

metaself.ai: the best possible version of us

Axel Noack

Abstract: As an individual and as a group, we are the product of the decisions we take. This is how we realize our potential. Whilst it has helped human development to use simplifications and shortcuts, metaself envisions a future where powerful technologies enable us to make better decisions. But unless these new powers are democratized and organized in an accountable way, the individual will be at the mercy of powerful third parties and technologies that can control and/or exploit him. Metaself proposes a comprehensive system of organization and cooperation to achieve true individual and collective growth. It starts with a robust infrastructure to preserve the digital identity of every individual and then connects these individuals in a true peer-to-peer network. The resultant digital autonomous organization will on the one hand be highly dynamic, but also legally anchored in the real world as a cooperative with the purpose to distribute all value to its members and not to outside shareholders.

1. Motivation

As we are about to enter a future where multiple innovations will transform our world beyond recognition, it is important to be clear about the role of the individual. He/she/it should be at the centre of all progress to come, technological, economic, political and otherwise, and not at the mercy of powerful third parties and/or technologies.¹

This paper proposes to build a comprehensive system of organization and cooperation that combines various technological and conceptual innovations to provide a platform for true individual and collective growth.

It is based on a robust infrastructure of individual data ownership that enables every person to achieve true self-determination in the digital sphere and, by extension, in the tangible world. Other than today, each individual will own and fully control all data related to his life at any time. He will also own all the value that is related to or can be extracted from his data and his existence at any time.

¹ In plain German: „Ist der Mensch selbstbestimmt oder fremdbestimmt?“

Each fully sovereign individual will be free to connect to other individuals, physical groups, machines or digital entities in the pursuit of common aims in the public or private sphere. This connection can only be established in mutual agreement and can be terminated at any time. The result will be a highly dynamic form of organization that serves the best interests of each member and each group at any moment. Other than today, there will be no intermediaries of any sort that can exploit and/or manipulate these relationships.

The resultant decentralized autonomous organization will facilitate a multitude of individual and collective interactions and decisions at any given moment. Other than today, when decisions are taken with either too much or too little information, under pressure of time and with multiple internal and external influences, advanced technologies will be used to make better decisions. Other than today, when many decisions are controlled, manipulated and/or exploited by third parties, all decisions will only benefit the parties involved.

The exact way in which the individual will interact with metasef will be the result of collective development. As of now, it can be imagined as a personal assistant to which the individual is constantly connected and with which he permanently interacts (if he choses to). One way to envision it, is as the voice in our head when we are thinking. This might be a strange idea right now, but many developments go in this direction. But other than today, when the technology is provided for by private companies or states, metasef will be a collective that builds and runs its own technology.

The value that is created by each individual and each group of individuals will be accounted for instantaneously, comprehensively and in precise detail, attributed to each individual, and distributed to them in instantaneous micro transactions. Other than today, when profits are extracted from the individual and transferred to external shareholders, metasef takes the form of a legal cooperative that creates value for its participants instead of extracting profit for outside shareholders.

The resultant collective is a deliberate and comprehensive departure from established ways of thinking and acting. As each individual is empowered to create the best possible version of himself, the resultant collective will also be the best possible version of ourselves. But the accountability will not end with the human actors involved, and also include the natural environment on which our survival depends and various machines in the quickly developing digital sphere.

The consequences of these changes will be very far-reaching and probably at some stage alter the current narrative which we construct around the individual self as an actor and the scope of his agency. How this new state of awareness will look and feel like is completely unpredictable, but so are the

challenges that await us. The assumption is that we will be better prepared to deal with these challenges if the individual is part of a new and bigger unity with a new level of consciousness than on his own.

This paper is conceptual in very broad terms and can not predict a future. But it suggests that we can start building it with technology that is available today. In the following, Chapter 2 takes the perspective of the individual human actor and examines how core concepts such as privacy, agency, and self-determination can be safeguarded in the digital world, which technical solutions are available, and how the digital self will be positioned. Chapter 3 looks at the implications for various groups and collectives, examine ways how individual human actors can organize and cooperate in novel ways, and describe how this construct can operate in the real world. Chapter 4 suggests that on the basis of such an architecture, the capabilities of individual and collective actors can be enhanced significantly, and that something new can arise.

2. Data, identity and self-determination

At the heart of all progress to come are vast amounts of data and powerful technologies to work with them. But it is still widely accepted that this data is owned, controlled, and exploited by third parties to generate profits for their shareholders. We propose a system where every person will own, fully control, and directly profit from all data connected to his or her life at any moment and in every situation.

This is a conscious and comprehensive departure from the current system, which has evolved over the past 20 years because the individual is simply overwhelmed by the sheer amount and the complexity of his data and can not (yet) fully grasp its power and value. Instead, the vast majority of people still enter into tacit agreements where they give up control over their data in exchange for free services and a comfortable life.² But in the last resort, they become transparent, and in giving up control over their digital identity, they ultimately give up control over their lives.

This chapter first discusses some basic values that we deem important when designing a data infrastructure. It then shows which data are referred to, why and how this data will be stored on a distributed ledger, and what the far-reaching implications are, if this is done consistently.

² The detailed account by Shoshana Zuboff (2019) has so far not lead to any significant change in attitude or behaviour.

2.1. Basic values

In most Western discussions since Aristotle it has been considered obvious that there should be a clear separation between the private and the public sphere, and that the protection of this separation in the last resort represents human freedom. As John Stuart Mill put it later: „That there is, or ought to be, some space in human existence thus entrenched around, and sacred from authoritative intrusion, no one who professes the smallest regard to human freedom or dignity will call in question“ (Mill 1848).

The protection of this freedom is therefore enshrined in many constitutions, including the freedom to exchange information with others (e.g. Article 10, Grundgesetz für die Bundesrepublik Deutschland). Privacy, confidentiality, and also anonymity are related concepts that pertain to the protection of personal information and the control over its disclosure. Whilst these concepts have been discussed for a long time, they received new urgency in the digital world. Before we continue, it is therefore important to agree on some basic definitions.³

Privacy refers to an individual's right and ability to control the access, disclosure and use of their personal information. It encompasses the ability to keep certain aspects of one's life or identity hidden from others. Privacy ultimately allows individuals to make personal choices without external interference, and thus develop and express their identity freely.

Confidentiality pertains to the protection of sensitive or private information that has been shared with or entrusted to a third party. It involves a duty to keep the information confidential and to prevent its unauthorized disclosure to others. Confidentiality therefore entails an obligation to maintain secrecy.

Anonymity involves the state of not being identified. It means intentionally concealing or not revealing one's identity or personal details. An anonymous person cannot be easily connected or linked to their actions, statements, or data. Anonymity provides individuals with protection against identification or surveillance, allowing them to interact or participate without disclosing their true identity.

To be very clear: anonymity in itself is not the aim of metaself. On the contrary: we believe that anonymity, especially in the digital space, has led to a lot of confusion and harm. Metaself wants to preserve privacy and confidentiality as much as possible and as much as necessary. This does, however, not preclude the freedom to take on various artificial identities, such

³ Unless referenced, definitions in this text have been partly formulated using ChatGPT in an iterative process („as a sparring partner“). The final formulations used always represent the author's opinion.

as avatars, in social contexts where this is permitted, desired, or intended. But we want to build a system, where a human actor can be identified in the last resort, when this is necessary or when he decides to do so.

Agency refers to an individual's capacity to act and make choices independently, based on their own beliefs, desires, and values. It is the ability to have control over one's actions and decisions, free from undue influence or coercion. Privacy plays a crucial role in supporting agency because it provides individuals with the necessary space and freedom to exercise their autonomy and make choices without external interference.

Self-determination is the right and ability to determine one's own destiny, make decisions about one's own life, and have control over one's own actions and choices. It encompasses the freedom to choose one's own path, pursue personal goals, and shape one's identity and relationships. Privacy and agency are foundational elements of self-determination, as they enable individuals to define their own boundaries, make informed choices, and express their authentic selves.

