—constitutes a valuable asset. This massive apparatus is a form of property that members of a polity own together. Economists have recognized the value that has accrued to societies that have accepted co-ownership of the society's steering mechanisms.[47]

Similar value from co-ownership could emerge from DSTs and their tools. There are several examples of collective ownership of non-depleting goods on Web3. In theory, DAOs enable shared ownership of digital assets.[48] DAOs combine multi-signature wallets for collective asset ownership, internal, token-based governance systems such as token voting, and smart contracts for asset distribution to construct low-overhead, collective organizing structures. The aim of a DAO is to replace human-driven, hierarchical decision-making predicated on existing legal systems with a decentralized protocol unbound from regulation that specifies the bounds of human input and interaction with the organization, and that distributes internal capital via this protocol, enforced and executed by a blockchain. As work progresses on bridging digital-physical gaps with blockchain technology, they may eventually support the shared ownership of physical assets. Further, DAOs present certain operational efficiency as compared to existing legal entities: they are able to rapidly pool and deploy capital, implement low-cost decision-making schemes, and protect member assets for internal fraud detection.[49]

The values from the philosophical context capture core elements of an account of well-being. As we said at the outset, we do not think they capture all elements of well-being; they do, however, capture some subset of necessary elements. Well-being, anchored in positive freedom, will depend on political equality, non-domination, and anti-monopoly efforts while sustainably delivering broad-based economic security and an experience of community in societies that simultaneously embrace pluralism and egalitarianism. The tech concepts to which those philosophical concepts have an affinity constitute a set of core ideas that might help establish the guardrails for a socially valuable version of Web3 development. Identifying the design solutions that embody these values will depend on practices of experimentalism.

4.2 Specific Case: Prudent Vigilance Applied to Blockchain Layers

To understand the requirements of prudent vigilance for decentralized architectures, it will be helpful to look at a specific case. We turn now to an illustration from the blockchain context. To reckon with architectures that are primarily focused on blockchains, we need to attend to the different layers of Web3 infrastructure, the building blocks that provide tools for launching Web3 applications, and the stakes of each. As mentioned above, those layers are: Layer 0 = Network Protocols; Layer 1 = Blockchain; Layer 2 = Scaling Solutions; and Layer 3 = Applications.



We will, for the purposes of this paper, designate as Layer 0 the network frameworks that run beneath a blockchain, whose components

47 Nobel Prize–winning economist Herbert Simon made a similar argument that a democracy's political institutions constitute an asset owned by the people as a whole. Moreover, he argued that this commonly owned asset was the source of significant wealth generation in developed democracies, and that that wealth might consequently be allocated to a universal basic income on the grounds that the public owns the asset that generated it. See Simon, "Response to Philippe van Parijs on Universal Basic Income."

48 Sims, "Blockchain and Decentralised Autonomous Organisations (DAOs): The Evolution of Companies?"

49 Wright, "The Rise of Decentralized Autonomous Organizations: Opportunities and Challenges."

include protocols, connections, hardware, miners, and other connections that enable Layer 1 networks to run smoothly. We also include the underlying internet backbone and the world's power systems, and any other infrastructure needed to run the blockchain in the first place.[50]

Layer 1 is the underlying blockchain network on which transactions can be validated and finalized. Layer 1 is the base or main implementation layer that sets the network's rules and parameters, including the consensus algorithm, programming languages, block time, dispute resolutions, and transaction throughput. Examples of Layer 1 blockchain architecture include Bitcoin, Etherum, Binance Smart Chain, Cardano, Solana, Avalanche, Algorand, Cosmos, etc.

The blockchain trilemma is the belief that at any given time decentralized networks can only provide two of three elements with respect to decentralization (whether a chain can run without any trust dependencies on a small group of centralized actors), security (whether a chain can resist a large attack of participating nodes), and scalability (whether a chain can process more transactions than a singular regular node can verify).[51]

Layer 2 refers to various protocols and solutions built on top of the Layer 1 blockchain to improve speed and functionality by often processing transactions off-chain.[52] Layer 2 protocols provide higher transaction speed and throughput, decongesting the Layer 1 blockchain and reducing gas fees. Examples of Layer 2 protocols for Bitcoin include Lightning Network. Examples of Layer 2 protocols for Ethereum include Polygon, Optimism, StarkNet, and Arbitrum.

Layer 3 consists of blockchain-based decentralized protocols and applications (DApps), including games, distributed storage, and decentralized finance (DeFi).[53] Many DApps have cross-chain integrations, which allow users to access various blockchain platforms within a singular app. Examples of DApps on Ethereum include Uniswap, OpenSea, MetaMask, and Aave.

