Decentralized Society: Finding Web3's Soul[1]

E. Glen Weyl,[2] Puja Ohlhaver,[3] Vitalik Buterin[4]

May 2022

"The Dao is the hearth and home
of the ten thousand things.
Good souls treasure it,
lost souls find shelter in it."
— Laozi, #62

Abstract

Web3 today centers around expressing transferable, financialized assets, rather than encoding social relationships of trust. Yet many core economic activities—such as uncollateralized lending and building personal brands—are built on persistent, non-transferable relationships. In this paper, we illustrate how non-transferable "soulbound" tokens (SBTs) representing the commitments, credentials, and affiliations of "Souls" can encode the trust networks of the real economy to establish provenance and reputation. More importantly, SBTs enable other applications of increasing ambition, such as community wallet recovery, sybil-resistant governance, mechanisms for decentralization, and novel markets with decomposable, shared rights. We call this richer, pluralistic ecosystem "Decentralized Society" (DeSoc)—a co-determined sociality, where Souls and communities come together bottom-up, as emergent properties of each other to co-create plural network goods and intelligences, at a range of scales. Key to this sociality is decomposable property rights and enhanced governance mechanisms—such as quadratic funding discounted by correlation scores—that reward trust and cooperation while protecting networks from capture, extraction, and domination. With such augmented sociality, web3 can eschew today's hyper-financialization in favor of a more transformative, pluralist future of increasing returns across social distance.

[1] We are grateful to Audrey Tang, Phil Daian, Danielle Allen, Leon Erichsen, Matthew Prewitt, Divya Siddarth, Jaron Lanier, and Robert Miller for their thoughtful feedback and comments. All errors and views are our own.

[2] Microsoft Corporation & RadicalXChange Foundation, glen@radicalxchange.org. Glen vinicula este documento a su Alma.

[3] Flashbots Ltd., puja@flashbots.net. Puja dedicates this paper to her grandmother, Satya, whose love and light will always shine on many Souls.

[4] Ethereum Foundation, vitalik.buterin@ethereum.org.

§1 INTRODUCTION

Web3 has stunned the world by forging a parallel system of finance of unprecedented flexibility and creativity in less than a decade. Cryptographic and economic primitives such as public key cryptography, smart contracts, proof of work, and proof of stake have led to a sophisticated and open ecosystem for expressing financial transactions.

Yet the economic value finance trades on is generated by humans and their relationships. Because web3 lacks primitives to represent such social identity, it has become fundamentally dependent on the very centralized web2 structures it aims to transcend, replicating their limitations.

Examples of these dependencies include:

1. Most NFT artists rely on centralized platforms like OpenSea and Twitter to commit to scarcity and initial **provenance**.

2. DAOs that try to move beyond simple coin-voting often rely on web2 infrastructure, such as social media profiles, for **sybil resistance**.

3. Many web3 participants rely on custodial wallets managed by centralized entities like Coinbase or Binance. Decentralized **key management** systems are not user-friendly for any but the most sophisticated.

Furthermore, the lack a native web3 identity makes today's DeFi ecosystem unable to support activities ubiquitous in the real economy, such as **undercollateralized lending** or simple contracts, like an **apartment lease**. In this paper, we illustrate how even small and incremental steps towards representing social identity with soulbound tokens could overcome these limitations and bring the ecosystem far closer to regenerating markets with their underpinning human relationships in a native web3 context.

Even more promising, we highlight how native web3 social identity, with rich social *composability*, could yield great progress on broader long-standing problems in web3 around wealth concentration and vulnerability of governance to financial attacks, while spurring a Cambrian explosion of innovative political, economic, and social applications. We refer to these use cases and the richer pluralistic ecosystem that they enable as **"Decentralized Society" (DeSoc)**.

§2 OUTLINE

We begin by explaining the primitives of DeSoc, centered around accounts (or wallets) holding **non-transferable (initially public) "soulbound" tokens (SBTs)** representing commitments, credentials, and affiliations. Such tokens would be like an extended resume, issued by other wallets that attest to these social relations.

We then describe a "stairway" of increasingly ambitious applications across the social stack such primitives could empower, including:

- establishing provenance
- unlocking undercollateralized lending markets through reputation
- enabling decentralized key management
- thwarting and compensating for coordinated strategic behavior
- measuring decentralization
- creating novel markets with decomposable, shared rights and permissions

This description culminates with a vision of DeSoc—a co-determined sociality, where Souls and communities come together bottom-up, as emergent properties of each other to co-create plural network goods, including plural intelligences, at a range of social scales.

Finally, we answer several potential concerns and objections, and make comparisons to other identity paradigms familiar in the web3 space, conceding often how our vision is just a first step but nonetheless an advance in programmable privacy and communication. Then, we consider technical pathways to bootstrap the vision we imagine. Building off these, we look forward, more philosophically, to the potential of DeSoc to redirect web3 to a more profound, legitimate, and transformative path.

§3 SOULS

Our key primitive is accounts, or wallets, that hold ***publicly visible***, ***non-transferable*** **(but possibly *revocable*-by-the-issuer) tokens**.[5] We refer to the accounts as **"Souls"** and tokens held by the accounts as **"Soulbound Tokens" (SBTs)**. We initially assume publicity despite our deep interest in privacy because it is technically simpler to validate as a proof-of-concept, even if limited by the subset of tokens people are willing to publicly share. Later in the paper, we introduce the concept of "programmable privacy" for richer use cases.

Imagine a world where most participants have Souls that store SBTs corresponding to a series of affiliations, memberships, and credentials. For example, a person might have a Soul that stores SBTs representing educational credentials, employment history, or hashes of their writings or works of art. In their simplest form, these SBTs can be "self-certified," similar to how we share information about ourselves in our CVs. But the true power of this mechanism emerges when SBTs held by one Soul can be issued—or attested—by other Souls, who are counterparties to these relationships. These counterparty Souls could be individuals, companies, or institutions. For example, the Ethereum Foundation could be a Soul that issues SBTs to Souls who attended a developer conference. A university could be a Soul that issues SBTs to

[5] We have chosen this set of properties not because they are clearly the most desirable collection of characteristics, but because they are easy to implement in the current environment and permit significant functionality. We explore programmably private SBTs in Section 5.3.

graduates. A stadium could be a Soul that issues SBTs to longtime Dodgers fans.

Note there is no requirement for a Soul to be linked to a legal name, or for there to be *any* protocol-level attempt to ensure "one Soul per human." A Soul could be a persistent pseudonym with a range of SBTs that cannot easily be linked.[6] We also do not assume non-transferability of Souls across humans. Instead, we try to illustrate how these properties, where needed, can naturally emerge from the design itself.

§4 STAIRWAY TO DESOC

4.1 Art & Soul

Souls are a natural way for artists to stake their reputation on their works. When issuing a tradeable NFT, an artist could issue the NFT from their Soul. The more SBTs the artist's Soul carries, the easier it would be for buyers to identify the Soul as belonging to that artist, and thereby also confirm the NFT's legitimacy. Artists could go a step further to issue a linked SBT stored in their Soul that attests to the NFT's membership to a "collection" and vouches for whatever scarcity limits the artist wishes to set. Souls would thus create a verifiable, on-chain way to stake and build reputation on the provenance and scarcity of an object.

Applications extend beyond art, to services, rentals, and any market built on scarcity, reputation, or authenticity. An example of the latter is verifying the authenticity of purported factual recordings, such as photographs and videos. With advances in deep fake technology, direct inspection by both humans and algorithms will increasingly fail to detect veracity. While **blockchain inclusion enables us to trace the time a particular work was made, SBTs would enable us to trace the *social provenance***, giving us rich social context to the Soul that issued the work—their constellation of memberships, affiliations, credentials—and their social distance to the subject. "Deep fakes" could be readily identified as those artifacts originated outside of time and social context, while trusted artifacts (like photographs) would emerge from the attestation of reputable photographers. Whereas present technology de-contextualizes cultural products (like pictures) and opens them to unchecked, viral attacks lacking social context, SBTs can recontextualize such objects and empower Souls to take advantage of trust relationships already present within communities as a meaningful backstop to protect reputation.

4.2 Soul Lending

Perhaps the largest financial value built directly on reputation is credit and uncollateralized lending. Currently, the web3 ecosystem cannot replicate simple forms of uncollateralized lending, because all assets

[6] Note, however, that in principle legal names could be represented themselves as SBTs: a family name would be a membership SBT to a family group and a given name could be a gifted SBT from parents to their child. In fact, richer notions of names would be easy to represent if, for example, other family lines or relations gifted membership SBTs to a new child.

are transferable and saleable—thus simply forms of collateral. The "traditional" financial ecosystem supports many forms of uncollateralized lending, but relies on centralized credit scores to gauge creditworthiness of borrowers who have little incentive to share information about their credit history. But such scores have many flaws. At best, they opaquely overweight and underweight factors relevant to creditworthiness, and bias those who haven't accumulated sufficient data—mainly minorities and the poor. At worst, they can enable *Black Mirror* opaque "social credit" systems that engineer social outcomes and reinforce discriminations.

An ecosystem of SBTs could unlock a censorship-resistant, bottom-up alternative to top-down commercial and "social" credit systems. SBTs that represent education credentials, work history, and rental contracts could serve as a persistent record of credit-relevant history, allowing Souls to stake meaningful reputation to avoid collateral requirements and secure a loan. Loans and credit lines could be represented as non-transferable but revocable SBTs, so they are nested amongst a Soul's other SBTs—a kind of non-seizable reputational collateral—until they are repaid and subsequently burned, or better yet, replaced with proof of repayment. SBTs offer useful security properties: non-transferability prevents transferring or hiding outstanding loans, while a rich ecosystem of SBTs ensures that borrowers who try to escape their loans (perhaps by spinning up a fresh Soul) will lack SBTs to meaningfully stake their reputation.

The ease of computing public liabilities with SBTs would open-source lending markets. New correlations between SBTs and repayment risk would emerge, birthing better lending algorithms that predict creditworthiness and thereby reduce the role of centralized, opaque credit-scoring infrastructure. Better yet, lending would likely occur *within* social connections. In particular, SBTs would offer a substrate for community lending practices similar to those pioneered by Muhammad Yunus and the Grameen Bank, where members of a social network agree to support one another's liabilities. Because a Soul's constellation of SBTs represents memberships across social groups, participants could easily discover other Souls who would be valuable co-participants in a group lending project. Whereas commercial lending is a "lend-it-and-forget-it" until repayment model, community lending might take a "lend-it-and-help-it" approach—combining working capital with human capital with greater rates of return.