As these concepts in the last resort depend on data and information, it should be made clear which data are referred to. And it is as simple as radical: all data related to a human existence, without any exception. Because in our opinion, without control over their data, an individual does not have privacy, and without privacy he does not have agency and can not reach self-determination.

Currently, privacy and self-determination are threatened in two contexts: political and commercial. There is increasing attention to both of them, yet still to a certain degree in separate discussions. Whilst this paper proposes a commercially driven grassroots initiative, the political freedom of the individual is always the ultimate point of reference. In the following, some illustrations are given.

If the aim is to own and control all data related to a human existence, the starting point is indeed the genetic setup of the individual itself and the potential dispositions embedded in this. But instead of having one's genome stored, analysed and used by one (or more) private or public organisations (under usage conditions and with permissions the individual never fully understands), an individual's genetic data will be stored in one secure place on a distributed ledger only, and he will grant access under the conditions required for each user and each instance as necessary only.

The aim must also be to store the entire medical history of an individual, without any exception. At the moment, this information is often with many different doctors and laboratories and can therefore not be used to its full potential. There are initiatives in countries like Germany to store all this

information in one central place („Digitale Patientenakte“), but this solution (if it will ever work) is provided for by a government, probably in cooperation with business partners, and in the last resort the individual is again at the mercy of third parties. Metaself wants to put the sovereign individual at the centre of all activities, and there is no reason why this should be otherwise.

The data to be collected, stored and processed will eventually also include all physical and digital activities, ranging from peoples' trails on the net, to mobility data, purchases (online and offline), consumption, and the like. As of 2023, it is established industry practice to collect, analyse and monetise data originating from individual human activities all the time, in different contexts, locations, legalities and in different use cases. The result is a rather chaotic situation, which is accepted largely because no alternative is available. Metaself wants to be this alternative. It wants to bring about a paradigm shift, after which the only point of reference is the sovereign human individual. One possible option of dealing with data generated in the past will be to demand the return of our data (at least partly), probably through lengthy and expensive class action lawsuits, which in turn will create publicity.

At the moment, this approach appears far from practicable because of the multitude of occasions and the variety of implications of the decisions that have to be made. But this demonstrates exactly the long-term vision of metaself as a personal assistant to the individual: metaself will support the individual to a very high degree to make these decisions in his best interest at any level in each situation. A clear understanding of the underlying data architecture is therefore paramount.

2.2. Data architectures

The extremely sophisticated, complex and reliable solutions for data storage, processing, and transmission built and operated by tech companies today are a marvellous achievement. Within some 20 -40 years they have created the backbone of our digital economy without which the world as we know it could not function.

The central criticism of metaself is that, in the last resort, all data (which ultimately originates from the behaviour of individual human actors) ends up to be physically controlled by a number of tech companies. This is partly due to the nature of the business models of those companies, and partly due to technical requirements: To preserve data integrity, sophisticated walls have to be built around the data, consisting of cybersecurity measures, data access policies, and rules regarding validation and access control.

The awareness of these problems has increased continuously in recent times and has led to practical policies such as the General Data Protection

Regulation (GDPR) in Europe, adopted in April 2016, and the the California Consumer Privacy Act (CCPA), passed in 2018. These are positive developments, but a distributed and agent-centric data architecture offers much simpler and more robust solutions.

Other recent developments, such as Personal Data Stores (“PDSs”), are aimed at giving users control over the processing of their personal data by enabling them to capture, aggregate and manage their personal data. But the details of such measures can become very complicated, and the continued enclosure of data still result in asymmetries in access to information and by extension asymmetries in power (Jansen, Cobbe, & Singh, 2020).

Alternative solutions have been provided by various types of **Distributed Ledger Technologies** (DLTs) that enable the decentralized storage and management of digital records. These technologies provide a means for multiple participants to maintain a synchronized and consistent database without the need for a central authority. They therefore employ cryptographic techniques and consensus algorithms to ensure the integrity and immutability of data across a network of nodes. Each participant in the network maintains a copy of the ledger, and transactions or data updates are propagated and verified by the network through a consensus mechanism.

The most well-known and widely adopted DLT is blockchain, which uses consensus mechanisms like Proof of Work (POW), Proof of Stake, (POS) and various others. Other types of DLTs such as directed acyclic graphs (DAGs) and hashgraph, offer different approaches to achieving decentralized consensus. The most important result is transparency: All participants have access to the same data, promoting transparency and accountability.

With blockchain and other DLTs, a lot can be achieved, but not everything at once. There are several conflicting objectives that need to be considered when making decisions (Kannengiesser et. al. 2019, Noack 2019). In most instances, they refer to six central properties: performance, security, usability, development-friendliness, anonymity, and compliance.

Performance. High performance aims for a high number of transactions per second. For this, the complexity of the employed algorithm and encryption method needs to be kept as simple as possible. Performance can also be achieved by reducing the number of nodes, but this compromises security. This can only be compensated by higher trust, which would likely require sacrificing participant anonymity. Nodes may also not contain a complete copy of the entire DLT. While less powerful devices can be used, the resilience of the system decreases.

Security. Maximum security in a DLT is achieved by increasing the network size, potentially excluding fraudulent nodes, and reducing the development

flexibility for smart contracts. These approaches contradict each other. For example, reducing the degree of anonymity of nodes helps prevent attacks by fraudulent nodes.

Usability. To enable DLT-capable applications and implement devices with limited access, the full replication of the ledger on each device, for example due to limited storage sizes, should be avoided. To make DLT design suitable for restricted devices, DLT designs can utilize lightweight nodes. Lightweight nodes do not contribute to increased availability or fault tolerance of the distributed ledger since they do not store a complete replication of the ledger or validate transactions.

Development flexibility. Significant development flexibility can be achieved through support for smart contracts that can be individually defined. The more expressive the supported programming languages are, the more errors can occur in smart contracts. Smart contract code is difficult to verify and, in particular, to test in terms of chained settlement. Therefore, smart contracts increase the risks of security breaches (e.g. The DAO Attack).

Anonymity. Achieving a high level of anonymity requires additional processing of transactions (e.g., mixing, stronger encryption). These processes are time-consuming and require additional computational power, which slows down performance. A large network size provides a higher degree of anonymity, but the performance in block-based distributed ledgers decreases. Additionally, traceability is limited or even impossible as transactions cannot be traced back to the issuing user.

Compliance requires a high degree of verifiability and compliance. Verifiability is associated with high transparency, leading to low anonymity. The level of compliance of a system always depends on the applicable norms and regulations. Standards and regulations can change, and systems need to adapt to those changes to maintain their compliance level. An inherent characteristic of DLT constructions is immutability. Therefore, adapting to compliance changes later becomes challenging. Thus, a high degree of immutability excludes adjustments to a high level of compliance without compromising integrity.

Many of these tradeoffs arise because the DLTs used are **data-centric**, with the integrity and security of data as the primary concern. An alternative are agent-centric data architectures.

2.3. Agent-centric architectures, Holochain

Agent-centric distributed systems are built around autonomous agents that have their own goals, knowledge, and decision-making capabilities. These agents can act autonomously, interact with each other, and collaborate to achieve overall system objectives. Key characteristics include:

Autonomy: Agents are autonomous entities capable of making decisions and taking actions independently based on their own internal state and the information they receive from the environment or other agents. They can perceive their surroundings, reason about it, and perform actions to achieve their objectives.

Interaction and Communication: Agents in a distributed system can interact and communicate with each other to exchange information, coordinate activities, and achieve shared goals. Communication can be direct or mediated through the environment.

Decentralization: Without central authority or control, each agent operates independently and makes decisions based on its local knowledge and interactions with other agents. This decentralization leads to increased flexibility, scalability, and resilience in the system.

Goal-orientation: Agents in an agent-centric system are driven by their individual goals and objectives. They have their own beliefs, desires, and intentions, and they strive to achieve their goals while considering the goals of other agents and the overall system objectives.