The development of Layers 0, 1, and 2 should be guided by experimentation protocols that explicitly invoke and seek to realize the principles we have laid out above—user control, sovereignty, decentralization, bridging, and regeneration—or revised versions of them (and of further concepts like trustlessness) pending further normative work. These should be additional design principles beyond scalability, decentralization, and security. Whenever the latter principles are in tension with the former, developers should pause to seek out methods for realigning their action with all of the principles.[54]

Given their global scale, Layers 0, 1, and 2 should also be subject to a governance body established to bring prudent vigilance to the interactions between emergent socio-technical infrastructure and the socio-technical infrastructure of existing societies. The goal of this governance body—whether a new body

50 Dang, "Understanding the Blockchain Layered Architecture to Solve the Scalability Challenges."

51 Buterin, "Why Sharding Is Great: Demystifying the Technical Properties."

52 Marcobello, "What Are Layer 2s and Why Are They Important?"

53 George, "What Is a Dapp?"

54 Lewis, "Where Bioethics Meets Machine Ethics."

or a branch of an existing organization, and whether public or civil—should not be to block the emergence of transformative infrastructure but rather to provide a context for adjudication as unequal impacts are felt by different stakeholders. National governments would be prudent to seek out the potentialities of these new tools to address long-standing challenges and to communicate design desiderata aligned with the human flourishing principles and an analysis of the public good as well as to make investments supportive of those goals. National governments should also prepare to navigate potential jurisdictional fluidity, both subnationally and transnationally. Again, that navigation should be guided by the human flourishing principles articulated here.

The apps populating Layer 3 may or may not scale to constitutively consequential levels. Protocols for assessing the ethics of experimentation involved in a Layer 3 app should be tied to investment decisions. When investment bets are placed on the expectation of national or global scaling for an app, that app should integrate human flourishing design principles to the models being scaled up.

What does prudent vigilance of this sort look like in practice? What does it mean to seek experimentation with techno-normative principles with a view to achieving stronger alignment to the philosophical principles? Our final section concludes by reviewing some more specific cases, analyzed in terms of the framework we have laid out here.

4.3 Prudent Vigilance Applied to DAOs

4.3.1 The Opportunity

Enabling self-determination and non-domination: Sandboxes for governance experiments are few and far between, and DAOs may present one such opportunity. It is clear that the decision-making processes that pass for democracy in most nation-state contexts rarely execute on the will of the people; DAO governance presents one opportunity to change this. Already, experiments are underway testing more granular and technologically sophisticated methods for deliberation, compromise, and execution within DAOs, from quadratic voting to impact certificates to liquid democracy.

Expanding coordination capacity: As smart contracts may become more programmable, auditable, and effective, these experiments may gain traction in the global economy. Coordination and collective action failures are at the core of many of humanity's greatest challenges: collectively mitigating climate change, pandemics, and risks from transformative technologies, addressing health and financial inequities, and the like. DAOs that balance human input, democratic decision-making, and shared ownership with coordination capacity augmented by AI algorithms could form a new frontier in tackling these challenges.