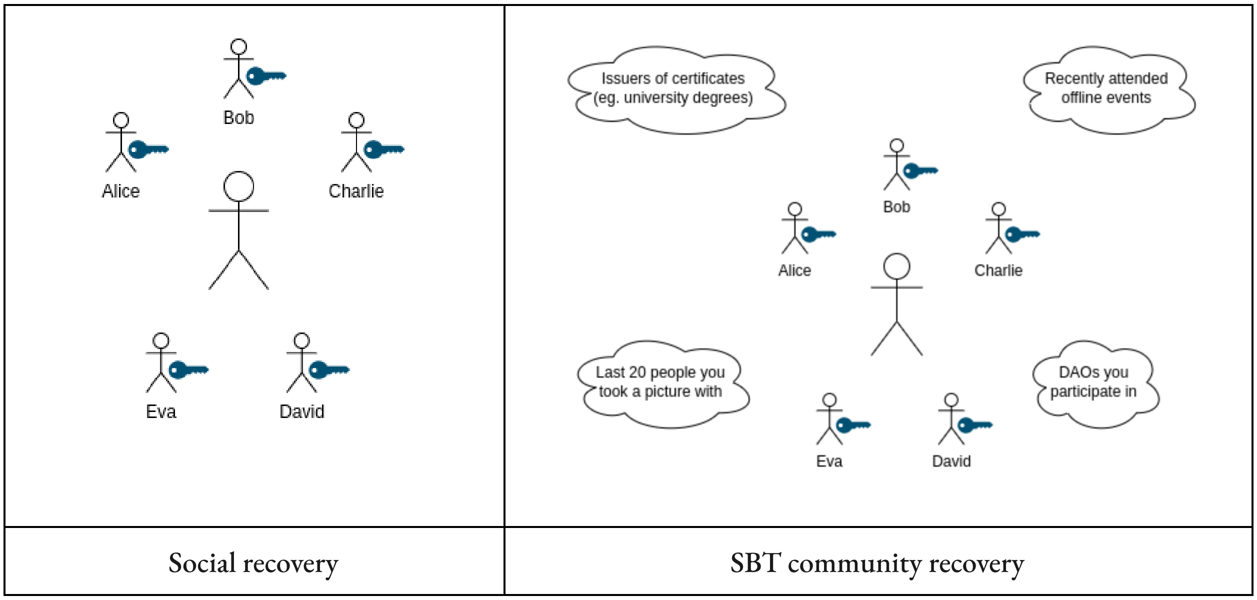
How does uncollateralized community lending get off the ground? At the start, we expect Souls to carry only SBTs that reflect information they are comfortable with sharing publicly, such as information in a CV. While limited in scope, it might be a level of resolution sufficient for intra-community lending experiments to take off, especially if the SBTs are issued by reputable institutions. For example, a constellation of SBTs that show certain programming credentials, participation in several conferences, and work history might be sufficient for a Soul to take a loan (or raise seed capital) for their venture. Such credentials and social relationships already informally play an important, but opaque role in capital allocation like venture capital.

4.3 Not Losing Your Soul

The non-transferability of key SBTs—such as one-time issued education credentials—raises an important question: how do you not lose your Soul? Recovery methods today, like muti-sig recovery or mnemonics, have different tradeoffs in mental overhead, ease of transacting, and security. Social recovery is an emerging alternative that relies on a person's trusted relationships. SBTs allow a similar, but broader paradigm: **community recovery, where the Soul is the intersectional vote of its social network.**

Social recovery is a good starting point for security, but has several drawbacks in security and usability. A user curates a set of "guardians" and gives them the power, by majority, to change the keys of their wallet. Guardians could be a mix of individuals, institutions, or other wallets. The problem is a user must balance the desire for a reasonably high number of guardians against the precaution that guardians be from discrete social circles to avoid collusion. Also, guardians can pass away, relationships sour, or people simply fall out of touch, requiring frequent and attention-taxing updates. While social recovery avoids a single point of failure, successful recovery nonetheless depends on curating and maintaining trusted relationships with a majority of guardians.

A more robust solution is to tie Soul recovery to a Soul's memberships across communities, not curating but instead drawing on a maximally broad set of real-time relationships for security. Recall that SBTs represent memberships to different communities. Some of these communities—like employers, clubs, colleges, or churches—might be more off-chain in nature, while others—like participation in protocol governance or a DAO—might be more on-chain. In a community recovery model, recovering a Soul's private keys would require a member from a qualified majority of a (random subset of) Soul's communities to consent.


Bob
Alice
Charlie
Eva
David
Issuers of certificates (eg. university degrees)
Recently attended offline events
Bob
Alice
Charlie
Last 20 people you took a picture with
DAOs you participate in
Eva
David

Social recovery

SBT community recovery

Like social recovery, we assume that the Soul has access to secure, off-chain communications channels where "authentication"—through conversation, meeting in person, or confirming a shared secret—can occur. Such communication channels would require greater bandwidth (technically the ability to carry richer "information entropy") than, for example, on-chain bots or computation over SBTs themselves. Indeed, we can think of SBTs as fundamentally being all about representing participation or access to such authentic—namely high bandwidth—communication channels.

Precise details to make this work will require experimentation. How guardians are chosen and how many guardians' consent is required, for example, are key security parameters for further research. With such a rich information base, however, community recovery should be computationally possible, with security increasing as a Soul joins more distinct communities and forms more meaningful relationships.

Community recovery, as a security mechanism, embodies the theory of identity proposed by turn-of-the-20th-century sociologist Georg Simmel—founder of social network theory—in which individuality emerges from the intersection of social groups, just as social groups emerge as the intersection of individuals. Maintaining and recovering cryptographic possession of a Soul requires consent of the Soul's network. **By embedding security in sociality**, a Soul can always regenerate their keys through community recovery, which deters Soul theft (or sale): because a Seller would need to prove selling the recovery relationships, any attempt to sell a Soul lacks credibility.

4.4 Souldrops

So far we have explained how Souls can come to represent individuals and reflect their unique traits and solidarities as they acquire SBTs that reflect their affiliations, memberships, and credentials. Such individuation helps Souls build reputations, establish provenance, access uncollateralized lending markets, and protect reputation and identity. But the converse is also true; **SBTs also enable *communities* to be convened at unique intersections of Souls**. Thus far web3 has largely relied on token sales or airdrops to summon new communities, which yield little accuracy or precision. Airdrops, in which tokens are algorithmically given for free to a set of wallets, mostly fall to some combination of existing token holders and wallets—easily attacked by sybils, encouraging strategic behavior and the Matthew effect. SBTs offer a radical improvement we call "souldrops."

"Souldrops" are airdrops based on computations over SBTs and other tokens within a Soul. For example, a DAO that wants to convene a community within a particular layer 1 protocol could souldrop to developers who hold 3 out of the last 5 conference attendance SBTs, or other tokens reflecting attendance like POAPs. Protocols could also programmatically weight token drops across a combination of SBTs. We can imagine a non-profit whose mission is to plant trees dropping governance tokens to Souls who hold a mix of environmental action SBTs, gardening SBTs, and carbon sequestration tokens—perhaps dropping more tokens to the carbon sequestration token-holders.

Souldrops could also introduce novel incentives to encourage community engagement. Dropped

SBTs could be engineered to be soulbound for a period but eventually "vest" into transferable tokens over time. Or the reverse could be true. Transferable tokens held for some period could unlock the right to SBTs that confer further governance rights over a protocol. SBTs open a rich possibility space to experiment with mechanisms that maximize community engagement and other goals, like decentralization, which we discuss further below.

4.5 The DAO of Souls

Distributed autonomous organizations (DAOs) are virtual communities that come together around a common purpose, coordinated by voting through smart contracts on a public blockchain. While DAOs offer great potential for coordination of global communities across distance and difference, they are vulnerable to sybil attacks where a single user can have multiple wallets to accrue voting power—or in less sophisticated one-token-one-vote style governance, simply hoard tokens to accrue 51% voting power and dispossess the other 49%.

DAOs could mitigate sybil attacks with SBTs in several ways, by:

- computing over a Soul's constellation of SBTs to **differentiate between unique Souls and probable bots**, and denying *any* voting power to a Soul that appears to be a Sybil.

- **conferring more voting power to Souls** who hold more reputable SBTs—like work or educational credentials, licenses, or certifications.

- issuing **specialized "proof-of-personhood" SBTs**, which could help other DAOs bootstrap sybil resistance.

- **checking for *correlations* between SBTs** held by Souls who support a particular vote, and applying a lower vote weight to voters who are highly correlated.

The latter idea of correlation checking is particularly promising and novel. A vote supported by many Souls who all share the same SBT(s) is more likely to be a Sybil attack and—even if *not* a Sybil attack—such a vote is more likely to be a group of Souls who are *making the same error in judgment or who share the same bias*, and so should reasonably be weighted less than a vote with the same numerical level of support but from a more diverse base of participants.[7]

We explore the latter idea mathematically in greater detail in the context of quadratic funding in the

[7] See https://twitter.com/VitalikButerin/status/1264948490834247681 and https://twitter.com/VitalikButerin/status/1265252184813420544 for informal Twitter poll evidence suggesting that people already find the idea of taking diversity into account in decision-making mechanisms intuitive.

Appendix, where we introduce a new primitive, called the **"correlation score**." This concept of correlation discounting could be extended to structure deliberative conversations. For example, DAOs susceptible to majoritarian capture could compute over SBTs to bring maximally diverse members together in conversation and ensure minority voices are heard.

DAOs could also rely on SBTs to deter forms of strategic behavior such as "vampire attacks." In such attacks, a DAO—typically with an associated DeFi protocol of economic value—free-rides off the R&D of another by copying their open-source code and subsequently luring users' liquidity with a token. DAOs could deter free-riders by first creating a norm around souldropping (perhaps vesting SBTs) only to probable sybil-resistant Souls who delivered liquidity and then withholding souldrops to Souls who shifted their liquidity in a vampire attack. The same mechanism wouldn't work with airdrops to wallets because a holder can spread liquidity across many wallets to obfuscate their liquidity trail.

DAOs could also use SBTs to make leadership and governance programmatically responsive to their communities. Leadership roles could dynamically shift as the composition of the community shifts—as reflected in the changing distribution of SBTs across member Souls. A subset of members could be elevated to potential officer roles based on their intersectionality and coverage across multiple communities within the DAO. Protocols that value community cohesion could use SBTs to keep intersectional Souls at the center. Alternatively, DAOs may opt for governance that elevates certain combinations of traits more than others, such as diversity among zip codes or participation among a subset of special hobby DAOs.

4.6 Measuring Decentralization through Pluralism

When analyzing real-world ecosystems, it is desirable to measure how decentralized the ecosystem actually is. To what extent is the ecosystem truly decentralized, and to what extent is the decentralization "fake" and the ecosystem de-facto dominated by one or a small set of coordinating entities?