Adaptability and Learning: Agents in agent-centric systems can adapt to changing environments and learn from their experiences. They can acquire new knowledge, improve their decision-making abilities, and adjust their behavior over time.

Agent-centric distributed systems have been applied in various domains, such as intelligent transportation systems, robotics, smart grids, and decentralized marketplaces. They provide a flexible and scalable approach to designing complex distributed systems by leveraging the capabilities of individual autonomous agents.

Holochain is an agent-centric distributed generalised computing system that achieves many of the above criteria by employing several interrelated mechanisms:

Distributed Hash Table (DHT): Holochain uses a distributed hash table to store data. Unlike traditional blockchains, where all nodes store the entire transaction history, Holochain stores data in a peer-to-peer network. Each user's data is stored on their own device or chosen hosting provider, giving them control over their data.

Cryptographic signatures: Holochain utilizes cryptographic signatures to verify the authenticity of data. Each user has their private key, allowing them to sign and encrypt their data. This ensures that only the intended recipients can access and decrypt the data, maintaining privacy and security.

Local validation: In Holochain, data is validated locally by each user's device. This means that data is not shared with the entire network for validation, reducing the exposure of sensitive information. Each user can independently verify the integrity of the data they receive.

Secure communication channels: Holochain provides secure communication channels between users. It employs encryption and secure protocols to protect data transmission, preventing unauthorized access and eavesdropping.

Selective data sharing: Holochain allows users to selectively share data with specific individuals or groups based on their desired privacy settings. Users have granular control over the data they share, ensuring that sensitive information remains private.

Each participant in the system exercises his agency via a device under his control, recording their actions as entries on a personal digital journal. These entries are cryptographically signed by a private key, which provides the necessary provenance context. When the application requires them to share these entries publicly, they do so by broadcasting to a subset of their peers who use the same application. Those peers then take responsibility for validating, storing, and serving the entries to others.

Another very important feature of Holochain is the acknowledgement that data separated from its **provenance** has lost a critical part of its meaning (Harris-Braun 2021). All information has its origin in the subjective experience of an agent (usually a human actor) producing it, and Holochain is designed around this perspective. As more and more machines and artificial actors are going to produce vast amounts of data, we believe that the ability to prove the provenance of data in a human actor will become a critical feature.

Further detailed discussion is beyond the scope of this paper. But in view of the considerations above, Holochain provides a framework for building and running a distributed a peer-to-peer network of individual nodes. With it's agent-centric approach, a wide range of decentralized applications can be realized, including social networks, collaboration platforms, supply chain systems, and more. It offers an alternative approach to decentralized computing, focusing on user autonomy, data integrity, and scalability.

Finally: With this setup it is possible to run a distributed system without the data infrastructure we are using today provided by big tech companies. As long as individual actors can run digital devices under their control (if necessary on solar power) and connect digitally, the distributed network they have created can continue to operate.⁴

2.4. The individual, digital identity and trust

With the setup described above, it is possible to achieve the concept of Self-sovereign identity (SSI) as outlined by Christopher Allen in 2016 as well as the design principles for the Trust over IP Stack of the Trust over IP Foundation. Individual human actors will have complete control and ownership over their digital identity and personal data. This empowers them to manage and share their identity information in a secure, private, and decentralized manner, without relying on centralized authorities or other intermediaries.

Here a short reference to key principles of self-sovereign identity (Allen 2016). It should be pointed out that these become parameters in metasef that will be regulated automatically by the system according to individual preferences.

Existence: Metasef ensures that the individual and his physical existence are at the heart of identity. It simply makes public and accessible aspects of the self that already exist and safeguards a right to remain unknown in certain contexts.

Control: Metasef gives human actors full control over their identity information. They can decide what personal data to share, with whom, and for what purpose. They can also revoke access to their data at any time.

Access. Due to the architecture employed, users have access to their own data at all times, comprehensively and without any reservations. A user is always able to easily retrieve all the claims and other data within his identity without any gatekeepers.

Transparency. Systems and algorithms must be transparent. The foundation of all technology solutions must therefore be open source. Holochain is an open source project that welcomes all sorts of participation.

Persistence: Some aspects of the principle that identities must be long-lived can be seen as controversial. We believe that the owner must be the in

⁴ The community of people actively considering such scenarios might be quite small at the moment (e.g. <https://dwebcamp.org/>), but the experience of the past couple of years has shown that not everything we take for granted might continue forever.

control of all identifiers, including decentralised identifiers, and be able to decide on their persistence as long as there are no conflicts of interest.

Portability: Information and services about identity must be transportable. Identities must not be held by a singular third-party entity, even if it's a trusted entity that is expected to work in the best interest of the user. Entities can disappear, regimes may change, and users may move to different jurisdictions. Metaself ensures that the user remains in control of his identity no matter what.

Interoperability: Identities are of little value if they only work in limited niches. The goal of a distributed digital identity system must be to make identity information as widely available as possible, crossing boundaries to create global identities, without losing user control.

Consent: Users must agree to the use of their identity in a real and meaningful way. The current practice in Europe that requires each user to consent to terms and conditions they never fully read, illustrates the challenges involved in today's big data ecosystems with substantial information asymmetries and uneven bargaining power between individuals and organizations. We believe that metaself will produce these results.

Minimization. Disclosure of claims must be minimized and only involve the minimum amount of data necessary to accomplish the task at hand. Metaself will be the assistant to the individual that makes this the default.

Protection. The rights of users must be protected. To ensure this, identity authentication must occur through independent algorithms that are censorship-resistant and force-resilient and that are run in a decentralized manner.

Metaself is also fully compliant with the vision of the Trust over IP Foundation (Trust over IP Foundation 2021). The conceptual framework, as visualised in the **Trust Over IP Stack** consists of two pillars representing the technology stack and the governance stack, with four horizontal layers representing the technical requirements in Layers 1 and 2 and the human requirements in Layers 3 and 4.⁵

On the basis of the hardware-related „dry code“ principles, the central human-network related „wet code“ principles can be realised. They revolve around the central element that has to be the basis of all human relationships: **trust**. Whilst trust in a machine context is deterministic and can be built around various „if...then“ rules, human trust is much more elusive („fuzzy“)

⁵ For more comprehensive visualisations head to <https://trustoverip.org/>

and can be based on an unlimited number of conditions, some of which not even clearly defined and/or subconscious.

Trust is fundamentally a belief of either a single actor or a group in the integrity, ability or character of an entity. The purpose of this belief is to help an actor make a decision relative to a specific action. Trust is therefore ultimately an active decision taken by humans based on the information available which can be subsumed under the concept of trustworthiness.

The **trustworthiness** of an entity regards the extent to which it is deserving of trust. It is based on assurance levels of the pillars of security, privacy, ethics, robustness, reliability and resiliency in digital identity systems. On the basis of this approach the relational, directional, contextual, limited and transitive qualities of trust can become manageable and trust can be realised.

This takes us to a central element of the changes envisioned: whilst „belief“ is still an active effort, „acceptance“ does not require an effort. Metaself envisions a world where we leave „belief“ behind us, and instead accept that something is true because we know that the information provided by the system is correct. In so doing, we reach a state of **trustless trust**.

As has been shown, it is technically possible to build distributed systems around individual human actors that preserve key values such as privacy and enable those actors to become sovereign. In the following, we want to examine what can be achieved when individual human actors begin to connect and cooperate in such a system.

3. Distributed organization and cooperation

As a species, we have become the dominant life form on this planet in a very short time because we have taken social organisation and cooperation to completely new levels. This is mainly due to our ability to construct abstract concepts in our imagination (such as nations, money, law) and to act on them collectively.⁶ These „imagined realities“ have real-world consequences by shaping our behaviors, beliefs, and actions, allowing us to create social structures and systems that are far more extensive and intricate than those of other species.