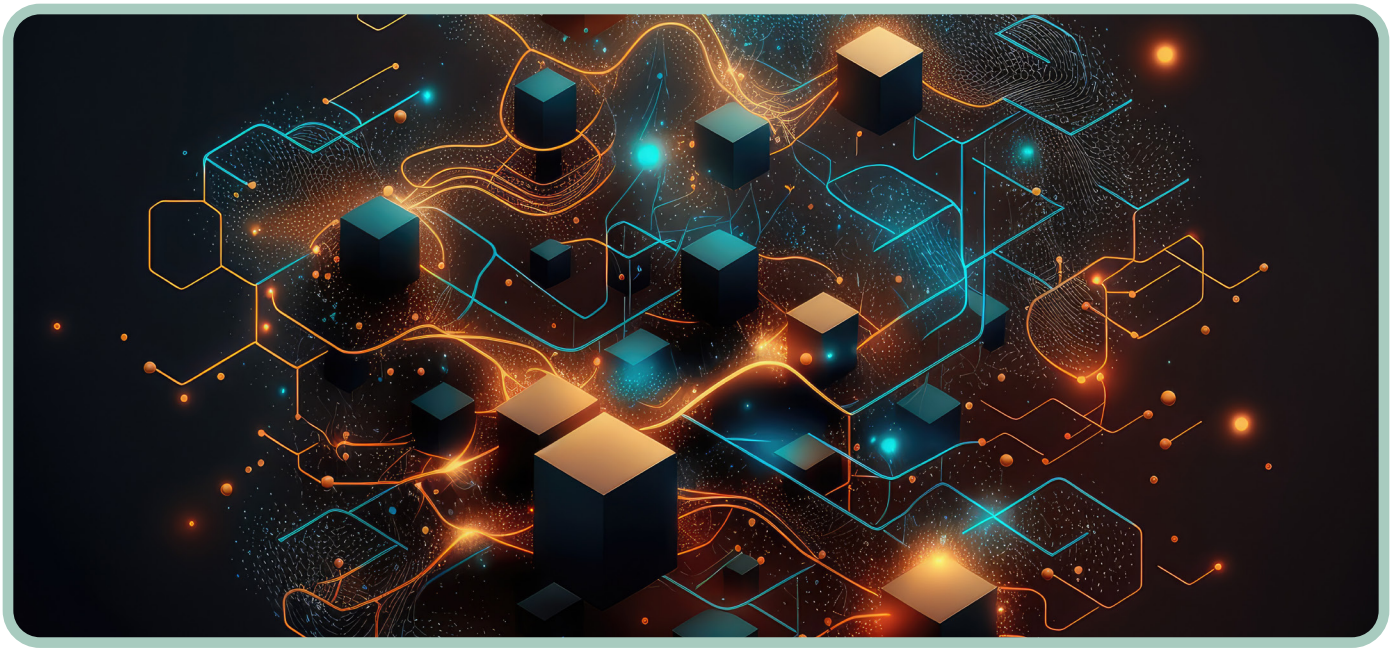
4.3.2 The Risks

Eroding self-determination and non-domination: Many DAOs accept venture capital investment and are also run on coin-based voting paradigms that assign governance rights to token-holders in proportion to tokens held without the usual regulatory provisions (e.g., poison pills, fiduciary responsibilities, antitrust) that prevent well-known attacks like tunneling and looting where someone gains control of 51% of shares and expropriates the other holders.[55] Risks in this direction currently manifest in venture capitalist voting rights in certain profit-focused, coin-voting-governed DAOs; in the future, one can easily imagine entire sectors run by DAOs and governed by venture capitalists, with little public or democratic oversight.

Shrinking collective ownership: More broadly, DAOs can enable sidestepping of nation-state regulation, tax provisions, and oversight. Carefully crafted global governance schemes (such as the minimum corporate tax rate) are sidestepped by proliferating DAOs. Assets become further concentrated in already-asset-owning classes, as shared ownership amongst the wealthy enables increasing returns to be consistently captured by a shrinking group.

Eroding egalitarian pluralism: This could see its logical conclusion in extreme concentration and automation. As artificial algorithms improve, AI-driven smart-contract based DAOs could gain increasing control over the global

55 Aponte-Novoa et al., "The 51% Attack on Blockchains: A Mining Behavior Study"; McShane, "What Is a 51% Attack?"



economy, with dwindling human or organizational oversight. There would be little view into what objectives autonomous entities might be optimizing for, greater automation leading to diminishing human autonomy. This could lead to catastrophic, 'flash-crash' style outcomes, runaway killer optimization, and exponentially increasing inequality and precarity.

4.3.3 Decision Points and Recommendations

In practice, as with many other philosophically driven technological concepts (such as AGI), DAOs are ill-defined. Many of the building blocks listed above are in the early stages of technological development (e.g., identity, internal dispute, governance mechanisms) or have thus far failed in practice (the imaginary of autonomous smart contracts running organizations is far from realized). Most DAOs that exist operate on a mix of on- and off-chain governance, have significant human intervention in their functioning, and may not even utilize collective wallets, smart contracts, etc.[56] As DAOs develop, technologists, developers, investors, and regulators will face a range of decisions with regards to their remit, scope, and structure. These decisions should be made to best mitigate risks and seize opportunities.

Governance structure: The coin-voting default for DAO governance is a recipe for runaway plutocracy. Democratic alternatives are necessary; at the same time, existing 'protocols' for democracy (majority vote, basic representative structures) require significant innovation to enable human flourishing. We recommend experimentation across possible collective intelligence building blocks, from crypto-economic incentives (non-transferrable tokens, reputation scores, voting mechanisms, token-based rewards), both as possible hedges against plutocracy and as pathways towards radically equitable political economies.

Interaction with existing institutions: Protocols must be 'sufficiently decentralized or functional' to avoid the securities label.[57] Sufficient decentralization is not yet defined precisely, and existing proposals largely cover technicalities such as the existence of a secondary market for tokens. In the interest of economic security and self-determination, we recommend that sufficient decentralization not only include provisions regarding protocol security and distribution,

56 Faqir-Rhazoui et al., "A Comparative Analysis of the Platforms for Decentralized Autonomous Organizations in the Ethereum Blockchain"; Santana and Albareda, "Blockchain and the Emergence of Decentralized Autonomous Organizations (DAOs): An Integrative Model and Research Agenda."

57 Hinman, "Digital Asset Transactions: When Howey Met Gary (Plastic)"; Peirce, "Running on Empty: A Proposal to Fill the Gap Between Regulation and Decentralization"; Crenshaw, "Statement on DeFi Risks, Regulations, and Opportunities."

but also social and economic decentralization, for example in the form of cross-sector and cross-community input.