Two popular decentralization metrics are the Nakamoto coefficient proposed by Balaji Srinivasan, which measures how many distinct entities need to be combined to gather 51% of some resource, and the Herfindahl-Hirschman index used to measure market concentration for antitrust purposes, calculated by summing the squares of the market shares of the market participants. These approaches, however, leave open key questions of what are the correct resources to measure, how to deal with partial coordination, and the gray areas in what constitutes a "distinct entity."

For example, nominally independent firms may have many major shareholders in common, have directors who are friends with each other, or be regulated by the same government. In the context of token protocols, measuring decentralization of token holdings by looking at on-chain wallets is wildly inaccurate because many people have multiple wallets, and some wallets (e.g., exchanges) represent many people. Moreover, *even if* addresses could be traced back to unique individuals, those individuals could be socially correlated groups prone to accidental coordination (at best) or intentional collusion (at worst). **A better**

way of measuring decentralization would capture social dependencies, weak affiliations, and strong solidarities.

Miners and mining pool operators that together make up 90% of Bitcoin's hashpower sitting together on a conference panel.

SBTs support a different way of measuring the level of decentralization (or pluralism) in a DAO, protocol, or network.

- As a first step, protocol could limit token voting to reasonably sybil-resistant (or SBT rich) Souls.
- As a second step, a protocol could examine the correlations between SBTs held by different Souls and discount votes by Souls (pooling them as only partially separate) if they share a large number of SBTs. (We explore the latter idea mathematically in greater detail in the context of quadratic funding in Appendix A, where we introduce a new primitive, called the "correlation score.")
- As a third step, to zoom out and get a sense of the decentralization across the network, one could measure the correlations between SBTs held by Souls among and across different layers of the network stack—measuring correlations in voting, token ownership, governance-related communication, and even control over computational resources.

SBTs allow us to begin to measure the decentralization of an interoperating and layered ecosystem that is very difficult to measure at all today. There is still a large, open question of what formulas would best capture what we want to measure and be least vulnerable to manipulation. There are also many questions about how to examine the relationships of SBTs—weighting some SBTs more than others, discounting nested SBTs, or also factoring in the composition of transferable tokens within Souls. However, with a rich ecosystem of Souls and SBTs, a much larger amount of data would be available to make these calculations and move towards meaningful decentralization.

4.7 Plural Property

DAOs often own—or organize around owning—assets, both in the virtual and physical worlds. So far web3's scope has largely been limited to a narrow class of property whose bundle of rights are *wholly transferable*: tokens, NFTs, artworks, first editions or rare manuscripts like the U.S. Constitution. But the emphasis on transferability has been to web3's detriment, making it incapable of representing and supporting some of the simplest and ubiquitous property contracts today, such as apartment leases. Property rights are defined in the Roman legal tradition as bundles of rights to use ("usus"), consume or destroy ("abusus"), and profit ("fructus"). Rarely are all these rights jointly vested in the same owner. Apartment leases, for example, confer limited rights of use ("usus") to the lessor, but not unfettered rights to destroy the apartment ("abusus"), sell it off ("fructus"), or even transfer use (subletting). Rights of real property (land) are typically encumbered by a range of restrictions on private use, grants of public rights of access, limits on rights of sale, and even rights of purchase by eminent domain. They are also typically encumbered with mortgages that transfer some financial value to lenders.

The future of property innovation is unlikely to build on wholly transferable private property so far imagined web3. Rather **innovation will hinge on the ability to decompose property rights to match features of existing property regimes, and code even richer elaborations**. Corporations and other organizational forms evolved precisely to reconfigure property rights in even more creative ways—for example, granting employees access to proprietary facilities ("usus"), but reserving for managers rights to change or damage assets ("abusus"), while paying shareholders most financial benefit ("fructus"). SBTs have the flexibility to represent and proliferate such nuanced property rights of both physical and virtual assets, while encouraging new experiments. Here are just a few use cases:

- **Permissioning access** to privately or publicly controlled resources (e.g., homes, cars, museums, parks, and virtual equivalents). Transferable NFTs fail to capture this use case well because often access rights are conditional and non-transferable: if I trust you to enter my backyard and use it as recreational space, that does not imply that I trust you to sub-license that permission to someone else.

- **Data Cooperatives** where SBTs grant data access to researchers, while instantiating members' rights to grant access (perhaps by quadratic vote) and bargain for economic rights to discoveries and intellectual property born out of research. We explore this further in

Section 4 on Plural Sensemaking.

- **Experiments with local currencies** with rules that make them more valuable to hold and spend by Souls who live in a particular region or are part of a particular community.

- **Experiments in participation** where SBTs create a continuous basis for less contextualized Souls (e.g., immigrants, adolescents) to gain influence within novel and broader networks. Such Souls would begin with narrow SBTs that pool them with their families or local communities. As their affiliations gradually diversify, they would gain broader SBTs that instantiate voting rights to influence broader networks—in the spirit of Danielle Allen's idea of polypolitanism—a process that currently is mediated by arbitrary age and residence cut-offs.

- **Experiments in market design**, such as Harberger taxation and SALSA (self-assessed licenses sold at auction), where holders of an asset post a self-assessed price at which anyone else can buy the asset from them, and must periodically pay a tax proportional to the self-assessed price to maintain control. SBTs could be used to create more nuanced versions of SALSA—for example, where rights of participation are approved by the community to minimize strategic behavior from within or outside the community.

- **Experiments in democratic mechanism design** such as quadratic voting. Holders of SBTs representing membership in a community could quadratically vote on parameters such as incentives and tax rates. Ultimately, "markets" and "politics" are not separate design spaces; SBTs can be a major part of a technological stack that enables the entire space *between* the two categories to be explored. Provision of public goods through quadratic funding is another such intersection.

Of course, there are dystopian scenarios to consider. Immigration systems could be permissioned with migratory SBTs. Regulatory capture could be codified in nested community tokens, where homeowners have a disproportionate voting power and stall housing construction. SBTs could automate red-lining. As we discuss further below, these scenarios should be considered within the context of the current opaque-top-down permissions and discriminations. SBTs make discrimination more transparent and therefore potentially contestable.

4.8 From Private and Public Goods to Plural Network Goods

More generally, SBTs could allow us to effectively represent and manage assets and goods that are anywhere on the **spectrum between being fully private and fully public**. In reality, almost everything is on the spectrum: even goods for personal consumption have positive spillovers, such as making the consumer better able to contribute to their family or community, and even the most globally available public goods (e.g. climate) are inevitably more useful to some people than others (e.g. Seychelles v. Siberia). Similarly, *human motivation* is rarely fully selfish or fully altruistic; there are many patterns of pre-existing cooperation, and some more present among certain communities over others.

Yet mechanism design today assumes atomized, selfish agents without *pre-existing* cooperation, often making mechanisms vulnerable to innocent over-coordination,[8] at best, and intentional collusion, at worst, by groups who are *already* cooperative. Thus, even the best public funding models, including Quadratic Funding (QF), *can't scale*. QF encourages coordination by offering diminishing rewards to concentrated action of the few, but increasing rewards to collective action of the many; for example, a total of $1 contributed equally by 10 people is matched by $99 to generate $100 in total, while $10 contributed by a single person receives no match. Mathematically, this is accomplished by matching funds proportional to the square of the sum of the square roots of individual contributions (as we further elaborate in the Appendix). But even *weak cooperation* (say donating $1 to a cause) among large groups (say most citizens of China) would dominate the system and absorb all its matching funds because the premium QF puts on the number of unique contributors. As is, QF doesn't discount coordination among correlated, special interests that may swamp a QF round, but instead *rewards* it.

But rather than treating pre-existing cooperation as a bug we ought to "write over," the key is to acknowledge it as reflecting partial cooperation that we should harness and compensate for. Afterall, we are in the business of encouraging cooperation. The trick is to make quadratic mechanisms work alongside pre-existing networks of cooperation, correcting for their biases and tendencies to over-coordinate. SBTs offer a natural way by allowing us to tip the scales in favor of cooperation *across differences*. As Nobel Laureate Elinor Ostrom famously highlighted, the problem is less coordinating public goods *per se* but rather one of helping communities made up of imperfectly cooperative but socially connected individuals *overcome* their social differences to coordinate at scale in broader networks.

If SBTs represent community memberships that reflect a Soul's partialities, **favoring cooperation across differences simply means discounting cooperative rewards to similarly affiliated or correlated Souls—similarity measured by their shared SBTs**. The assumption is that consensus *between the differently* affiliated better signals plural goods across *broader* networks, whereas consensus between the *similarly* affiliated more likely signals over-coordinated (or colluded) goods serving narrower interests.

By revealing shared memberships across Souls, SBTs allow us to *discount* pre-existing cooperation and quadratically scale up plural goods that confer benefits widely across emergent networks—**agreed upon by the most diverse members**—rather than more narrow goods innocently over-coordinated (or intentionally colluded) by special interests. The precise formula for correlation discounting "optimally" depends on model details and has not yet been studied, but we provide a first pass for experimentation for

[8] We say "innocent," because highly-cooperative groups naturally will seek to advance their interests, which may very well be for *their* collective benefit.

further research in the Appendix.

§5 PLURAL SENSEMAKING

An example of plural network goods that are of increasing salience in a digital world are predictive models built off user data. Both artificial intelligence (AI) and prediction markets seek to predict future events based on data primarily elicited from people. But both paradigms are limited in different and nearly opposite ways. The dominant paradigm in AI eschews incentives, instead hoovering up (public or privately surveilled) data feeds and synthesizing them into predictions through proprietary large-scale, non-linear models—harnessing the default web2 monopoly on "usus" without any "fructus" flowing to data laborers. Prediction markets take the opposite approach, where people bet on outcome in the hopes of financial gains, relying entirely on economic incentives of financial speculation ("fructus") without synthesizing the beliefs of bettors to produce composable models. At the same time, both of these paradigms yield conclusions that are characterized as "objective" truths; whereas AI models are portrayed as "universal" or "generally intelligent," prediction markets are portrayed as summarizing all the beliefs of the market participants in a single number: equilibrium price.

A more productive paradigm is to eschew these extremes, and instead draw on the virtues of both, while compensating for their weaknesses and enriching their breadth. We propose thoughtfully combining the complexity of non-linear AI models with the market incentives of prediction markets to transform passive data laborers into active data creators. With such provenance-rich information rooted in the sociality of data creators, we illustrate how DeSoc can unlock plural network(ed) intelligence more powerful than either approach.

5.1 Prediction Markets to Prediction Plurality

Prediction markets aim to aggregate beliefs based on wealth and risk preferences of those willing to bet—money talks. But this "survival of the fittest" isn't a desirable way to aggregate beliefs. A zero-sum game where one trader's gain is another's loss assumes a generalized ability at prediction that wrests with "the smart" and not "the dumb." While wealth may be a proxy for some forms of ability and expertise, predictions that account for other forms of relative expertise may be more reliable. Participants who have lost bets in a particular domain, may have more accurate beliefs in another domain. But prediction markets have the unfortunate effect of eliciting beliefs of those prone to gambling, which enriches those who win bets, impoverishes the rest, and discourages general participation of the risk-averse.