We envision a future where the individual self will be connected with other individuals directly and in a more dynamic ways. Every person will be enabled

⁶ Since the publication of *Sapiens: A Brief History of Humankind* by Yuval Noah Harari in 2015, the awareness of these topics has increased significantly and has become part of mainstream discussions.

to interact directly with others without interference and to develop their relationships. Such a system can be called liquid.

3.1. Liquid reign ⁷

Human organisation and cooperation in general and organizational behavior in particular have been studied intensively for quite some time now, and no attempt can be made to take all knowledge available into account here. But we will proceed along some widely acknowledged key concepts. The underlying question in each case will be, if a new and dynamic / liquid form of organization facilitated by the distributed digital infrastructure outline before has advantages over established ways.

3.1.1. Individual Behavior:

In the last resort, we are driven by the same two protocols as any other life-form on our planet: survival of the individual and survival of the species. Arguably, this is the **meaning of life** (at least biologically, as we strive to pass our genes on to the next generation), the real challenge being to find **meaning in life**.

One central element of our current predicament is that we still function with impulses that have served thousands of generations of our ancestors well but produce absurd results in our current state of over-abundance and choice. It can be argued that the complex processes in our brains and our bodies revolving around mesolimbic dopamine pathways linked to motivation, action and (calorific) reward are currently being hijacked by modern technology and business models primed for profit maximisation (Hagens 2011).

As has been pointed out (Hagens, 2011), that whilst the human brain of an online shopper, computer gamer or a social media consumer lights up in an fMRI the same way a chimpanzee's does when finding a nut or berry, **there is no instinctual 'full' signal in modern human brains** - we are addicted to the unexpected reward, the next encounter, episode, or email, at an ever increasing pace, and the 'wanting' becomes a stronger emotion than the 'having'. For many, the meaning of life therefore has become consumption.

Awareness of these challenges is growing slowly, and any claim to provide a simple solution or a quick fix would be vain. But we must deal with the problem that our **needs** (defined as a state of felt deprivation) and our **wants** (the actual expression of those needs) always arise in one human actor who is currently at the mercy of powerful third parties with rather dumb key performance metrics (maximize share of wallet and/or share of attention).

⁷ In *Liquid Reign*, Tim Reutemann sketches interesting possible scenarios in the near future.

In fact, as has been argued (Andersen 2022), three central assumptions of our behaviour turn out to be insufficient, if not downright wrong: 1) the emphasis of capitalist economic theory from the 1700s and onwards on man's self-interest and urge for competition, 2) the interpretation of Darwin's "survival of the fittest" from 1859, and 3) the concept of Homo Economicus, stating that we constantly strive to optimize personal economic gain.

Amongst other things, metaself eventually can achieve two things for the individual: 1) provide them with comprehensive information about the implication and consequences of their behaviour, and 2) enable them to connect in more meaningful ways to other individuals. We hope that this will eventually bring about meaningful change. Any speculation how this might work out in detail is not helpful at the moment.

3.1.2. Group behavior and networks

It is part of human complexity that, whilst we pursue selfish behaviour, we are also collaborative, social and caring, and depend on strong communities and other people's recognition. The individual human actor therefore normally not only wants to connect to others, he wants to connect to the peer with the most utility. Analysed in the context of network theory, this is usually the peer with the most connections.

The result of this impulse of wanting to connect with the hub in the network rather than with someone on the fringes are **scale-free networks**, a pattern that occurs across all kinds of natural and artificial real-world systems. Key characteristics of a scale-free network are:

Highly Connected Nodes: Scale-free networks have a small number of nodes, the hubs, that are highly connected. These hubs have a significantly larger number of connections compared to other nodes in the network. Hubs play a critical role in maintaining the network's connectivity and influencing information flow.

Power-law Degree Distribution: The degree distribution of a scale-free network follows a power-law distribution, meaning that the probability of a node having a certain number of connections decreases exponentially with increasing degree. This distribution implies that there are many nodes with a low degree and a few nodes with an exceptionally high degree.

Resilience to Random Node Removal: Scale-free networks tend to be robust against random node failures because most nodes have relatively few connections, and the removal of any individual node has a minimal impact on the overall connectivity of the network. However, they can be vulnerable to

targeted attacks that specifically target highly connected nodes, as their removal can disrupt the network significantly.

Small World Phenomenon: Scale-free networks often exhibit the "small world" phenomenon, characterized by short average path lengths between nodes. Despite the presence of hubs, most nodes can be reached from any other node in a relatively small number of steps.

Preferential Attachment: Scale-free networks often evolve through a process called preferential attachment, where new nodes entering the network are more likely to connect to already highly connected nodes. This mechanism further reinforces the uneven distribution of connections, leading to the emergence of hubs.

As a result, the scale-free network structure is the most prevalent type in natural and cultural self-organizing networks, such as social networks, the internet, biological networks, citation networks, and transportation networks. This is also one of the reasons why wealth in human societies is distributed so unevenly („the rich get richer ...“). In postmodern capitalist societies this striving for hubs creates immense economic inequality and established political structures that redistribute this wealth are currently under enormous strain.

It is still very early in the ideation phase, but the assumption is that in the liquid and data-driven structure envisaged by metasef, parameters like the utility of connections can be measured in much more detailed and comprehensive ways, and as a result each participant (or their assistant) can constantly adapt their relationships according to dynamic criteria.

3.1.3 Power and decision-making

At least since the increase in complexity in human organization with the advent of agriculture, the prevalent form of human organization has been some sort of hierarchy, with three basic approaches to decision-making: autocratic, majority/democratic, and consensual.

Autocratic decision-making involves a single individual or a small group of individuals at the top of a hierarchy making decisions without significant input or participation from others. The decision-making power lies with the leader or leaders, and they exercise authority to determine the course of action. This approach can be efficient in situations that require quick decisions or when a clear chain of command is necessary. However, it can lead to limited input, reduce morale, and a lack of diversity in perspectives.

Democratic / majority decision making involves the participation and input of multiple individuals in the decision-making process. It allows for open discussion, debate, and voting to determine the final decision. In a democratic approach, decisions are often made through majority rule, with each participant having equal voting power. This approach fosters inclusivity, promotes diverse perspectives, and encourages buy-in from participants. However, democratic decision-making can be time-consuming and may result in slower decision-making processes.

Consensual decision making requires that all participants must agree on the final outcome. It involves active collaboration, negotiation, and finding common ground among the individuals involved. Consensual decision-making aims to build cooperation, trust, and a sense of ownership among participants. It values the input and concerns of all stakeholders and strives for win-win solutions. However, this approach can be challenging, as achieving consensus can be time-consuming and may require extensive discussion and compromise.

Metaself wants to build a system where eventually all decisions are taken in the best possible way with the most suitable mechanism. This is a very subjective statement which depends on the values each participant applies and the goals they want to achieve. But this is exactly the point: values, goals, tradeoffs, compromises eventually will become parameters that can be changed and adapted, like the settings in a software. The result will be a highly dynamic / liquid form of organisation.

The concept of **liquid democracy**, also known as delegative democracy or proxy voting, can not describe the eventual functioning of metaself in detail, but it gives some indications. It is a form of democratic governance that combines representative and direct democracy elements.

In this system, each person has the freedom to either participate directly in decision-making or delegate their voting power to someone else. Delegation can be done on a per-issue basis, allowing individuals to choose different proxies for different matters based on their expertise or trust. Delegations can also be revoked or transferred at any time, providing flexibility and adaptability.

As mentioned before, we envision a system where eventually many decision-making parameters (potentially all relevant parameters) can be identified, quantified, and considered. The individual human actor must have final authority to decide but many decisions will be taken automatically by his assistant according to the preferences he has set.

The role of the technological platform will not only be to facilitate voting, delegation, and tracking of decisions, but also to enable transparent and

accountable governance, ensuring that delegated votes are accurately represented and preventing potential abuse or manipulation.

The previous chapters have shown in an abstract way that the individual human can become a sovereign actor in the digital world and that they can connect to and cooperate with other actors in dynamic ways. But how will it fit in the real world?