Enabling autonomous futures: It is crucial to invest in and stress-test smart contract auditing capabilities. As DAOs proliferate, and smart contract functionality becomes more self-sufficient, auditing is the first line of defense against runaway optimization. Famously, the first DAO was hacked and lost investors millions of dollars, leading to a fork of Ethereum, a core blockchain; future scenarios may be of greater consequence and far harder to track.[58] Audits are not a purely technical phenomenon. Understanding and transparency are crucial inputs to self-determination and non-domination. The failure of current technocracy can be the strength of decentralized alternatives, but only through transparent mechanisms.

4.4 Prudent Vigilance Applied to Identity and Reputation Infrastructure

The Web3 space has enabled a proliferation of solutions to digital and decentralized identity, a missing layer in our existing digital ecosystem that has been filled by a patchwork of large-scale corporations, nation-states, and start-ups. Proposed Web3 identity and reputation infrastructure might be built on revocable or irrevocable shared tokens, stored in custodial or non-custodial wallets, enabling the attestation of identity and reputation claims—for example, institutional affiliation, community memberships, endorsements, etc.[59]

4.4.1 The Opportunity

Enabling self-determination and egalitarian pluralism: The modern world is built on the ability to grant and revoke access to rights and responsibilities through identification, authentication, and authorization mechanisms. These mechanisms mediate access to public benefits and the ability to vote, travel and cross borders, attend school, prove education or employment credentials, purchase land or housing, even enter into many communities. Offline, complex capacities have evolved over centuries to build layered mechanisms for adjudicating and evaluating trust-based relationships, from the nation-state and institutional level down to the level of families, communities, and neighborhoods. Technological substrates for identity have not kept up. Decentralized and blockchain-based identity primitives can enable richer representations of socially programmable identity across contexts and use cases.[60]

Expanding coordination capacity: Officially recognized forms of ID have enabled various forms of coordination, decision-making, and collaboration at scale, from national democratic regimes to international travel and globalized finance. Further, the proliferation of identity and reputation systems on the web outside of the strict control of nation-states (from Facebook and Twitter to PGP keys) has accelerated communication and knowledge creation, forming a networked social infrastructure that has allowed for a new kind of participative politics. Expanding the scope of identity and reputation systems could provide a rich new layer for coordination and innovation across borders, sectors, and modalities.

4.4.2 The Risks

Enabling domination and eroding self-determination: Identity systems can express major vulnerabilities, most pertinently (1) privacy concerns and data misuse; (2) the risk of creating exclusions to the system, with significant adverse social and political effects; and (3) enabling of systematized exploitation or oppression (as with minority targeting, religious persecution, etc.). Publicizing, formalizing, and otherwise making available identity attributes may contribute to points 1 and 3, while moving in the direction of mandating certain identity infrastructures may cause exclusionary risk as in point 2.

4.4.3 Decision Points and Recommendations

Negative reputation: Negative reputation attestations serve an important societal function. For

58 del Castillo, "The DAO Attacked: Code Issue Leads to $60 Million Ether Theft,"

59 Weyl, Ohlhaver, and Buterin, "Decentralized Society: Finding Web3's Soul."

60 Weyl, Ohlhaver, and Buterin, "Decentralized Society: Finding Web3's Soul."

> “ Well-being, anchored in positive freedom, will depend on political equality, non-domination, and anti-monopoly efforts while sustainably delivering broad-based economic security and an experience of community in societies that simultaneously embrace pluralism and egalitarianism. The tech concepts to which those philosophical concepts have an affinity constitute a set of core ideas that might help establish the guardrails for a socially valuable version of Web3 development.”

example, in cases of police violence, it is crucial that the offending officer be somehow tagged with appropriate information about wrongdoing so as to alert future police precincts of past history; negative reputation is also useful for far lower-stakes situations, such as online reviews. However, negative reputation may also present dangers, as certain communities or individuals might become unwillingly and irrevocably tagged with negative information, reifying many of the risks above.

Privacy and contextual integrity: These systems have privacy tradeoffs. Useful reputation requires a level of visibility and interoperability—one must be able to see and use the information provided. At the same time, existing systems often make individuals and communities far too transparent, particularly to centers of power, like the state and corporations. To mitigate these tradeoffs, we recommend a principle of minimum disclosure and contextual integrity. Removing context leads to reliance on "universally secure" identifiers based on clean/universal features like biometrics, which often raise at least as many concerns as the centralized protocols they replace. We further recommend limiting linkability across attributes in the view of external organizations.