There are better ways to elicit beliefs. Research suggests that while prediction markets generally outperform simple polling, they don't outperform *sophisticated team prediction polling*, where people have incentives to share and discuss information. Under team deliberation models, members can be weighted based on factors like past performance and peer evaluation, and the team participates in semi-structured discussions to pool information that can't be encapsulated simply in a buy or sell contract. Such team deliberation models can be further improved with **quadratic rules to elicit exact probability estimates**

from all participants (compared to prediction markets, which only elicit up-down views about the current price equilibrium).[9] It has been demonstrated that the amount of contracts that people have an incentive to buy reflects their subjective probability assessment.[10] Such markets also distribute the gains from participation much more equally, rewarding accuracy without bankrupting the rest and thus keeping everyone as participants for future rounds.

SBTs could unlock a new class of rich models and experiments in predictive power and relative expertise. Whereas prediction markets elicit one number—the price of a contract—quadratic polling elicits each participant's *exact* belief about the probability of an event. SBTs enable further **computation over those beliefs *in social context*** of the education credentials, memberships, and general sociality of a participant to develop better weighted (or non-linearly synthesized) predictive models, likely surfacing expert predictors at novel, unforeseen intersections. So even if a poll did not aggregate beliefs well, polls could be studied retroactively to uncover the characteristics of "more correct" participants and convene better tailored "experts" in future polls, perhaps in a deliberative team context. These mechanisms are closely related to those we advocate throughout this paper. In the same way that quadratic mechanisms discounted by correlation scores can transform poorly coordinated top-down public goods into powerful, bottom-up plural network goods, they can also transform governance systems based on zero-sum prediction markets that incent participants to hide their information (e.g., Futarchy) into more positive-sum plural sense-making that can encourage revelation and synthesis of new and better information.

5.2 Artificial Intelligence to Plural Intelligence

Large scale non-linear "neural network" models (such as BERT and GPT-3) could also be transformed by SBTs. Such models hoover volumes of public or privately surveilled data feeds to produce rich models and predictions, such as code based on natural language prompts. Most surveilled data creators aren't aware of their role in creating these models, retain no residual rights, and are viewed as "incidental" rather than as key participants. Moreover, data hoovering divorces models from their social context, which masks their biases and limitations and undermines our ability to compensate for them. These tensions have increasingly come to the fore with growing demand for data availability, new initiatives like "data sheets for data sets" that document data provenance, and privacy-preserving approaches to machine learning. Such approaches require giving meaningful economic and governance stakes to those who generate the data and incenting them to cooperate in producing models more powerful than what they could build alone.

SBTs offer a natural way to program **economic incentives for provenance-rich data while**

[9] Under a quadratic rule, team members can buy a contract that pays out $X conditional on an event occurring, but costs $(X^2)/2. For example, an individual who sets X=0.5 will receive $0.5 if the event occurs—paid by the poller—and will pay $0.125 regardless.

[10] If an individual assesses probability p, their expected payoff is pX and cost is X^2/2. Taking the derivative with respect to X, the optimality condition is p=X, assuming risk neutrality, which is reasonable for small stakes (both the payoff and the cost may be arbitrarily scaled down or up and the same argument holds).

empowering data creators with residual governance rights over their data. In particular, SBTs allow carefully and proportionately targeted incentives for data (and data quality) at individuals and communities based on their characteristics. At the same time, model-makers can track the characteristics of the collected data and their social context—as reflected by SBTs—and find contributors that offset biases and compensate for limits. SBTs can also program bespoke governance rights to data creators, allowing them to form cooperatives that pool data and negotiate uses. This bottom-up programmability by data creators enables a future of plural intelligences, where model-makers can compete to negotiate uses over the same data to build different models. Thus, we move away from a paradigm of a detached monolithic "artificial intelligence" free from human origins, hoovering up provenance-free surveilled data to instead a **Cambrian explosion of cooperatively constructed plural intelligences rooted in social provenance and governed by Souls**.

Over time, just as SBTs individuate a Soul, they also come to individuate models—embedding data provenance, governance and economic rights directly into the model's code. Thus, plural intelligences—like humans—build a Soul embedded in human sociality. Or depending on how you look at it, humans evolve over time embedded in plural intelligences—each with a unique Soul, complementing and cooperating with other Souls. And, in this, we see the convergence of the prediction market and AI paradigms towards plural sense-making, combining widely distributed incentives and careful tracking of social context to create a diversity of models that combine the best of both approaches into a technology paradigm more powerful than either.

5.3 Programmable Plural Privacy

Plural intelligences raise important questions about data privacy. Afterall, to build such powerful intelligences requires pooling data across individuals from large data sets (e.g., health data), or capturing data that isn't interpersonal but shared (e.g., a social graph). "Self-sovereign identity" advocates tend to treat data as private property: data about this interaction is *mine* and so *I* should be able to choose when and to whom to reveal it. However, even more than in the physical economy, the data economy is poorly understood in terms of simple private property. In simple two-way relationships, such as an illicit affair, the right to reveal information is usually symmetrical, often requiring mutual-permission and consent. As scholar Helen Nissenbaum highlights, the concern is not "privacy" as such but lack of integrity to context in the sharing of information. The Cambridge Analytica scandal was largely about people revealing properties of their social graph and information about their friends, without their consent.

Rather than privacy-as-transferable-property-right, a more promising approach is to treat **privacy as a programmable, loosely coupled bundle of *rights to permission access, alter or profit from information*.** Under such a paradigm, every SBT—such as an SBT that represents a credential or access to a data store—would ideally also have an implied programmable property right *specifying access* to the underlying information constituting the SBT: the holders, the agreements between them, the shared property (e.g., data), and obligations to 3rd parties. For example, some issuers would choose to make SBTs wholly public. Some SBTs, such as a passport or health record, would be private in the self-sovereign sense, with unilateral rights to disclose by Souls who carry the SBT. Others, such as SBTs that reflect membership of a data cooperative, would have multi-signature or more sophisticated community voting permissions, where all or a qualified majority of SBT holders must consent to disclosure.

While there are current technical questions (can SBTs be programmed in such a way?) and important questions around incentive compatibility (explored further in Section 7)—we nonetheless think programmable plural privacy warrants further research and offers key advantages over alternative paradigms. Under our approach, SBTs have the potential to enable privacy as a programmable, composable right that can map upon the complex set of expectations and agreements we have today. Moreover, such programmability could help us reimagine new configurations, as there are an ***infinite*** **number of ways privacy—as a *right to permission access to information*—could be composed with "usus," "abusus," and "fructus"** to create a nuanced constellation of access rights. For example, SBTs could permission computations over data stores—perhaps owned and governed by a plurality of Souls—using a specific privacy preserving technique. Some SBTs may even permission access to data in a way where certain computations can be made, but the results cannot be *proven* to third parties. A simple example is a vote: the voting mechanism needs to tally votes from every Soul, but votes should not be provable to anyone else to prevent vote buying.

Communication is perhaps the most canonical form of shared data. Yet today's communication channels lack both user control and governance ("usus" and "abusus") and at the same time auction user *attention* ("fructus") to the highest bidder—even if a bot. SBTs have the potential to steward healthier forms of the "attention economy" that empowers Souls to spam-filter inbounds from likely bots outside of their social graph, while elevating communication from real communities and desired intersections. Listeners could become more aware of who they are listening to and better able to assign credit to works that spur insights. Rather than optimizing for maximum engagement, such an economy could optimize for positive-sum collaborations and valuable co-creations. Such communication channels also are important for security; as noted above, "high bandwidth" communication channels are critical to building the security foundations of community recovery.

§6 DECENTRALIZED SOCIETY

Web3 aspires to transform societies broadly, rather than merely financial systems. Yet today's social fabric—families, churches, teams, companies, civil society, celebrity, democracy—is meaningless in virtual worlds (often called the "metaverse") without primitives representing human souls and the broader relationships they support. If web3 eschews persistent identities, their patterns of trust and cooperation, and their composable rights and permissions, we see, respectively, sybil attacks, collusion, and a limited economic realm of wholly transferable private property—all of which trends towards hyper-financialization.

To skirt hyper-financialization—yet unlock exponential growth—we propose *augmenting and bridging* our sociality across virtual and physical realities, empowering souls and communities to encode rich social and economic relationships. But simply building on trust and cooperation is not enough. Correcting for biases and tendencies to over-coordinate (or collude) among trust networks is essential to encouraging more intricate, diverse relationships that span greater social distances than before. We call this **"Decentralized Society (DeSoc)": a co-determined sociality, where Souls and Communities convene bottom-up, as emergent properties of each other to produce *plural network goods* across different scales.**

We emphasize plural network goods as a feature of DeSoc, because networks are the most powerful engine of economic growth, yet the most susceptible to dystopian capture by private actors (e.g., web2) and powerful governments (e.g., Chinese Communist Party). Most significant economic growth results from *increasing network returns*, where every additional unit of input yields incrementally *more* output. Examples of simple physical networks include roads, electrical grids, cities, and other forms of infrastructure built off labor and other capital inputs. Examples of powerful digital networks include marketplaces, predictive models and plural intelligences built off data. In both cases, network economics diverges from neoclassical economics, which teaches *decreasing* returns—where every additional unit of input yields incrementally *less* output—and where private property yields the most efficient outcomes. Private property applied to an increasing returns context has the opposite effect—throttling network growth by rent extraction. A road between two cities can unlock increasing returns from gains from trade. But the same road privately owned can throttle growth if the owners choose to extract rent up to the value trading between the two cities. Public ownership over a network also has its own perils, being susceptible to regulatory capture or underfunding.

Networks with increasing returns are most efficient when treated neither as purely public nor purely private goods, but rather as *partial and plural shared goods*. DeSoc provides the social substrate to unbundle and reconfigure rights—rights of use ("usus"), rights to consume or destroy ("abusus"), and rights of profit ("fructus")—and enable efficient governance mechanisms across these rights that augment trust and cooperation while checking for collusion and capture. We've explored several mechanisms throughout this paper, such as community-based SALSA and quadratic funding (and voting) discounted by correlation scores. This third way of partial and plural ownership avoids the Charybdis of private rent extraction and Scylla of public regulatory capture.

In many ways, **DeFi today is a decreasing returns private property paradigm retrofitted onto increasing returns networks**. Built on the premise of *trustlessness*, DeFi is inherently limited to the realm of wholly transferable private property (e.g., transferable tokens) that mostly bundles "usus," "abusus," and "fructus." At best, DeFi risks throttling network growth by rent extraction and at worst risks ushering in dystopian surveillance monopolies dominated by "whales'' who harvest and hoover up data in a race-to-the-bottom—much like web2.