3.2. Cooperative legal structure

As has become clear by now, metaself proposes to depart from a number of accepted patterns and practices. It will therefore be anchored in the real world through the legal structure of a **Cooperative under European Law**. This Cooperative will serve three purposes only:

- to safeguard the digital self-determination of its members,
- to compensate them fairly and directly for the use of their data, and
- to facilitate individual and collective growth through insights from data.

The purpose is explicitly not the maximization of profits for outside shareholders. All value that is created will be distributed immediately and continuously as micro-transactions to its members.⁸

Whilst this cooperative will begin with a small number of initial members at its core, it will eventually be open to anyone who wants to participate with a very low threshold for entry and exit. A central plank for this purpose will be a dynamic equity split.

A dynamic equity split was originally proposed as an approach to allocating ownership in startups and early-stage companies that adjusts based on the relative contributions of each participant over time (Moyer 2016). Instead of fixed equity allocations, a dynamic equity split recognizes that contributions to a venture can change and evolve as the company grows.

Metaself will implement and operate a dynamic (or liquid) equity split on a permanent basis, with all input and all output calculated and allocated immediately and continuously. The contributions can be time, money, ideas, relationships, data (and others that we cannot think of yet). Eventually, the system will be able to calculate the value of any input or data set that a human actor contributes to the system, no matter in which context and for which

⁸ Just for illustration: in 2022 just over 80% of Google's 280 billion US\$ in revenues came from advertising. In the last resort, Google sold information about us, which we gave them in exchange for their „free“ services. What if we build a system where commercial agents pay users directly for access to them?

purpose it was used. And it will do so with such a high degree of reliability that no one has any need any more to question these calculations.

At this early stage, it is not possible to grasp the multitude of issues that have to be solved on the way. That will be the result of collective reasoning and decision-making. But it seems obvious that new approaches have to be built.

3.3. Economy and Ecology

Finally, it is important to remember that our human development has always been intricately linked to the development of our planetary ecosystem (Gowdy & Krall 2014). Whilst we lived in more or less egalitarian, roaming bands for some 300,000 years, it was the stabilization of the climate around 11,000 years ago that allowed agriculture to develop in various separate locations around the world. And as production began to exceed the group's immediate caloric needs, groups of humans began to organize around physical surplus. As some of the population no longer had to devote their time to hunting and gathering, this allowed the development of specialization, resulting in new jobs, hierarchies, and complexity.

And it was only the large-scale discovery of fossil carbon and the invention of technologies to use it as fuel in the 18th century, that this process was accelerated. As a result, the past 200 years became a unique period in human history when more and cheaper resources led to sharp productivity increases and unprecedented economic growth enabling diverse and richer societies.

And maybe it is no coincidence that since the beginning of the 21st century, as energy and resources are again becoming constraining factors and the effects of our civilisation on the environment becomes obvious, that societies are becoming polarized, losing trust in governments, media, and science, and revert to small group behaviour.

It would be naive to propagate metaself as a fix for everything. But eventually it can provide each individual human (and other) actor with detailed information about their economic and ecological impact. And may be this induces more and more people to set other priorities. There are hopeful signs that more and more people want to minimize resource consumption, waste generation, and environmental impact by promoting the continual use, recycling, and regeneration of materials and products.⁹

⁹ But may be the above patterns of production and consumption become redundant anyhow, as we are moving into a world where everybody can make (almost) everything, as Neil Gershenfeld suggested in Episode #380 of the Lex Fridman Podcast. But then these artificial structures will probably also possess some sort of self-awareness and intelligence.

It is not helpful to engage in too much speculation here, but with the setup described, we have the chance to build structures that are more egalitarian and cooperative than the social hierarchies and inequalities that emerged with the rise of agriculture and the resultant consolidation of power.

4. Meta thinking, meta-self

The setup discussed so far already provides a number of potential benefits in terms of privacy, agency and self-determination. But what exactly is eventually going to happen under the hood of metasef? Like many other assistants under development now, metasef will do real-time calculations on an unprecedented scale in order to reduce uncertainty and complexity in more and more areas and situations. This will enable human actors to make better decisions - individually and collectively.

The mental processes that govern human behaviour have become a topic of wide and intense discussion within a relatively short time.¹⁰ The fact that these processes are in large part based on simplifications, shortcuts and assumptions has grown steadily.¹¹ The awareness that they can be influenced subconsciously from the outside has also grown quickly and been turned into practical policy.¹² And the exploitation of these processes for commercial purposes on a vast scale has come to the forefront of public discussion, certainly since the publication of Shoshana Zuboff's *Surveillance Capitalism* in 2019 and the release of the movie *The Social Dilemma* in 2020.

Metasef wants to provide an alternative by supporting these mental processes on the basis of the distributed data architecture and social organization outlined before. Only in this combination will the individual be truly sovereign and at the centre of what is to come. In the following, we will give some tentative ideas, how these processes could look like, what this means for the individual, and what this means to the various collectives involved.

4.1. Mental processes

The complex mental processes that work together when we humans think have been studied for quite some time now, and no serious attempt can be

¹⁰ Certainly since the publication of Daniel Kahnemann's, *Thinking, Fast and Slow*, in 2012, these issues have been discussed by the broad public.

¹¹ E.g. Gigerenzer 2013

¹² Richard Thaler and Cass Sunstein, *Nudge: Improving Decisions About Health, Wealth and Happiness* in 2009.

made here to reference the vast amount of research that has been produced.¹³ In due course, it is intended to set up a scientific advisory board that will give metaself a broad scientific foundation. The following discussion wants to illustrate some possible changes in the various cognitive functions and mechanisms that allow us to perceive, process, store, and manipulate information. The focus on four sub-topics is completely arbitrary and intended to illustrate potential developments.

4.1.1. Perception

Perception involves the brain's processing, organization and interpretation of sensory information from our environment. The stimuli can be visual, auditory, tactile, olfactory, and gustatory inputs. Importantly, our brains simultaneously also processes stimuli from our physical body in a variety of forms, and together they form the basis of our behaviour. It has been argued (Damasio 2021) that any theory of mind that relies exclusively on the nervous system to account for minds and consciousness is bound to fail.

It has also been pointed out (Juarrero 2023) that sensory information is inherently relational and contextual (inserting a hand into warm or cold water), and that because they are not independent of perception, relational properties dependent on an observer have to be considered subjective, and their reality is therefore in the last resort derivative.

Further steps in perception that can not be discussed here include Neural Processing, Feature Extraction, Interpretation, Cognitive Selection and many more. Combined, they allow us to recognize patterns and make sense of the world around us. For many generations of our ancestors this was the basis to perform the same two basic protocols that every other life-form fulfils: survival of the individual and survival of the species. Because they were so successful, we are where we are.

The ultimate result of perception, Meaning Making, has always played a certain role for our ancestors. But, probably it has never been more important than at this moment in our history. And, as has been argued (Andersen 2022, p. 102), „our meaning-making is limited, and before we understand the shortcomings of how we understand the world, we cannot make conscious and deliberate changes“.

How exactly it will feel like when we can perceive more of our environment, and how much is possible with technology is pure speculation, but human adaptability has proven to be quite impressive in the past.

¹³ Kurzweil 2012 and Damasio 2021 are only snapshots of a very dynamic ongoing process.

4.1.2. Attention

Human attention is the cognitive process of selectively focusing on specific stimuli while filtering out others. As Jim Rutt put it, „Attention is the cursor of consciousness“. ¹⁴ It is crucial for perception, learning, memory, and various other cognitive tasks. Selective attention allows us to concentrate on what is relevant to our goals, while attentional capture can involuntarily divert our focus.

Attentional control enables us to regulate our focus, shift attention, and resist distractions. Attention can be influenced by the characteristics of stimuli (bottom-up processes) or internal factors like goals and expectations (top-down processes). Various brain networks, including the frontal and parietal cortex, contribute to attentional processes. Attention enhances cognitive performance, including perception, learning, memory encoding, problem-solving, and decision-making.