4.5 Prudent Vigilance Applied to Funding Mechanisms

The growth of Web3 infrastructure has brought with it experimentation in novel methods of resource allocation, collective financing, crowdfunding, and value accounting. The past several years have seen exponential growth in matching funds and grant programs, which are rapidly scaling, diversifying, and, at least in some cases, improving. These mechanisms may combine smart contract and wallet functionality with funds disbursed through democratic, delegative, or representative voting, allowing for communities to choose project recipients for public goods support. The development of crypto-native collective funding mechanisms (notably quadratic funding, retroactive public goods funding, impact certificates, etc.) have expanded the scope of possible financing innovation.[61] In addition, experiments are underway with ecosystem-level rules to create collective funding pools similar to taxation, codifying allocations or creating Layer 2 infrastructural solutions that explicitly dictate resource distribution to specific public goods.

4.5.1 The Opportunity

Enabling coordination capacity and economic security: Many digital goods exhibit "supermodularity," where it is (within some group) cheaper on a per-person basis to supply a good or service to many people than to few: the whole is greater than the sum of the parts. Supermodular goods are poorly provided by standard capitalism (as they tend to be monopolized and underfunded) and by nation-states (whose boundaries rarely line up with those who can be served). These goods are chronically underfunded—open-source software, basic science research, shared infrastructure, etc. Greater capacity to provide capital and oversight in these areas could have compounding network effects, underlying decommodified economic structures and efficient provision across sectors.

Expanding non-domination: The expanded possibilities here are significant. For example, imagine if democratic matching-fund mechanisms were available to for-profit as well as non-profit entities. A range of corporations may then receive at least some amount of matched funding, which could be accompanied by some form of governance rights. For-profit cooperatives might flourish, with partial philanthropic funding, community-managed enterprises might benefit, or even programs democratically determined to be positive-sum within traditional corporate structures. These funds would no longer be targeted to pure public goods, meaning that they would be enabling greater excludability than other funding opportunities. However, in return, the range of impact would be greatly widened—one can see this as a form of trading in perfect availability for greater applicability. This is a far more tractable way of engendering stakeholder or democratic governance over private entities in most cases than either nationalization or internal advocacy; further, it does not rely on a lack of monopolies, preferring accountability and oversight over disaggregation.

61 Buterin, Hitzig, and Weyl, "A Flexible Design for Funding Public Goods."

4.5.2 *The Risks*

Eroding coordination capacity and economic security: There is already very little public transparency with regards to the development, funding, and maintenance of public goods. Relying on Web3-based structures to fund basic infrastructure could exacerbate this issue, resulting in a funding ecosystem in which necessary goods are funded arbitrarily and at-will by non-transparent groups of token-holders. The systems could also be hijacked by collusive groups of trolls, such as extremist or terrorist organizations that are (rightly) not funded by existing public goods infrastructures.

5. Conclusion

We are aware that in this paper we have just begun to tackle the complex problem of defining the ethical landscape in which decisions about decentralized social technology must be made. Ultimately, there are a host of minute decisions to be made by technologists, funders, legislators, and users, among other stakeholders. In future work, we plan to build out guidance for specific categories of stakeholder. Moreover, decision-making contexts are various depending on whether the innovation grows out of the blockchain community, the Fediverse, or another Web3 context.

Here we have sought to paint on a big canvas, rather than tackling the many decision-making moments that are pertinent. We have sought to bring to the surface how the question of what to do, or what to build, is never a neutral question. Normative judgments are necessarily made in answering that question. But normative judgments are also made when we answer the question of *how* to build whatever it is we decide to build in the DST space. This is because the choices of *how* to build, once scaled to a global level, will have institutional consequences with significant downstream effects.

We believe the new technical possibilities hold such promise for the achievement of solutions to perplexing human collective action problems that a duty to experiment follows from that promise. Yet we do not endorse unbridled or unconstrained experimentation. Rather we propose the constraints of prudent vigilance.

The goal is to lay out hypotheses about what to build, and how to build it and to test them for alignment to the principles laid out above; these are bedrocks for human flourishing. We should feel confident that there is a strong case for alignment before we begin to build, but if we find that we are in error, and that our hypotheses have been proven false, we should be prepared to abandon them and try again.

6. Works Cited

Ackerman, Bruce. *We the People: Foundations*. Cambridge, MA: Belknap Press, 1991.

Alexander, John M. *Capabilities and Social Justice: The Political Philosophy of Amartya Sen and Martha Nussbaum*. Aldershot: Ashgate Publishing, 2008.

Ali, Omar, Ashraf Jaradat, Atik Kulakli, and Ahmed Abuhalimeh. "A Comparative Study: Blockchain Technology Utilization Benefits, Challenges and Functionalities." *IEEE Access* 9 (2021): 12730–749

Allen, Colin, Wendell Wallach, and Iva Smit. "Why Machine Ethics?" *IEEE Intelligent Systems* 21, no. 4 (2006): 12–17.

Allen, Danielle. *Justice by Means of Democracy*. Chicago: University of Chicago Press, 2023.