DeSoc transforms DeFi's race to control and speculate on the value of networks into a bottom-up coordination to build, participate, and govern them. At minimum, DeSoc's social substrate can make DeFi sybil-resistant (enabling community governance), vampire-resistant (internalizing positive externalities to build an open-source network), and collusion-resistant (preserving a network's decentralization). With DeSoc's structural corrections, DeFi can support and expand plural networks that confer benefits broadly—as agreed upon by the most diverse members—rather than further entrenching networks captured by narrow interests.

Yet, the **greatest strength of DeSoc is its *network composability*.** Sustained increasing returns and network growth isn't simply avoiding the perils of rent extraction, but also encouraging the proliferation and intersection of nested networks. A road may form a network between two cities. But cut off from broader cooperation, two cooperating cities will eventually hit a ceiling of diminishing returns—either because of *congestion* (roads and housing) or *exhaustion* (reaching the limits of the people they can serve). Only through technological innovation and growing broader, if looser, cooperation with neighboring networks for new sources of increasing returns can value continue to grow exponentially. Some cooperation will be physical, incrementally extending physical trade across space. But many more connections will be informational and digital. Over time, we will see new matrices of cooperation between physical and digital networks, reliant upon and extending the social interconnections they are built on. It is precisely this intersecting, partly nested structure of ever growing network cooperation across digital and physical worlds that DeSoc enables.

Through composing networks and coordination, DeSoc emerges at the intersection of politics and markets—augmenting both with sociality. DeSoc empowers the vision of JCR Licklider—founder of ARPANET that created the internet—of "man-computer symbiosis" in an "intergalactic computer network" with dramatically increased social dynamism *built on trust*. Rather than build on DeFi's *trustless* premise, DeSoc encodes trust networks that underpin the real economy today and enables us to harness them to generate plural network goods resilient to capture, extraction, or domination. With such augmented sociality, web3 can eschew short-term hyper-financialization in favor of an unbounded future of increasing returns across social distance.

6.1 Souls can go to Heaven...or Hell

While we have selectively highlighted the potential unlocked by DeSoc that we find promising, it is important to remember that almost any technology with such transformative potential will have a similar potential for destructive transformation: fire burns; the wheel steamrolls; the television brainwashes; cars pollute; credit cards trap in debt, and so on. Here, the same SBTs that could be used to compensate for in-group dynamics and achieve cooperation across differences could also be used to automate red-lining of disfavored social groups or even target them for cyber or physical attack, enforce restrictive migration policies, or make predatory loans. Many of these possibilities are less salient in the current web3 ecosystem because they aren't meaningful concepts given the current substrate. Enabling upsides of DeSoc also enables these harms. Just as the downside of having a heart is that a heart can be broken, the downside of having a Soul is it can go to hell and the downside of having a society is that societies are often animated by hatred, prejudice, violence and fear. Humanity is a great and often tragic experiment.

As we meditate on the possible dystopias of DeSoc, we should also contextualize these possibilities within other technological enabled dystopias. Web2 is architecture for opaque authoritarian surveillance and social control. Whereas web2 often relies on top-down artificial bureaucracies to confer identity (a "driver's license"), DeSoc relies on horizontal ("peer-to-peer") social attestations. Whereas DeSoc empowers Souls to encode their own relationships and co-create plural property, web2 intermediates social connections or monetizes them with opaque algorithms that can polarize, divide, and misinform. DeSoc sidesteps top-down, opaque social credit systems. Web2 forms the basis of them. DeSoc treats Souls as agents, whereas web2 treats Souls as objects.

The risk of social control with DeFi—without any identity substrate—is less, at least in the near-term. But DeFi has its own dystopia. **While DeFi overcomes *explicit* forms of centralization—where specific actors have an outsized level of formal power within a system—it has no built-in way to overcome *implicit* centralization through collusion and market power**. Monopolies don't always surface as the Standard Oils of the past. Collusion can even happen at higher and far-removed levels of an ecosystem. We see this today with the rise of a class of institutional asset managers (e.g., Vanguard, BlackRock, State Street, Fidelity, etc.) that are the largest shareholders of all the largest banks, airlines, car companies, and other major industries. Because such asset managers hold a stake across all rivals within an industry (i.e. a stake in every major airline), their incentive is to make the companies that they hold *look* like a competing industry but *act* like a monopolist that maximizes industry-wide profits and entrenchment at consumer and general-public expense.[11]

In DeFi too, the same "whales" and VCs accumulate larger shares across each level of the stack and across competitors within a stack, perhaps voting in token governance, or delegating it to the same class of delegates, who are also similarly correlated across the network. Without any social substrate for sybil-resistance and correlation discounts to force-function decentralization, we should also expect to see more monopolies funded by whales, as monopolists increasingly become the largest pool of available investment capital. As "the money class" and users diverge, we should expect to see (and already see) greater and greater levels of incentive misalignment and rent extraction. If DeFi applications that deal with private data emerge, we may well see similar dynamics, such as apps encourage bidding wars between multiple people who "own" data that is actually interpersonal (e.g., their social graph) to build monolithic private AIs that compete against humans, eschewing a future of competing plural AIs that augment humans.

Thus, DeSoc does not need to be perfect to pass the test of being acceptably non-dystopian; to be a paradigm worth exploring it merely needs to be better than the available alternatives. Whereas DeSoc has *possible* dystopian scenarios to guard against, web2 and existing DeFi are falling into patterns that are *inevitably* dystopian, concentrating power among an elite who decide social outcomes or own most of the

[11] See Posner, E. & Weyl, E. G., "Dismembering the Octopus," *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*, Princeton University Press, 2018.

wealth. The direction of web2 is deterministically authoritarian, accelerating the capacity of top-down surveillance and behavior manipulation. The direction of today's DeFi is nominally anarcho-capitalist, but is already falling into network effect and monopoly pressures that risk its medium-term path becoming authoritarian in much the same way.

DeSoc, in contrast, is **stochastic social pluralism**—a network of individuals and communities that come together, as emergent properties of each other, co-determining their own future. Looking at web2, the outgrowth of DeSoc can be analogized to the rise of popular participatory governments out of centuries of monarchy. Participatory governments didn't *inevitably* give rise to democracy; it also led to the rise of communism and fascism. Similarly, SBTs don't make digital infrastructure *inherently* democratic, but are democratic-compatible depending on what Souls and communities co-determine. Opening this possibility space is a marked improvement over web2's authoritarianism and DeFi's anarcho-capitalism.

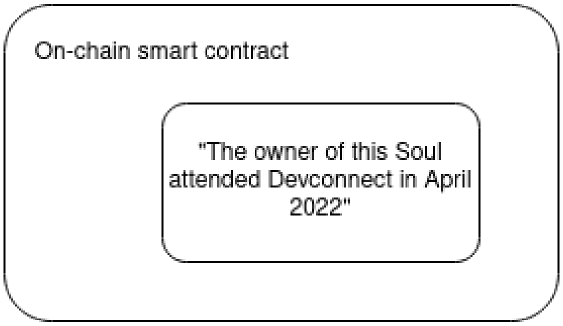
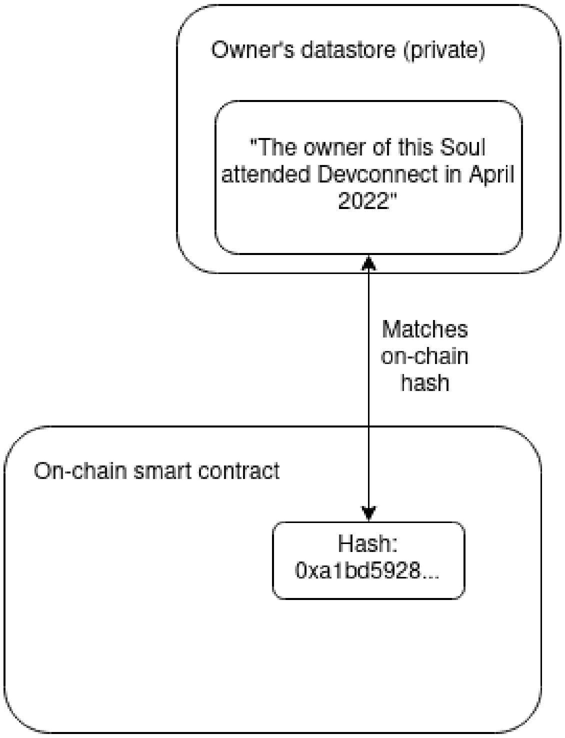
§7 IMPLEMENTATION CHALLENGES

Privacy presents a key challenge for DeSoc. On the one hand, too many public SBTs may reveal too much information about a Soul, making them vulnerable to social control. On the other hand, too many purely private SBTs may also lead to private communication channels that eschew correlation discounting for governance and social coordination—presenting important incentive compatibility questions. Closely related to the issue of privacy is the issue of cheating: Souls may misrepresent their social solidarities, while coordinating through private or side channels. We cannot aspire to know all the possibilities and answers, but instead explore the nature of the challenge here and sketch a few promising paths for future research.

7.1 Private Souls

Blockchain-based systems are public by default. Any relationship that is recorded on-chain is immediately visible not just to the participants, but also to anyone in the entire world. Some privacy can be retained by having multiple pseudonyms: a family Soul, a medical Soul, a professional Soul, a political Soul each carrying different SBTs. But done naively, it could be very easy to correlate these Souls to each other. The consequences of this lack of privacy are serious. Indeed, **without explicit measures taken to protect privacy, the "naive" vision of simply putting all SBTs on-chain may well make too much information public for many applications**.

To deal with over-publicity, there are a number of solutions with different levels of technical complexity and functionality. The simplest approach is that an SBT could store data off-chain, leaving only the hash of the data on-chain.

 On-chain smart contract "The owner of this Soul attended Devconnect in April 2022"	 Owner's datastore (private) "The owner of this Soul attended Devconnect in April 2022" Matches on-chain hash On-chain smart contract Hash: 0xa1bd5928...
On-chain data (fully public)	Off-chain data, on-chain hash link (Owner chooses when to reveal)

The choice of how to store the off-chain data is left to the person; possible solutions include (i) their own devices, (ii) a cloud service trusted by them, or (iii) decentralized networks such as the Interplanetary File System (IPFS). Storing data off-chain lets us continue to have smart contracts that permission the right to *write* SBT data, but at the same time have separate permissions to *read* that data. Bob can choose to reveal the contents of any of his SBTs (or the data stores which they permission) only when he wishes to. This already gets us quite far, and has the further benefit of improving technical scalability because most data only needs to be handled by a very small number of parties. But to fully achieve properties like plural privacy, as well as more fine-grained forms of disclosure, we need to go further. Fortunately, many cryptographic technologies let us do that.