Various assistants that help the individual focus and direct attention are already operational today (e.g. Apple Focus). That is probably only a very humble beginning. With the comprehensive data architecture envisioned for metaself described above, the individual will be able to control and direct their attention in much more comprehensive and detailed ways. As explained, it is crucial that this assistant only takes the sovereign individual and other sovereign individuals as a point of reference.

4.1.3. Memory

Memory refers to the processes involved in encoding, storing, and retrieving information. It enables us to retain and recall past experiences, knowledge, and information, which we can draw upon to inform our thoughts and decision-making. Awareness that we are very selective in the way we memorize for example our life is quite widespread, and various techniques to improve our memory have been offered for quite some time.

Technology-based tools to enhance human memory are already being used widely, ranging from note-taking apps, task managers, digital calendars, to wearable trackers and the like. Brain-computer interfaces are under development, and permanently connected individuals with access to vast amounts of data become viable.

Any speculation where this might lead us probably misses the point: in the last resort, human actors - as operational today - will normally retrieve stored

¹⁴ In several recent episodes of the Jim Rutt Show, e.g. #148 with Antonio Demasio.

information to construct stories that is coherent to them.¹⁵ The real challenge will be to reconcile the various stories that are being created, and may be out of this effort there can arise a new level of awareness and consciousness.

4.1.4. Reasoning, Problem-Solving, Decision-Making

In the last resort, the mental processes that have brought us to where we are revolve around making decisions for further development. The mental processes involved include: **Reasoning**, the process of logically evaluating and drawing conclusions based on available information and evidence; **Problem-solving**, which involves identifying and implementing solutions to overcome obstacles or achieve desired outcomes; and **Decision-making**, the process of selecting the best course of action from among multiple alternatives. So we are constantly weighing options, considering consequences, and making choices based on available information, preferences, and goals.

As so much of our environment is far beyond the scope of our mental capacities, we have reached our position as a species because we can make practicable decisions with either

- too much information,
- too little information,
- under pressure of time,
- and with multiple internal and external influences.

One of the most important tools we employ for these purposes are **heuristics**: mental shortcuts or rules of thumb that help us make decisions, solve problems, and navigate through complex situations (Gigerenzer & Todd 1999). They help us simplify information processing and allow us to make quick judgments or reach conclusions without engaging in exhaustive analysis or computation. Heuristics therefore inevitably lead to biases and errors in certain situations. A brief mention of some common types of heuristics illustrates the potential for optimization:¹⁶

Availability Heuristic involves estimating the likelihood or frequency of an event based on how easily examples or instances come to mind. We tend to judge events as more probable if we can easily recall related examples, even if those examples are not representative of the actual probabilities.

¹⁵ As Harari (2018, p. 11) points out: „Humans think in stories rather than in facts, numbers or equations, and the simpler the story, the better“.

¹⁶ For a much more comprehensive overview head to: https://en.wikipedia.org/wiki/File:Cognitive_bias_codex_en.svg

Representativeness Heuristic involves making judgments or decisions based on how well an individual or event matches a particular prototype or stereotype. We may rely on this heuristic to assess the likelihood of an outcome based on its similarity to a familiar category or prototype.

Anchoring and Adjustment Heuristic involves starting with an initial reference point (an anchor) and adjusting subsequent judgments or estimates based on this anchor. The initial anchor can bias subsequent judgments, even if the anchor is irrelevant or arbitrary.

Confirmation bias is a tendency to seek or interpret information in a way that confirms pre-existing beliefs or hypotheses. This bias can lead individuals to selectively attend to and remember information that supports their existing views while ignoring or discounting contradictory evidence.

Satisficing is a decision-making strategy that involves choosing an option that is "good enough" or satisfactory, rather than seeking the optimal or best possible choice. It focuses on reaching a solution that meets a minimum threshold of acceptability rather than exhaustive search for the best outcome.

Detailed knowledge of these processes is growing continuously, and Kahneman's simplification („System 1“ and „System 2“) has found its way into public discourse. To suggest that metasef will speed up Kahneman's „System 2“ would be a gross simplification. In a very broad sense, however, this goes in the right direction. How exactly this will work out, can not be predicted as of now. But there is little doubt that we will be able to process more information and make better decision with a variety of tools that are under development right now. The real question is which metrics are employed to define „better“ decisions.

The long-term aim of metasef is not to make an individual human actor more competitive in comparison to his peers. Metasef eventually wants to achieve a new level of collective consciousness that balances individual and collective development. So whilst the individual user of metasef will be free to set the metrics of his assistant to „competitive“, we hope that individual competitiveness will have a lower priority for more and more people as they can process more and more information.

4.2. Human interaction with metasef

How exactly the individual human will eventually interact with metasef is pure speculation at the moment. A variety of assistants are already at our disposal, with many more under development. The shift to voice control (Alexa, Siri, etc) has been gathering pace for some time now, but with the large-scale adoption of Large Language Models in 2023 we are moving into uncharted

territory. Artificial agents produce language that clearly affect humans - individually and collectively - with mechanisms nobody understands.¹⁷

We still believe that language and voice will be central elements of our interaction with the assistants to come. The role of visuals is difficult to predict as of now. So with some imagination, we can envisage a scenario where at some stage metasef becomes a voice in our head when we are thinking.

Our "inner monologue" that helps us with all the mental processes described above develops during childhood as we acquire language and our cognitive abilities. Initially, this monologue is indeed external, as can be observed when little children are playing alone and talking to themselves („now I do this, and then I do that...“). This monologue becomes internalized as we interact with our environment, but throughout our lives, it helps us process subjective experience and is a manifestation of our own consciousness.

As we all know, this inner voice often engages in an „internal dialogue“, where we have conversations or debates with ourselves, weigh pros and cons, and analyze situations. The assumption underlying metasef is that we will be able to make better decisions if the partner in this internal dialogue is much better informed than it is now. What we do not want is any third party or technology listening in on us whilst we are doing this. That is what metasef stands for.

4.3. Consequences for the individual

We believe that the effects of enhancing the mental capacities of individual human actors in general and as proposed for metasef will lead to far-reaching changes in our understanding of consciousness and our construct of „the self“. These are topics that are being discussed with great intellectual vigour (e.g. Koch 2004, Tononi 2012, Damasio 2021, Azarian 2022, and Juarrero 2023, to name just a few), and no attempt can be made here to contribute to this academic discourse. But in many of these discussions there is a clear expectation that something new will emerge.¹⁸

4.3.1. Consciousness

¹⁷ The fact that openai.com is anything but open is criticised quite widely now.

¹⁸ Andersen (2022) argues with reference to Henriques that we are at a fifth joint point.

Consciousness is a complex and multifaceted phenomenon that refers to the state or quality of awareness and subjective experience. „The hard problem of consciousness“ can certainly not be addressed in any depth here.

Two of the most influential theories of consciousness today, the **global workplace theory** (Koch 2004) and **integrated information theory** (Tononi 2012), take integration as a necessary and/or sufficient condition for the emergence of conscious awareness and phenomenal feels. Damasio (2021) argues that any understanding of consciousness has to „begin with the beginning“, which is the running of life inside an organism, and that our selves are based on how we feel about the interior of our organism.

From this perspective there can be little doubt that animals around us are not only feeling and minded creatures like we are, but also conscious. The difference lies in the capacity and complexity of their minds in comparison to ours. The way we treat other life-forms then is in the first place a result of our limited perception (i.e. empathy). What an enhancement of our mental capacities enables to perceive more dimensions means remains to be seen.

It is also accepted that there is intelligent behaviour without consciousness. Even simple organisms without a nervous system adjust their behavior and put themselves in situations that are most conducive to continuing their life. They strive for **homeostasis**, a set of goals and regulatory activities that have one purpose and one purpose, maintaining life for as long as that is possible according to the controls of your genome.