Allen, Danielle. "A New Theory of Justice: Difference without Domination." In *Difference without Domination: On Justice in Diverse Democracies*, 27–58. Chicago: University of Chicago Press, 2020.

Allen, Danielle. *Talking to Strangers: Anxieties of Citizenship since* Brown v. Board of Education. Chicago: University of Chicago Press, 2004.

Allen, Danielle. "Toward a Connected Society." In *Our Compelling Interests: The Value of Diversity for Democracy and a Prosperous Society*, edited by Earl Lewis and Nancy Kantor, 71–106. Princeton, NJ: Princeton University Press, 2016.

Allen, Danielle, and Rohini Somanathan, ed. *Difference without Domination: Pursuing Justice in Diverse Democracies*. Chicago: University of Chicago Press, 2020.

Allen, Danielle, et al., ed. *A Political Economy of Justice*. Chicago: University of Chicago Press, 2022.

Anderson, Michael, and Susan Leigh Anderson, ed. *Machine Ethics*. Cambridge: Cambridge University Press, 2011.

Aponte-Novoa, Fredy Andres, et al. "The 51% Attack on Blockchains: A Mining Behavior Study." *IEEE Access* 9 (2021): 140549–64. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9567686

Barocas, Solon, and Andrew Selbst. "Big Data's Disparate Impact," *California Law Review* 104 (2016): 671–732.

Birhane, Abeba, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. "The Values Encoded in Machine Learning Research." In *FAcct: 2022 ACM Conference on Fairness, Accountability, and Transparency*, 173–84. New York: Association for Computing Machinery, 2022. https://doi.org/10.1145/3531146.3533083

Bogusz, Claire Ingram, et al. "Tracking the Digital Evolution of Entrepreneurial Finance: The Interplay Between Crowdfunding, Blockchain Technologies, Cryptocurrencies, and Initial Coin Offerings." *IEEE Transactions on Engineering Management* 67, no. 4 (2020): 1099–108.

Brundage, Miles. "Limitations and Risks of Machine Ethics." *Journal of Experimental & Theoretical Artificial Intelligence* 26 (2014): 355–72.

Buterin, Vitalik. "Why Sharding Is Great: Demystifying the Technical Properties." Posted April 7 2021. vitalik.ca/general/2021/04/07/sharding.html.

Buterin, Vitalik, Zoë Hitzig, and Glen Weyl. "A Flexible Design for Funding Public Goods." *Management Science* 65, no. 11 (2019): 4951–5448.

Chouldechova, Alexandra. “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments.” *Big Data* 5, no. 2 (2017). https://doi.org/10.48550/arXiv.1610.07524

Chouldechova Alexandra, and Aaron Roth. “The Frontiers of Fairness in Machine Learning.” ArXiv, posted Oct. 20, 2018. https://doi.org/10.48550/arXiv.1810.08810

Consumer Financial Protection Bureau. “Complaint Bulletin: An Analysis of Consumer Complaints Related to Crypto-Assets.” November 2022. https://files.consumerfinance.gov/f/documents/cfpb_complaint-bulletin_crypto-assets_2022-11.pdf

Crenshaw, Caroline A. “Statement on DeFi Risks, Regulations, and Opportunities.” *International Journal of Blockchain Law* 1 (Nov. 2021).

Crenshaw, Kimberlé. “Demarginalizing the Intersection of Race and Sex.” In *Living with Contradictions: Controversies in Feminist Social Ethics*, 1st ed., by Alison M. Jaggar, 39–52. New York: Routledge, 1994.

Dang, Sanjit Singh. “Understanding the Blockchain Layered Architecture to Solve the Scalability Challenges.” *Forbes*, October 24 2022. www.forbes.com/sites/sanjitsinghdang/2022/10/24/understanding-the-blockchain-layers-to-solve-the-scalability-challenges

Davidson, James Dale, and William Rees-Mogg. *The Sovereign Individual: How to Survive and Thrive During the Collapse of the Welfare State.* New York: Simon & Schuster, 1997.

De Filippi, Primavera, and Aaron Wright, *Blockchain and the Law: The Rule of Code*. Cambridge, MA: Harvard University Press, 2018.

de Soto, Hernando. *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else.* New York: Basic Books, 2000.

del Castillo, Michael. “The DAO Attacked: Code Issue Leads to $60 Million Ether Theft.” *CoinDesk*, June 17 2016. www.coindesk.com/markets/2016/06/17/the-dao-attacked-code-issue-leads-to-60-million-ether-theft

Dewey, John. *The Public and Its Problems.* New York: Holt Publishers, 1927.