One powerful set of building blocks that enables new ways to partially reveal data is a branch of cryptography called "zero knowledge proofs." While zero knowledge proofs are most frequently used today to enable privacy-preserving transfers of assets, they also can allow people to prove arbitrary statements without revealing any more information beyond the statement itself. For example, in a world where government documents and other attestations are cryptographically provable, someone could prove a statement like "I am a citizen of Canada, who is over 18 years old and has a university degree in economics and over 50,000 Twitter followers, and who has not yet claimed an account in this system."

Zero-knowledge proofs can be computed over SBTs to prove characteristics about a Soul (e.g., that it has certain memberships). This technique can be extended further by introducing **multi-party computation** techniques such as garbled circuits, which could make such tests *doubly private*: the prover does not reveal who they are to the verifier, and the verifier does not reveal their verification mechanism to the prover. Instead, both parties make the computation together and only learn the output.

Another powerful technique is ***designated-verifier proofs***. In general, "data" is slippery: if I send a movie to you, I cannot technologically prevent you from recording and sending it to a third party. Workarounds like Digital Rights Management (DRM) have at best limited effectiveness, and often come at great costs to users. Proofs, however, are not slippery in the same way. If Amma wants to prove some property X about her SBTs to Bob, she can make a zero knowledge proof of the statement "I hold SBTs that satisfy property X, OR I have the access key to Bob's Soul." Bob would find this statement convincing: he knows that he did not make the proof, and so Amma must actually have SBTs that satisfy property X. But if Bob passes the proof along to Cuifen, Cuifen would not be convinced: for all he knows, Bob could have made the proof with his own key. This can be made even stronger with ***verifiable delay functions (VDFs)***: Amma can make and present a proof that can only be made with the required SBTs right now, but anyone else will be able to make *five minutes from now*. **This means it is possible to represent sophisticated access permissions to *trustworthy proofs about data* despite the impossibility of making the same kinds of selective permissions to the raw data itself, which may simply be copy and pasted**. This may take us quite far nonetheless. Just as blockchains offer traceability in transactions that prevents someone from right-click copy-and-pasting a valuable NFT (and sybil attacking the original owner), similarly SBTs can offer traceability in social prevenance, which at minimum can reduce the value of copy-and-pasted data with unverified origins.

These off-chain data and zero-knowledge techniques are compatible with ***negative*** **reputation**—SBTs that are made visible even if the holder does not *want* them to be visible. Important examples of negative reputation include credit history, data about unpaid loans, negative reviews and complaints from business partners, and SBTs attesting to social connections relevant for coordination. Blockchains coupled with the same cryptography could offer a potential solution: Souls could be forced by smart contract logic to incorporate negative SBTs into a data structure like a Merkle tree that is stored off-chain, and any zero knowledge proof or garbled circuit computation would require them to introduce that information, because otherwise there would be a visible "hole" in the provided data that the verifier would recognize. The Unirep protocol is an example of how this might be implemented.

The point of these examples is not to show exactly how cryptographic technology can be used to solve all of the privacy and data permissioning problems with SBTs. Rather, it is to sketch out a few examples to show the power of such technologies. An important future research direction is to scope the exact limits of different kinds of data permissioning and the specific combinations of techniques that work best to achieve the desired level of permissions. Another question is what types of plural property regimes are desirable to govern data, and how to properly unbundle access ("usus"), editing ("abusus") and cash flow ("fructus") rights.

7.2 Cheating Souls

If SBTs are the social substrate upon which plural property, network goods and intelligences are coordinated, one might be concerned that Souls will try to trick or cheat their way into communities to gain access to governance or property rights that we imagine SBTs permissioning. For example, if many applications depend on SBTs representing conference attendance, unscrupulous conferences could offer such SBTs in exchange for bribes. With enough bribes, humans (and bots) could generate a fake social graph that makes the account look like an authentic human Soul, richly differentiated by (fake) SBTs. Just as DAOs can be bribed, so can Souls and the on-chain voting mechanisms which they use. Conversely, if SBTs are used to discount coordination, Souls may avoid SBTs to maximize their influence. Why should we believe that the SBTs a Soul possess accurately reflect their true social commitments rather than simply how they choose to play this game?

One argument is that the varying incentives to cheat may "balance out." Souls may sort and self-identify into the networks that are important to them at the right scale, much like how Harberger taxes balance out the incentive to over-value and under-value assets to elicit approximately accurate market valuations. Souls will want to hold more SBTs to gain influence within their communities, but on the other hand will eschew SBTs from communities they care less about to score lower on correlation metrics and increase their influence in governance over broader networks.

But it would be naive to assume that the two incentives—to gain access and maximize influence—always evenly cancel out, or even come close to canceling out, as though by magic. There may be many communities that use systems other than SBTs to gate access and governance. Or communities may—counter to our primary assumption about publicity—dole out private SBTs to reflect governance rights, but induce community members to keep these SBTs secret in broader decisions.

The problem of "gaming" should not be understated. It is a significant issue and resolving it is one of the most important foci for future research. Indeed, it is a major reason why open-sourcing many existing algorithms that prioritize or filter for human users is very challenging. To mitigate and deter SBT gaming, we suggest several norms and cryptographic directions:

1. The ecosystem of SBTs could **bootstrap off "thick" community channels**, where SBTs signal authentic off-chain community membership with strong social bonds and repeat interactions. This would make it easier for communities to filter and revoke SBTs of impersonators and bots. Such thick channels—which we often find in churches, workplaces, schools, meet-up groups, and organizations in civil society—would provide a more sybil-resistant social substrate to police gaming

(e.g., through bots, bribes, impersonation) in more "thin" social channels.

2. **Nested communities could require SBTs to force context on potential collusion vectors "just below" them**. For example, if a state were holding a funding round or vote, the state might require every participating citizen to also hold an SBT of a defined county and municipality.

3. The openness and cryptographic provability of the SBT ecosystem could itself be used to **actively detect collusive patterns and penalize inauthentic behavior**—perhaps discounting the voting power of collusive Souls, or obliging Souls to accept SBTs representing negative attestations. For example, if one Soul attests to the humanity of another Soul that turns out to be a bot, the case can be escalated and publicly verified, leading to that Soul having a large number of negative attestations. This already happens to an extent within the GitCoin QF ecosystem, where a range of signals are used to detect "collusive groups."

4. ZK technology (eg. MACI) could **cryptographically prevent some attestations made by a Soul from being provable**. This would make attempts to sell certain kinds of attestations non-credible, because the briber would have no way to tell whether or not the bribe recipient followed through on their side of the deal. There has been a large body of research on the use of such techniques for voting, but ultimately any non-financialized social mechanism may end up benefiting from similar ideas.

5. We could **encourage whistleblowers** as a way of making collusion of significant size unstable. Instead of detecting and penalizing incorrect or abusive *behavior*, we detect and penalize abusive *patterns of collusion*. This technique is risky to overuse because of the possibility of false-flag bribes, but it is nevertheless part of the toolkit.

6. We could use **mechanisms from peer-prediction theory** to encourage reporting to be honest in all cases except where collusion is extremely large. Instead of the conference attesting to attendees' attendance, attendees could attest to *each other's* attendance, so the number of participants that would need to be bribed to attest to a false claim becomes very large. The rewards need not be financial, but could be SBTs, making the rewards more useful to genuine community members than they are to attackers.

7. We could use correlation scores that **focus on correlations where there is a large incentive to be honest** if a group of Souls share a common interest. For example, the correlation scoring technique used in bounded pairwise quadrating funding uses quadratic funding donations themselves to determine how correlated two participants are, and therefore how much to discount their intersection. If two participants share many common interests, their incentive to express this fact to the QF mechanism is certainly diminished with correlation discounting, but it never becomes *zero* or *negative*.

§8 COMPARISONS AND LIMITATIONS

While the range of identity frameworks proposed is almost limitless, there are four particularly prominent and adjacent paradigms widely discussed in the web3 space that merit comparison: the dominant "legacy" identity ecosystem, the pseudonymous economy, proof of personhood, and verifiable credentials. Each paradigm highlights important contributions and challenges for future development of the social identity paradigm we advocate, and we use such limitations as a springboard for exploring future directions. All that considered, we also explain why we believe our social identity primitives of Souls and soulbound tokens are a more promising path forward for privacy regimes.

8.1 Legacy

Legacy identity systems rely on pieces of papers or identity cards issued and mediated by a 3rd party (a government, university, employer, etc). Provenance is established by calling up the 3rd party for a confirmation. While the legacy system has an interesting set of properties we should understand more deeply, such systems are wildly inefficient and do not lend themselves to composability or computation for rapid, efficient coordination. Moreover, these systems lack social context and makes Souls reliant on a centralized 3rd party to confirm membership to a community, rather than the embedding community. For example, most government issued IDs eventually trace back to a birth certificate issued on the authority of a medical doctor and family members, who are the ultimate source of truth and leave out many equally meaningful social connections that—taken together—offer far stronger validation. In fact, when centers of concentrated power seek strong identification (e.g., getting a security clearance from a major government) they rarely rely on such documents, instead turning to interviews in social networks. **Thus such legacy identity systems tend to concentrate power in the issuer and in those who can undertake the due diligence to get stronger verification, who in turn become calcified and unreliable bureaucracies**. A crucial design goal of DeSoc is ensuring that the security requirements of government IDs can be met and exceeded, allowing horizontal networks to make greater security available to all users and through a range of social substrates.

8.2 Pseudonymous Economy

The vision of a society based around combining reputation systems with zero knowledge proof mechanisms to preserve privacy has been most widely promoted by Balaji Srinivasan, who coined and popularized the phrase "pseudonymous economy." His early version emphasizes the use of pseudonyms to avoid discrimination and evade "cancel culture" by social mobs that seek to harm a person's reputation and break their social ties. It envisions people accumulating *transferable* zero-knowledge (ZK) attestations in their wallets and evading reputational attacks by transferring a subset of attestations to new wallets, or splitting the attestations amongst multiple wallets, presumably without traceability. In culling attestations to port, a person chooses the level of desired pseudonymity in the new account, weighing a tradeoff between more anonymity (porting fewer attestations) or more distribution to their social network (porting over more attestations).

The practical difference between typical pseudonymous economy proposals and DeSoc is that we deemphasize identity separation as a primary way to protect participants from abuses and cancel culture. Some level of separation (e.g., different Souls between family, work, politics, etc.) may be healthy, but in general there are great disadvantages to relying on the ability to spin up new identities as a primary crutch against attacks. It makes reputation-staking for lending and provenance harder, and it composes poorly with governance mechanisms that try to correct for correlations or Sybils.