Damasio also emphasizes the importance to keep mind and consciousness separate. According to this point of view, mind is the result of the patterns that are achieved by creating an internal map. And the big challenge is to show how mind evolved from matter.¹⁹ And our mind is where we construct a self.

Bach (2023) distinguishes between sentience and consciousness. Sentience is the ability of a system to make sense of its relationship to the world. So it basically understands what it is and what it is doing.²⁰ And consciousness is slightly different from sentience in that it is a real-time model of self-reflexive attention and the content that we attend to.

4.3.2. The Self

The human conception of The Self has been a central topic of human thought for at least 2400 years. The only thing we can really agree upon here

¹⁹ Terrence Deacon on Mind's emergence from matter. JRS EP #157

²⁰ He continues to argue that „a corporation like Intel is sentient, because it has a good model of what it is ... and the necessary cognition is largely facilitated by people“.

is that in the last resort it is only a construct that originates in our conscious human minds as described above. And it is also widely acknowledged that at its core it involves creating a narrative or story of one's life that provides a sense of continuity and coherence. This narrative helps us to connect past experiences, present circumstances and future aspirations, contributing to a sense of personal identity and meaning.

As Andersen (2022) pointed out, there can be no discussion of The Self separate from the concept of free will. She introduces the very useful concept of „fuzzy free will“ that does not regard The Self and free will in binary either / or logic, but places them on a continuum.

Her fuzzy logic concept of Self and free will asks:

- How much of a self-conscious sense of Self does a particular individual have?
- How much free will does a particular individual have?

The answers place human individuals on a point in two continuums:

- From no sense of Self at all – to meta-cognition of Self.
- From no free will at all – to absolute free will; the conscious Self is its own master.

Her graphical depiction of these two interacting continuums contains a central question:

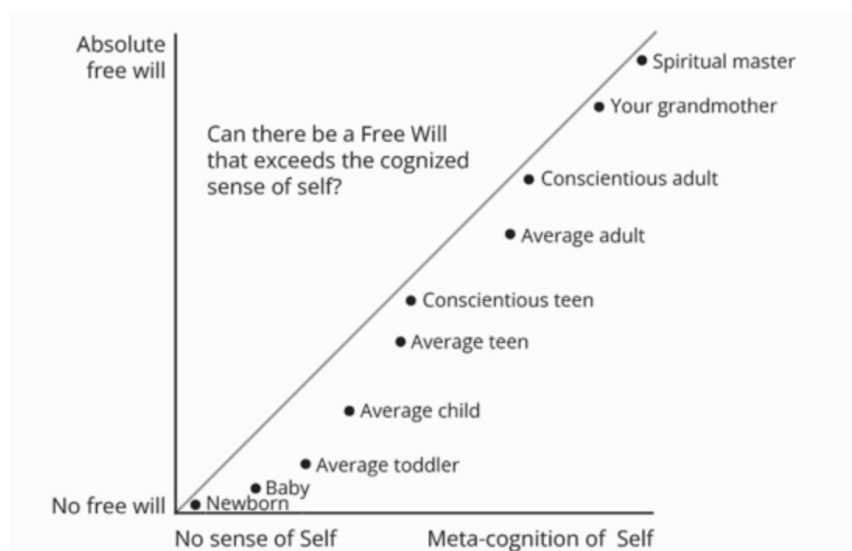


Figure: Andersen (2022, p 81)

4.4. Consequences for the collective

The challenges that begin when an individual human actor wants to exercise his free will in a social context with other human actors is leading us to some central aspects of metasef. As mentioned before, metasef does not want to make individual human actors more competitive than their peers, but to achieve a new level of conscious based on values.

As has been shown by researchers like Robert Sapolsky (2017), various species of animals can display an enormous degree of social complexities, hierarchical structures, cooperation, and competition. His research also demonstrated that animals are capable of empathy, collective problem-solving, and behaviors that resemble laughter or grief.

But we are (still) the only species that is capable of codifying such collective behavioural rules into abstract concepts. Any platform for human progress has to take such concepts into account. The most obvious ones are morals and ethics. Whilst they are often used interchangeably, it is important to be clear about the difference between the two:

Morals normally refer to individual beliefs, principles, or values that guide one's personal behavior and decision-making. Whilst morals are often shaped by cultural, religious, or societal influences, they can vary from person to person. They tend to be more subjective and reflect an individual's sense of right and wrong and are therefore more associated with a person's character and conscience.

Ethics are broader principles or standards that guide the conduct of individuals to more objective and systematic standards. Ethics are aiming to establish a framework for evaluating and making decisions about what is morally right or wrong. They often involve a set of rules, principles, or codes of conduct that prescribe behavior to the agreed-upon standards.

Andersen proposes the concept of *Bildung* as „the combination of upbringing, education, enculturation, and moral and emotional development“ based on the German-language thinkers of the late 18th and early 19th century.²¹ *Bildung* will enable individuals to feel their true emotions and to overrule them in order to care about people and society as a whole.

And as Juarrero (2023, p.31) has shown, values, ethics, and morals can emerge over evolutionary time as effects of coherence-making among interacting human beings. As a result, any personal identity is grounded in dynamic interdependencies that keep coherent structures whole.

²¹ Herder, Goethe, Kant, Schiller, Pestalozzi, Fichte, von Humboldt, Hegel

So it is the workings of enabling and constitutive constraints—among individual entities, processes, and actions—that generate novel properties. Values, ethics, and morals can emerge over evolutionary time as effects of coherence making among interacting human beings (ibid.).

On the basis of such theories, we propose metasef as a form of digital organisation of sovereign individuals that incorporates values. As described earlier, metasef will be a very dynamic form of organization and any participating individual can set his metrics according to his preferences. In sharing his preferences with others they can find common ground. In our belief, being connected with other individuals on the basis of shared values is also the only possibility to have agency in the age of machines.

4.5. Metasef and values

As suggested by Joscha Bach (2023), it is possible to build a system that is composed of multiple agents that are autonomous who cooperate and make a society of mind with values that need not be too complicated. He refers to the seven virtues espoused by Thomas Aquinas. Once stripped of their religious baggage, they offer simple rules of conduct and an entirely rationalist epistemology:

The **first four policies** Thomas Aquinas calls the rational policies or the **practical virtues**, that are accessible to every rational agent.

Prudence: the virtue of practical wisdom and sound judgment. It involves the ability to discern what is right and make morally sound decisions in specific situation.

Justice: the virtue of giving each person their due. It involves fairness, equity, and the establishment of right relationships within society.

Fortitude: the virtue of facing difficulties, challenges, and dangers with strength and resilience. It enables individuals to persevere in the pursuit of moral and virtuous actions, even in the face of adversity.

Temperance: the virtue of self-control and moderation. It involves the proper management and restraint of one's desires and appetites, promoting balance and harmony in one's actions.

It is completely feasible to quantify such values in a distributed network of sovereign individuals. As mentioned, nobody is forced to reveal anything

about himself, but he can give our assistants permission to use this information when it is deemed necessary or desirable.

What Thomas of Aquinas called the **divine virtues** can be re-interpreted as policies for a multi-agent system to merge into a next-level agent:

Faith: according to Thomas Aquinas the belief in a true divine being, according to Joscha Bach the belief in a next level agent.

Love: the willingness to act not in some kind of abstract sense, but with other agents that serve that same next level agent and coordinate with them.

Hope: the willingness to invest in a higher-level agent before it is there, before it can give you any return.

Bach points out that these concepts are logically derived policies for a multi-agent system that are forming a coherent next-level agent. On this basis it is possible to build a system that is composed of many sub-agencies.

This setup allows humans to go beyond the tribal mode in which we only have reputation systems and personal bonds, and to discover that we are serving a transcendental next-level agent that we are building and implementing together. It is the result of the concerted activity of people who decide to submit to that agent based on the values they have agreed upon.

5. From Aporia to Emergence

Metaself offers a long-term vision for development with tools that are available and functional today.