Estlund, David. “Debate: Liberalism, Equality, and Fraternity in Cohen’s Critique of Rawls.” *Journal of Political Philosophy* 6, no. 1 (1998), https://doi.org/10.1111/1467-9760.00048

Faqir-Rhazoui, Youssef, et al. “A Comparative Analysis of the Platforms for Decentralized Autonomous Organizations in the Ethereum Blockchain.” *Journal of Internet Services and Applications* 12, no. 1 (2021). https://doi.org/10.1186/s13174-021-00139-6

Farrell, Henry, and Cosma Rohilla Shalizi. “Pursuing Cognitive Democracy.” In *From Voice to Influence: Understanding Citizenship in a Digital Age*, edited by Danielle Allen and Jennifer S. Light, 209–31. Chicago: University of Chicago Press, 2015.

Fletcher, Emma. “Reports Show Scammers Cashing in on Crypto Craze.” *Federal Trade Commission* June 3, 2022. www.ftc.gov/news-events/data-visualizations/data-spotlight/2022/06/reports-show-scammers-cashing-crypto-craze#crypto1

George, Benedict. “What Is a Dapp? Decentralized Apps Explained.” *CoinDesk*, Nov. 7, 2022. https://www.coindesk.com/learn/what-is-a-dapp-decentralized-apps-explained

Granovetter, Marc S. “The Strength of Weak Ties.” *American Journal of Sociology* 78, no. 6 (1973): 1360–80.

Gregory, Chris. *Star Trek: Parallel Narratives*. New York: Palgrave Macmillan, 2000.

Hagendorff, Thilo. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds and Machines* 30 (2020): 99–120. link.springer.com/article/10.1007/s11023-020-09517-8

Hinman, William. "Digital Asset Transactions: When Howey Met Gary (Plastic)." U.S. Securities and Exchange Commission, June 14, 2018. www.sec.gov/news/speech/speech-hinman-061418

Investing.com. "Ethereum Processes 4.5x More Transactions Than Visa by CoinEdition." Cryptocurrency News, Feb. 21, 2022. www.investing.com/news/cryptocurrency-news/ethereum-processes-45x-more-transactions-than-visa-2768337

Ishmaev, G. "Sovereignty, privacy, and ethics in blockchain-based identity management systems." *Ethics Inf Technol* **23**, 239–252 (2021). https://doi.org/10.1007/s10676-020-09563-x

Kaspersen, Anja, and Wendell Wallach. "Why Are We Failing at the Ethics of AI?" Carnegie Council, Artificial Intelligence & Equality Initiative, Nov. 10 2021. www.carnegiecouncil.org/media/article/why-are-we-failing-at-the-ethics-of-ai

King, Thomas C., et al. "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions." *Science and Engineering Ethics* 26 (2020): 89–120.

Knight, Will. "Biased Algorithms Are Everywhere, and No One Seems to Care." *MIT Technology Review*, July 12, 2017. www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care

Lanier, Jaron, and Glen Weyl. "How Civic Technology Can Help Stop a Pandemic." *Foreign Affairs*, March 20, 2020. www.foreignaffairs.com/articles/asia/2020-03-20/how-civic-technology-can-help-stop-pandemic

Lewis, Anna C. F. "Where Bioethics Meets Machine Ethics." *American Journal of Bioethics* 20, no. 11 (2020): 22–24. https://doi.org/10.1080/15265161.2020.1819471

Licklider, J.C. R. "Computers and Government." In *Computers and the Future: A Twenty-Year View*, edited by Michael L. Dertouzos and Joel Moses, 87–110. Cambridge, MA: MIT Press, 1979.

Licklider, J. C. R. "Topics for Discussion at the Forthcoming Meeting." ARPA, April 23, 1963. Archived at https://perma.cc/JCT5-846G

Lippert-Rasmussen, Kasper. "Pluralist Egalitarianism." In *Relational Egalitarianism: Living as Equals*, 181–210. Cambridge, MA: Cambridge University Press, 2018.

Marcobello, Mason. "What Are Layer 2s and Why Are They Important?" *CoinDesk*, June 29 2022. www.coindesk.com/learn/what-are-layer-2s-and-why-are-they-important

Mazières, David. "The Stellar Consensus Protocol: A Federated Model for Internet-level Consensus." Stellar Development Foundation, Dec. 6, 2017.

McShane, Griffin. "What is a 51% Attack?" *CoinDesk*, Oct. 12, 2021. www.coindesk.com/learn/what-is-a-51-attack

Metz, Cade, and Karen Weise. "A Tech Race Begins as Microsoft Adds A.I. to Its Search Engine." *New York Times*, February 7, 2023, sec. Technology. www.nytimes.com/2023/02/07/technology/microsoft-ai-chatgpt-bing.html

Moessinger. Pierre. "An Overview and a Response to Vitalik Buterin." *Medium*, blog posted Oct. 23, 2020. moresoon.medium.com/trustless-an-overview-and-a-response-to-vitalik-buterin-

16f2eb1aa2a3

Nussbaum, Martha Craven. *Creating Capabilities: The Human Development Approach*. Cambridge, MA: Belknap Press, 2011.