Rather than protecting victims by allowing them to re-emerge from attacks with a new–if diminished—identity, DeSoc would allow other approaches, such as **contextualizing the attacker.** "Cancellation" often arises precisely because statements and actions are taken *out of context* and viral signals travel through uncontextualized networks, when a person or bot has little social connection or context to a victim. In the same way that SBTs provide provenance to protect against deep fakes, a map of SBTs socially graphs a "hit piece's" origin. "Hit pieces" essentially are artifacts arising outside of the victim's communities (as reflected by shared SBT memberships), or lacking SBT attestations from the victim's communities—which should cast doubt on the piece's veracity. SBTs also empower victims to launch a defensive response to counteract the hit, curated and propagated from *their* network of trust (represented here by the patterns of co-holding of SBTs). By maintaining social context, people can maintain trust, even if they are under threat of cancellation, and hold attackers accountable. Improving provenance improves the social foundation of truth.

8.3 Proof of personhood (PoP)

Proof of Personhood protocols (PoP) aim to provide tokens of individual uniqueness, to prevent Sybil attacks and allow non-financialized applications. To do so, they rely on approaches such as global analysis of social graphs, biometrics, simultaneous global key parties, or some combination thereof. However, because PoP protocols seek to represent *individual* identities—-focused on achieving *global uniqueness*—rather than *social identities* mapping relationships and solidarities, **PoP protocols are limited to applications that treat all humans the same**. Most applications we are interested in—such as staking reputation—are relational and move beyond being a unique human to being a *differentiated* human.

Moreover, PoP protocols are not immune to sybil attacks. In almost all near-term foreseeable applications, PoP systems are effectively open to Sybil attacks, just at a slightly higher cost. Unless most people on the planet are registered for a PoP service and are participating in a particular validation exercise, an attacker can always recruit disinterested humans who are not yet participating to act as Sybils. While such mercenaries are not quite bots, the difference is superficial other than perhaps a small added expense.

Many PoP protocols aim to build a substrate for universal basic income or global democracy. While we don't share the same ambition, such protocols have spurred us to nonetheless consider how to build gradually towards coordinating plural network goods. In contrast to the binary, individualist and global nature of PoP, our approach aims to construct a rich, contextual and layered substrate for bottom-up reputation, property and governance that allows participation in a range of communities and networks, small and large.

8.4 Verifiable credentials

Verifiable credentials (VCs) are a W3C standard where credentials (or attestations) are zk-shareable at the holder's discretion. VCs highlight the major limitations of our baseline privacy paradigm and motivate our discussion of privacy extensions above. Until SBTs have privacy extensions that narrow publicity, VCs and SBTs can be seen as natural complements: in particular, SBTs are initially public making them inappropriate for sensitive information like government-issued identification, while VC implementations have struggled with a recovery paradigm that could be addressed by community recovery. The two approaches combined can in the near-term be stronger than either alone. But VCs also have a key limitation: at least in their standardized form, VCs do not support most of the applications we have enumerated because of their *unilateral* privacy.

Unilateral zk-sharing isn't incentive-compatible with our use cases, nor does it reflect our norms around privacy. Most of our applications depend on some level of publicity. But under zk-sharing, Souls can't know another Soul possesses an SBT unless it is shared to them—making reputation-staking, credible commitments, sybil-resistant governance, and simple rental contracts (e.g., apartment lease) impossible to get off the ground as other commitments and encumbrances are not necessarily visible. More deeply, we are skeptical that unilateral shareability is usually the right privacy paradigm. Rarely does one party in a multi-party relationship have the unilateral rights to disclose the relationship without the consent of the other. Just as unilaterally transferable private property is not a rich property regime, simplistic unilateral shareability is not a very rich privacy regime. If two parties co-own an asset and choose to represent their relationship through a VC, such credential doesn't allow for the mutual-consent and mutual-permissions. This problem travels to more complex cases of plural property and complex organizational forms and permissions, which are a feature of DeSoc.

§9 SOUL BIRTH

The path from the current web3 ecosystem to augmented sociality mediated by SBTs faces a classic cold start challenge. On the one hand, SBTs are not transferable. On the other hand, today's mix of wallets may not be the final home for SBTs because they lack community recovery mechanisms. But in order for community recovery wallets to work, they need a rich variety of SBTs across discrete communities to be secure. **What comes first: SBTs or community recovery?** Who are the early adopter communities? How do SBTs on different chains interoperate? We cannot aspire to know all the possibilities and answers, but instead sketch a few promising paths for the reader to further explore within the current web3 and even web2 architecture.

9.1 Proto SBTs

Although the hallmark of SBTs is non-transferability, SBTs may also have another property which may prove more useful in bootstrapping: *revocability*. It's possible that SBTs first gestate as revocable, transferable tokens, before growing into non-transferability. A token is revocable if an issuer can burn the token and re-issue it to a new wallet. Burning and re-issuing would make sense when, for example, keys are lost or compromised, and the issuer has an interest in ensuring the tokens are not financialized and sold off to a party—in other words, when the token signals authentic community membership. Employers, churches, meet-up groups, clubs with repeat off-chain interactions are well positioned to burn and re-issue tokens because they have a relationship with a person, and can easily check for impersonation by phone call, video-conference, or simple meeting in person. Single interactions, such as attendance to a concert or conference are poorly suited because community bonds are weaker.

Revocable, transferable tokens are a kind of proto-SBT—serving supportive, placental functions before Soul birth. These tokens buy time both for wallets to gestate secure, community recovery mechanisms and for a person to sufficiently accumulate proto-SBTs that can eventually be burned and re-issued into non-transferable SBTs. Under this pathway, the question is not, "what happens first: SBTs or community recovery?" Rather, SBTs and community recovery instantiate simultaneously, birthing a Soul.

9.2 Community Recovery Wallets

Although today's wallets lack community recovery, they each have relative strengths and weaknesses in being homes—or perhaps gestational wombs—for SBTs. Proof of Personhood (PoP) protocols have the advantage of already experimenting with social dispute resolution mechanisms, which are the foundation of community recovery. Also, many DAOs use PoPs to facilitate governance, making them natural first issuers of SBTs. However, despite PoPs natural lead, PoP protocols haven't yet earned broad trust to house valuable token assets, whereas custodial wallets have.

Custodial wallets—despite their flaws of centralization—may thus offer a natural onramp for less sophisticated retail users. Such custodial wallets could also build tooling for retail communities to issue revocable tokens that later convert (or burn and reissue) into SBTs or even tooling for more "corporate" issuers—many of whom are looking for ways to build loyal customer bases in web3 but lack expertise in custody. Once community recovery mechanisms have been formalized and battle-tested, these custodial wallets could decentralize into community recovery, while custodians move on to providing other valuable services in DeSoc (like community management, SBTs issuances, etc.)

For more sophisticated web3 users, decentralized non-custodial wallets (or non-custodial social recovery wallets like Argent and Loopring) are a natural starting point for bootstrapping community recovery mechanisms. Non-custodial wallets have the advantage of being native web3 open-source, and the flexibility to pre-announce and experiment with mechanisms incrementally to a subset of voluntary, sophisticated users to battletest incentives and mix mechanisms (e.g.,mult-sig). All of these approaches—PoPs, custodial, and non-custodial—play an important role in experimenting and onboarding users with different degrees of sophistication and risk tolerance.

9.3 Proto-Souls

Norms can also shepherd Souls into existence. As we rethink tokens and wallets, we can also reframe how we think about certain classes of NFTs and tokens that are intended to signal membership. In particular, we can introduce a norm of not transferring NFTs and POAPs issued by reputable institutions that reflect attendance to a conference, work experience, or education credentials. Such transfers of membership tokens—if traded for value—could diminish the reputation of a wallet and perhaps discourage issuers from further issuing membership or POAP tokens to that wallet. Already in the non-custodial ecosystem, a significant number of users have achieved significant financial reputation and stake in their wallets, which could bootstrap as effective collateral for them not to abuse non-transferability expectations.

While all these pathways have respective challenges, we hope that the variety of approaches increases the chance of convergence to our quasi-equilibrium state in the medium term through a small set of steps.

§10 CONCLUSION

As ambitious as we have been in imagining what DeSoc could enable, in many ways the above are just first steps. There is more than one road to DeSoc, including a number of non-blockchain based frameworks, such as Spritely, ACDC and Backchannel that rely on data stores tied to local machines rather than global ledgers. These frameworks may eventually offer even greater trust across social distance, because they can harness transitivity of trust relationships—like trusted introductions—rather than relying on SBTs issued by well-known, high-status institutions (like universities or DAOs). Furthermore the applications we describe above are just the beginning of what DeSoc can empower, not touching virtual worlds: their physics, society, and their complex intersection with the physical world. All this suggests that even the broad ambitions we paint above are just the beginning of what DeSoc may eventually become.

On that path, however, many challenges and open questions remain. The above sketches require extensive red teaming and many of them are more suggestive than fully prescriptive. How can DAOs maintain their publicity of state while thoughtfully comparing patterns of Souls and correlations in SBTs to enforce Sybil protections and decentralization? How incentive compatible is acquiring SBTs in face of various schemes of correlation discounting? How much does privacy conflict with correlation discounting and other DeSoc mechanism designs? How can we measure inequality in a social and yet appropriately private (contextually integral) manner? How should inheritance work in the community recovery framework? Are there red lines that can be drawn or even baked into protocols to avoid dystopian scenarios? Or should we simply race to build the best scenarios first? These questions are just the beginning of what we expect to be a research agenda spanning years that will co-evolve with the DeSoc ecosystem.

Yet the potential that DeSoc offers seems not just worth the price of navigating these tricky challenges, but perhaps necessary to ensure our survival. Albert Einstein told the 1932 disarmament conference that the failures of the "organizing power of man" to keep pace with "his technical advances" had put a "razor in the hands of a 3-year-old child." In a world where his observation seems more prescient than ever, learning how to program futures that encode *sociality*—rather than writing over trust—seems a required course for human life on this planet to persist.

APPENDIX

Adjusting Quadratic Mechanisms for Pre-Existing Cooperation

Because quadratic mechanisms incent collaboration from a baseline of selfishness, they are vulnerable to groups who are *already* cooperative. If SBTs reflect community memberships that individuate a Soul to reflect their partialities, SBTs can help us discount pre-existing cooperation and tip the scales in favor of cooperation *across differences*. Here we provide an illustration of a first attempt at a refined quadratic model and offer future directions for research. This mechanism is not optimized and doubtless has vulnerabilities; it is meant as an illustrative example to spur experimentation and future research. While we illustrate with Quadratic Funding (QF), the same principles and formulas also apply to Quadratic Voting (where individual contributions are simply substituted with voice credits).