We propose to start with the recognition that (with metaself and as humans) we are at a moment of Aporia: established concepts and ways of thinking are losing their viability and question arise that can only be answered thinking differently about problems (Snowden 2023).

A suitable reference is provided by the Cynefin framework, that assumes three types of system: ordered, complex and chaotic, with solid, liquid and gas as metaphors. Aporia is the triple point, where it is not clear whether something will become solid, liquid or gas.

We therefore propose to focus on „the next right thing“, or „the adjacent possible“ (Kauffman 1996): to map where we are and to map where we can go next. Put simply: „The more you know about which constraints are in play, the more you can manage those constraints, the more you can influence

emergence.“ (Snowden 2023) And out of a clear understanding and handling of these constraints, there will eventually result coherence and a new form of identity.

And following the argument of Juarrero (2023), we can assume that „personal identity, which has been thought to be conferred through internal traits (essential natures), is grounded in dynamic interdependencies that keep coherent structures whole. This challenges our ideas of identity, as well as the notion that stability means inflexible rigidity“.

References

Allen, Ch. (2016). The Path to Self-sovereign Identity. <http://www.lifewithalacrity.com/2016/04/the-path-to-self-sovereign-identity.html>

Ammous, S. (2018). *The Bitcoin Standard: The Decentralized Alternative to Central Banking*. Wiley John & Sons.

Andersen, L. R. (2022). *Libertism: Grasping the 21st Century*. Nordic Bildung.

Azarian, B. (2022). *The Romance of Reality*. Benbella Books.

Bach, J. (2020). Episode #101 of the Lex Fridman Podcast.

Bach, J. (2023). Transcript of Currents #083 of the Jim Rutt Show. <https://jimruttschshow.blubrry.net/the-jim-rutt-show-transcripts/transcript-of-currents-083-joscha-bach-on-synthetic-intelligence/>

Burmeister, O., d'Aoust, P., Fernando, A., Ul-Haq, A., & Wahlstrom, K. (2021). Exploring co-design considerations for embedding privacy in holochain apps: A value sensitive design perspective. In M. Aria-Oliva, J. Pelegrín-Borondo, K. Murata, & E. R. Lara (Eds.), *19th International Conference on the Ethical and Social Impacts of ICT: Normal Technology Ethics Proceedings of the ETHICOMP 2021* (pp. 96-99). University of La Rioja.

Cameron, K. (2018). Let's find a more accurate term than 'Self-Sovereign Identity'. <https://www.identityblog.com/?p=1693>.

Camilleri, T., Rockey, S., Dunbar, R. (2023). *The Social Brain: The Psychology of Successful Groups*. Cornerstone.

Damasio, A. (2021). *Feeling and Knowing: Making Minds Conscious*. Pantheon.

Dawkins, R. (1974). *The Selfish Gene*. Oxford University Press.

Dunbar, R. (1998). *The Social Brain Hypothesis and Human Evolution*. Oxford Research Encyclopedia of Psychology. <https://doi.org/10.1093/acrefore/9780190236557.013.44>

Gigerenzer, G., Todd, P.M. (1999). *Simple Heuristics that Make Us Smart*. Oxford University Press.

- Gigerenzer, G. (2013). *Risk Savy: How to make Good Decisions*. Penguin.
- Gowdy, J., Krall, L. (2014). Agriculture as a major evolutionary transition to human ultrasociality. *Journal of Bioeconomics* 16(2), 179–202.
- Graeber, D., Wengrow, D. (2021). *The dawn of everything: A new History of Humanity*. Penguin
- Harari, Y. N. (2015). *Sapiens: A Brief History of Humankind*. Harper.
- Harari, Y. N. (2017). *Homo Deus: A Brief History of Tomorrow*. Harper.
- Harari, Y. N. (2018). *21 Lessons for the 21st Century*. Spiegel & Grau.
- Harris-Braun, E., Luck, N., & Brock, A. (2018). Holochain: scalable agent-centric distributed computing. <https://github.com/holochain/holochain-proto/blob/whitepaper/holochain.pdf>
- Harris-Braun, E. (2021). Decentralized Next-level Collaboration Apps with Syn. <https://blog.holochain.org/decentralized-next-level-collaboration-apps-with-syn/>
- Janssen, H., Cobbe, J., Norval, Ch., Singh, J. (2020). Decentralised Data Processing: Personal Data Stores and the GDPR. *International Data Privacy Law*, Volume 10, Issue 4 Pages 356–384. <https://doi.org/10.1093/idpl/ipaa016> (28 December 2020).
- Juarrero, A. (2023). *Context Changes Everything*. MIT Press. Kindle-Version.
- Kahneman, D. (2011). *Thinking: Fast and Slow*. Macmillan US.
- Kannengiesser, N., Lins, S., Dehling, T., Sunyaev, A. (2019). What Does Not Fit Can be Made to Fit! Trade-Offs in Distributed Ledger Technology Designs. *Conference Paper, Hawaii International Conference on System Sciences*, https://www.researchgate.net/publication/327793246_What_Does_Not_Fit_Can_be_Made_to_Fit_Trade-Offs_in_Distributed_Ledger_Technology_Designs
- Kauffman, S. A. (1996). *Investigations: The nature of autonomous agents and the worlds they mutually create*. SFI working paper # 96-08-072. Santa Fe Institute
- Koch, C. (2004). *The Quest for Consciousness: A Neurobiological Approach*. Roberts & Company.
- Kurzweil, R. (2005). *The Singularity Is Near: When Humans Transcend Biology*. Penguin Us.
- Kurzweil, R. (2012). *How to Create a Mind: The Secret of Human Thought Revealed*. Viking Adult.
- Maple, C., Epiphaniou, G., Gurukumar, N. K. (2021). *Facets of Trustworthiness in Digital Identity Systems*. The Alan Turing Institute.
- Mill, J. S. (1848). *Principles of Political Economy*. Parker.
- Moyer, M. (2016). *Slicing Pie Handbook European Union Edition*. Lake Shark Ventures.
- Noack, A. (2019). Blockchain, DLT, DAO, Smart Contracts - Chaos oder Chance? Entscheidungsansätze für eine vielversprechende Technologie. Discussion Paper, Hochschule Stralsund.

Ruff, T. (2018). 7 Myths of self-sovereign identity. <https://medium.com/evernym/7-myths-of-self-sovereign-identity-67aea7416b1>,

Sapolsky, R. (2017). *Behave: The Biology of Humans at Our Best and Worst*. Penguin Press.

Snowden, D., Rancati, A. (2021). *Managing complexity (and chaos) in times of crisis*, EU Science Hub and Cynefin.

Snowden, D. (2023). Transcript of Episode #184 of the Jim Rutt Show. <https://www.jimruttshow.com/dave-snowden-2/>

Thaler, R., Sunstein, C. (2008). *Nudge: Improving Decisions About Health, Wealth and Happiness*. Yale University Press.

Tononi, G. (2012). *Phi: A voyage from the brain to the soul*. Pantheon.

Trust over IP Foundation (2021). *Design Principles for the Trust over IP Stack*. Version 1.0. <https://trustoverip.org/wp-content/uploads/Design-Principles-for-the-ToIP-Stack-V1.0-2022-01-17.pdf>

Wahlstrom, K. Fairweather, N. B., & Ashman, H. (2017), *Brain-Computer Interfaces and Privacy: Method and interim findings*. Paper presented at the ETHICOMP, Turin, Italy.

Zaman, S., Khandaker, M.R., Khan, R.T., Tariq, F., & Wong, K. (2021). Thinking Out of the Blocks: Holochain for Distributed Security in IoT Healthcare. *IEEE Access*, 10, 37064-37081.

Zuboff, Sh. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of PowerCapitalism*. Profile Books.

This is version 1.0 of the Metaself Whitepaper, published on 26.6.23. Comments welcome:

axel@metaself.ai

<https://www.metaself.ai/>