Parens, Erik, Josephine Johnston, and Jacob Moses. *Ethical Issues in Synthetic Biology: An Overview of the Debates*. Woodrow Wilson International Center for Scholars, June 2009. www.wilsoncenter.org/sites/default/files/media/documents/publication/synbio3.pdf

Peirce, Hester M. "Running on Empty: A Proposal to Fill the Gap Between Regulation and Decentralization." U.S. Securities and Exchange Commission, Feb. 6, 2020 [speech]. https://www.sec.gov/news/speech/peirce-remarks-blockress-2020-02-06

Pettit, Philip. *Just Freedom: A Moral Compass for a Complex World*. New York: Norton Books, 2014.

Presidential Commission for the Study of Bioethical Issues. *New Directions: The Ethics of Synthetic Biology and Emerging Technologies*. Report, Dec. 2010. https://bioethicsarchive.georgetown.edu/pcsbi/sites/default/files/PCSBI-Synthetic-Biology-Report-12.16.10_0.pdf

Rahman, K. Sabeel. *Democracy Against Domination*. New York: Oxford University Press, 2017.

Rodrik, Dani, and Charles Sabel. "Building a Good Jobs Economy." In *A Political Economy of Justice*, edited by Danielle Allen, Yochai Benkler, Leah Downey, Rebecca Henderson and Josh Simons, pp. 61–95. Chicago: University of Chicago Press, 2022.

Santana, Carlos, and Laura Albareda. "Blockchain and the Emergence of Decentralized Autonomous Organizations (DAOs): An Integrative Model and Research Agenda." *Technological Forecasting & Social Change* 182 (2022).

Sen, Amartya. "Democracy as a Universal Value." *Journal of Democracy* 10, no. 3 (1999): 3–17.

Seyedsayamdost, Elham, and Peter Vanderwal, "From Good Governance to Governance for Good: Blockchain for Social Impact." *Journal of International Development* 32, no. 6 (2020): 943–60. https://doi.org/10.1002/jid.3485

Simmel, Georg. *Conflict & The Web of Group-Affiliations*. Translated by Kurt H. Wolff and Reinhard Bendix. 1922. Reprint, New York: Free Press, 1964.

Simon, Herbert. "Response to Philippe van Parijs on Universal Basic Income." *Boston Review*, July 17, 2000. bostonreview.net/forum/basic-income-all/herbert-simon-ubi-and-flat-tax

Simons, Josh. *Algorithms for the People: Democracy in the Age of AI*. Princeton, NJ: Princeton University Press, 2022.

Sims, Alexandra. "Blockchain and Decentralised Autonomous Organisations (DAOs): The Evolution of Companies?" *New Zealand Universities Law Review* 28, no. 3 (Nov. 1 2019). https://doi.org/10.2139/ssrn.3524674

Srinavasan, Balaji S. *The Network State: How to Start a New Country*. 2022.

Stephenson, Neal. *Snow Crash*. New York: Bantam Books, 1992.

Szreter, Simon, and Michael Woolcock. "Health by Association? Social Capital, Social Theory, and the Political Economy of Public Health." *International Journal of Epidemiology* 33, no. 4 (2004): 650–67.

Trozze, Arianna, Josh Kamps, Eray Arda Akartuna, et al. "Cryptocurrencies and future financial crime." *Crime Science* 11, no. 1 (2022). https://doi.org/10.1186/s40163-021-00163-8

Vaughan, Jennifer Wortman, and Hanna Wallach. “A Human-Centered Agenda for Intelligible Machine Learning.” *Machines We Trust: Perspectives on Dependable AI*, edited by Marcello Pelillo and Teresa Scantamburlo, 123–38. Cambridge: MIT Press, 2020.

Waldrop, Mitch. *The Dream Machine: J. C. R. Licklider and the Revolution That Made Computing Personal*. New York: Viking Penguin, 2001.

Wallach, Wendell, and Anja Kaspersen. “Creative Reflections on the History & Role of AI Ethics.” Carnegie Council, Artificial Intelligence & Equality Initiative, May 26 2021. www.carnegiecouncil.org/media/series/aiei/20210426-creative-reflections-history-role-artificial-intelligence-ethics-wendell-wallach

Weyl, E. Glen, Puja Ohlhaver, and Vitalik Buterin. “Decentralized Society: Finding Web3’s Soul.” SSRN, posted May 11, 2022. http://dx.doi.org/10.2139/ssrn.4105763

Wright, Aaron. “The Rise of Decentralized Autonomous Organizations: Opportunities and Challenges.” *Stanford Journal of Blockchain Law & Policy* (June 30, 2021). https://stanford-jblp.pubpub.org/pub/rise-of-daos/release/1