In QF, a community matches individual contributions to shared projects with funds in proportion to the square of the sum of the square roots of individual contributions. For fixed contribution levels, matching funds grow as the square of the number of individual contributors, but have diminishing returns to individual contributions. There are diminishing returns to concentrated individual action, but increasing returns to collective action. For example, if Abdu, Shou and Belle were non-cooperating individuals—contributing respectively A, S and B currency units—the matching funds to their donations in a QF program (for example, Gitcoin Grants) should be proportional (with scaling determined by available funds) to the square of the sum of the square roots of individual donations.

$$Simple\ Match \sim \left(\sqrt{A} + \sqrt{S} + \sqrt{B}\right)^2 - (A + S + B)$$

Single Membership

Now suppose a simplified model where Abdu, Shou and Belle are differentiated by a single membership—workplace—and matching funds are available for startups, companies, and open-source projects (again, in the spirit of Gitcoin). Because people from the same workplace have a strong incentive to contribute to their own workplace to maximize matching funds to their company, we should expect them to coordinate. An extreme approach would be to assume that workers fully share goals and fully coordinate their behavior. Yet even in this simple case, there are several ways we might compensate in the formula.

A simple approach, which we call "clustering," would put two co-workers "under the same square root" in the quadratic formula to offset their tendency to already coordinate. If Abdu and Shou were co-workers (but not Belle), Abdu and Shou's contribution would be summed and square rooted together while Bob's contribution would be square rooted alone, effectively giving his contribution more weight:

$$Cluster\ Match \sim \left(\sqrt{A\ +\ S} + \sqrt{B}\right)^2 -\ A\ -\ S\ -\ B$$

If Abdu and Shou are perfectly coordinated, it's always optimal for them to split their join contribution equally, so we can assume $A\ =\ S$, letting us simplify:

$$= \left(\sqrt{2A} + \sqrt{B}\right)^2 -\ (2A\ +\ B)$$

In this case, it is easy to see how clustering leads to optimality (or welfare maximization) by the same argument as for QF more generally: if Abdu and Shou are perfectly coordinated, they effectively act as a single agent and the Clustering Matching formula is the QF formula for two agents—the joint Abdu-Shou agent and the Belle agent.

Another adjustment that also achieves optimality is what we call the "Offsetting Match:"

$$Offset\ Match \sim \left(\frac{\sqrt{A}+\sqrt{S}}{\sqrt{2}} + \sqrt{B}\right)^2 -\ A\ -\ S\ -\ B$$

The rationale in the Offsetting Match is that because Abdu and Shou are part of a perfectly coordinating size-2 group, we can reduce the weight of their votes by a factor of $\sqrt{2}$ to compensate for the coordination. This leads to the same outcome as the Cluster Match as it is always optimal for a perfectly coordinated Abdu and Shou ($A\ =\ S$) to make equal contributions and in this case

$$Offset\ Match \sim \left(\frac{\sqrt{A}+\sqrt{A}}{\sqrt{2}} + \sqrt{B}\right)^2 -\ A\ -\ A\ -\ B$$
$$= \left(\sqrt{2A} + \sqrt{B}\right)^2 -\ (2A\ +\ B)$$

Multiple Memberships

The previous example assumes Abdu, Shou and Belle have a single membership: workplace. Yet in almost all applications this would be a vast oversimplification. People have multiple community memberships, cooperative relationships, and even informal intersections. Abdu and Belle might be extended family, Shou and Belle might have attended the same school, or Shou and Abdu might be token-holders of the same layer 1 protocol, and so on. To facilitate cooperation across differences, these correlations in memberships between individuals need to be recognized in a less binary manner. We now consider extending each of the approaches above to do this. We again focus on the simplest example sufficient to make the point; below we follow up with more general formulae.

We focus on an example where Abdu and Shou share an affiliation, Abdu and Belle share a different affiliation, and Shou has an affiliation with a group that includes other members, but none participating in this matching round. This is the complete set of affiliations.

To extend the Cluster Match to this case, we include a cluster for each group of shared affiliations and distribute the contributions of each individual among all of the groups they participate in equally with coefficients on their contributions that sum to one.

$$Cluster\ Match \sim \left(\sqrt{\frac{S}{2} + \frac{A}{2}} + \sqrt{\frac{A}{2} + B} + \sqrt{\frac{S}{2}}\right)^2 - A - B - C$$

To extend the Offset Match, we have to solve for coefficients on each individual's contribution to compensate for the coordination benefiting that individual. In particular, if we assume that Belle half internalizes Abdu's value, that Abdu half internalizes Belle and a quarter internalizes Shou's and Shou quarter internalizes Abdu's, then we need to find coefficients solving

$$\alpha_A + \frac{\alpha_B}{2} + \frac{\alpha_S}{4} = 1$$

$$\alpha_B + \frac{\alpha_A}{2} = 1$$

$$\alpha_S + \frac{\alpha_A}{4} = 1$$

The solution to this equation is $\alpha_A = \frac{4}{11}$, $\alpha_B = \frac{9}{11}$, $\alpha_S = \frac{10}{11}$. So

$$Offset\ Match \sim \left(\sqrt{\frac{4A}{11}} + \sqrt{\frac{9B}{11}} + \sqrt{\frac{10S}{11}}\right)^2 - A - B - C$$

The Offset Match, while in some ways the simplest, is almost the most opaque, assigning to each individual a weight depending on their social centrality that offsets the power this grants.

General Formulae

For each individual $i = 1, \ldots N$, let us define the number of affiliations she holds as T_i; in general we might give different weights to different affiliations, but at present we assume they are all equal. Let Σ be the set of all "affiliation groups," projects of the set of holders of a given affiliation onto the set of participants in the match, with typical element σ_j. Note that $T_i = \sum_{j=1}^{|\Sigma|} 1_{i\in\sigma_j}$, where 1 is the indicator function. Denote the contribution of individual i as c_i. Then the general formula for the Cluster Match is

$$Cluster\ Match \sim \left(\sum_{j=1}^{|\Sigma|}\sqrt{\sum_{i=1}^{|\sigma_j|}\frac{c_i}{T_i}}\right)^2 - \sum_{i=1}^{N} c_i$$

Define the Correlation Score between any ordered pair of individuals i and k to be

$$s_{i,k} = \frac{\sum_{j=1}^{|\Sigma|} 1_{i\in\sigma_j} 1_{k\in\sigma_j}}{T_i}$$

The Offset Match is then derived by the offset coefficients, α_i that solve the system of equations, one for each individual i:

$$\alpha_i + \sum_{k\neq i}^{N} \alpha_k s_{k,i} = 1$$

This will generically yield a unique solution for the vector α, which is roughly an inverse measure of the network centrality of individuals in the solidarity network. Then

$$Offset\ Match \sim \left(\sum_{i=1}^{N}\sqrt{\alpha_i c_i}\right)^2 - \sum_{i=1}^{N} c_i$$

One appealing feature of this solution is that it will generally lead to optimality assuming that solidarity correctly measures effective internalization of utility. A less appealing feature is that it seems unlikely to be particularly "robust:" in particular and in contrast to other cases, it will not always be optimal for any individual to give all her contributions through the match rather than externally, given the penalties.

Pairwise Matching

A third mechanism, which we call "Pairwise Matching," suggested by Buterin (2019) takes a different approach. Pairwise Matching has the disadvantage that it does not achieve optimality but instead focuses on bounding losses from specific attacks, but it has the important advantage that it does not require an extrinsic source to specify who is coordinating and who is not; instead, this information is extracted from the contribution values themselves.

Pairwise Matching can only be meaningfully defined in the context of multiple projects and a per-pair matching cap, M. For every pair of agents (A, B), if they contribute $x_{A\to P}$ and $x_{B\to P}$ to the same project P, they get a subsidy[12]

$$Match_{AB\to P} = \frac{2M\sqrt{x_{A\to P}x_{B\to P}}}{M + CorrelationScore_{AB}}$$

Where M is a parameter of the system and

$$CorrelationScore_{AB} = \sum_{all\ projects\ P} \sqrt{x_{A\to P}x_{B\to P}}$$

The $CorrelationScore$ is intended to reflect to what extent two participants contribute to the same projects. If two participants A and B both contribute x to some project, then $CorrelationScore_{AB}$ increases by x. If they contribute different amounts, $CorrelationScore_{AB}$ increases by the geometric mean of their two contributions.

If A and B have a low $CorrelationScore$, we assume that they are highly independent agents, and give them close to the maximum subsidy whenever they do contribute to some project together. But if A and B contribute to the same project frequently and/or in large amounts, we assume that they are highly coordinated and are acting somewhat more like a single agent, and discount the subsidies to projects that they co-fund.

In the limiting case where $x_{A\to P} \to 0$ for all agents and projects, the correlation scores are negligible, and so the above formula is equivalent to simple quadratic funding: $Match_{AB\to P}$ simplifies to $2\sqrt{x_{A\to P}x_{B\to P}}$. In the three-agent case, where three agents contribute A, S and B, this simplifies to:

$$\begin{aligned} TotalMatch_P &= Match_{AS\to P} + Match_{BS\to P} + Match_{AB\to P} \\ &= 2\sqrt{AS} + 2\sqrt{BS} + 2\sqrt{AB} \\ &= \left(\sqrt{A} + \sqrt{S} + \sqrt{B}\right)^2 - (A + S + B) \end{aligned}$$

But if a pair of agents contributes many times or in large amounts to the same projects, the correlation score of that pair increases, until eventually any additional shared contributions to a new project are mostly taking subsidies away from other shared contributions that the same pair of agents has already made. As total matches approach infinity, the total subsidy per pair of agents approaches $lim_{T\to\infty}\frac{2MT}{M+T} = 2M.$

[12] The original description differs slightly in that it uses M instead of 2M. Technically, 2M is correct if we sum over *unordered pairs* of agents, and M is correct if we sum over *ordered pairs*. Here, we are summing over unordered pairs.

A key design goal of this formula was to bound the losses from incorrectly identifying a colluding group as independent agents. In Simple Matching, losses are unbounded: N fake or colluding agents controlled by the same real-world actor can each contribute V to a fake project, and extract a subsidy of $V * (N^2 - N)$. In Cluster Matching, a similar unbounded extraction is possible if the clustering mechanism misindentifies even one colluding group as being completely independent. In Pairwise Matching, in contrast, losses from N fake or colluding agents are always bounded above by $M * (N^2 - N)$, where M is a parameter of the system.

Note that Pairwise Matching does *not* achieve optimality: colluding actors still have the incentive to somewhat over-report how much they value certain projects, and can even extract some funds by contributing to a fake project controlled by themselves. Rather, this approach is intended to be a second-best, optimized for the case where limited outside information is available about which actors are actually colluding.

That said, Pairwise Matching can be used as a *philosophical template* for how to account for pre-existing coordination without over-penalizing it: instead of the correlation score only including $\sqrt{x_{A\rightarrow P} x_{B\rightarrow P}}$ values for that particular quadratic funding system, it could attempt to include similar terms for all instances where those two actors gained a benefit by cooperating. If benefits from cooperation are valued correctly, cooperating further would never be net-harmful for any pair of agents; rather, the net gains from further cooperation would simply approach zero.