INTELLIGENT COOPERATION

FORESIGHT INSTITUTE

Aaron King

Keynote Speakers

Gillian Hadfield
Kate Sills
David Brin
Daniel Ellsberg
Richard Craib
Jim Epstein
Primavera De Filippi
Alex Tabarrok
Balaji Srinivasan
Peter Norvig
Anders Sandberg

Brewster Kahle
Robin Hanson
Anthony Aguirre
Paul Gebheim
Thomas Pfeiffer
Chris Hibbert
Martin Koeppelmann
Vernon Smith
Mark Miller
David Friedman
Zhu Xiaohu

Matan Field
Jazear Brooks
Patrick Joyce
Esteban Ordano
Tyler Golato
Zooko Wilcox
Howard Wu
Andrew McAfee
Christine Peterson
James Bennett
Randy Farmer

Chip Morningstar
Meng Weng
Federico Ast
Glen Weyl
Marc Stiegler
Arthur Breitman
Audrey Tang
Tyler Cowen
David Krakauer
Gernot Heiser
Christine Lemmer-Webber

Table of Contents

Introduction

Foresight Intelligent Cooperation Group

A group of researchers, engineers, and entrepreneurs in computer science, ML, cryptocommerce, and related fields who leverage those technologies to improve cooperation across humans and ultimately Artificial Intelligences. Keynotes roughly follow an unpublished book draft that proposes Intelligent Voluntary Intelligent Voluntary Cooperation as a path for different intelligences to peacefully pursue a diversity of goals while reducing potential conflicts.

We explore:
• What is good about systems of voluntary cooperation.
• Which technologies strengthen them.
• Which factors threaten them.
• How to extend them to other intelligences.

This report gives an overview of our 2021 recorded seminars, including a favorite slide, and a link to the full written summary and recording for those who wish to learn more.

Foresight Institute

Foresight Institute is a 30+ year-strong San Francisco-based institute to advance crucial science and technology for the long-term flourishing of life. We believe that, in addition to directly addressing existential risks, one relatively neglected area for impact is to directly support differential technology development in areas that make great futures more likely. We focus on working groups to advance:

- Molecular Machines for atomically precise control of matter
- Biotech & Health Extension to reverse aging and improve cognition
- Computer Science to secure decentralized human AI cooperation
- Existential Hope to catalyze beautiful futures

I invite you to apply to join, support our work, or contact me with feedback, questions, and suggestions.

Thank you for your interest in our work,

Allison Duettmann
President, Foresight Institute
a@foresight.org

Intelligent cooperation intro

Mark Miller
Christine Peterson **May 18, 2021**

Summary

The private space and longevity industries are taking off, AI is progressing, and the developing world is getting connected to the internet bringing education and employment. Environmental problems are getting tackled, robotics and drone technology are maturing. Neuroscience is poised for a period of major advancement and our command of biology is more complete every day.

Global violence has decreased and cooperation has ramped up and allowed for amazing growth, but we need better infrastructure to support these global complex cooperative arrangements. Centralizing and decentralizing forces are in play in the world that aid and impede agreements. While nation-states centralize power, the Internet makes geographic governance less and less relevant to cyber activities.

Blockchain technology allows us to experiment with digital governance at a much lower cost than physical governance experiments (which often cost many lives). As in biological evolution, most governance experiments in the digital realm will fail, but there will be some gems that survive, thrive, and become important tools for cooperative society moving forward.

It's very hard for individuals to hold any private information, most of which is entrusted to companies with terrible information security ripe for predation by independent and state-affiliated threat actors.

The software infrastructure of the world is not just insecure, it's insecurable. We know HOW to build secure systems, we have since the 1970s, but markets aren't rewarding security, only functionality. We will soon have automated AI hacking software in the wild constantly stress-testing all systems to a degree that hasn't even been approached yet. Because of the Darwinism of the cryptosphere, the systems that survive will be the most secure.

To get out in front of the issue and understand the radical nature of change on the horizon of information security, we have reached out to many cryptography experts worldwide.

Opportunities

On intelligizing cooperation: how AI is being pursued in the world is quite different than the normal framing of AI risk. We're building many intelligent systems for doing very particular tasks that are embedded in a society that is coevolving with these systems to cooperate with them. As these things become more and more intelligent, they'll find themselves in the same situation we do: attempting to accomplish our goals in an environment that is largely shaped by other agents working toward their own goals.

Access the full summary and recording

Value drift & futarchy: Vote values but bet beliefs

Robin Hanson, George Mason University **February 8, 2021**

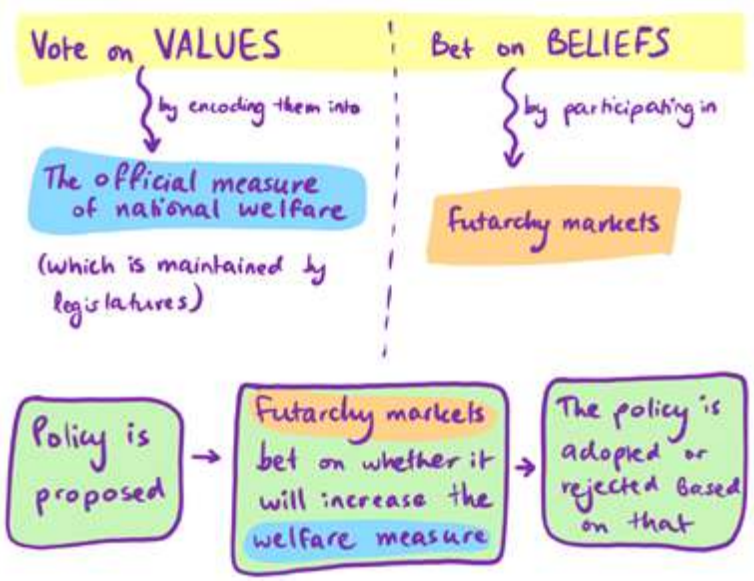
Summary

Ordinary decision theory separates values and beliefs about facts as the two components of decisions. The latter should change as you learn about the world, but your values are assumed to be constant, the anchor for decision-making. So in these contexts, the idea of "value drift" seems scary: they are supposed to be anchors!
Often when folks talk about values they aren't talking about these core anchors, they're talking about something more fluid that may be faction-based, or about symbols we're coordinating around.

It's commonly accepted that your human descendants will have values that differ from their ancestors: children branch out and relate to the world differently, and values change over the course of a lifetime. This is value drift. Sharing all the same values is not how we tend to keep the peace: instead, we use norms and laws.

If you think of governance as "making good decisions", you could say that what we need to do to make good decisions is to set an objective, aggregate information about the consequences of possible actions towards that objective, then take the actions which would seem to best accomplish our goal. From that point of view, decision markets and prediction markets may work well for this. We can set objectives based on our values, then open markets for folks to predict the outcomes of potential actions.

Governance is actually a way in which we struggle for power some win, and the rest of us submit and pretend that the winners were right all along. We are very indulgent of people we give power to, and let them get away with a lot. It's an age-old observation that "we have kings, but they over there have tyrants", usually to justify invading them over there.



Access the full summary and recording

Paretotropism

Marc Miller, Agoric

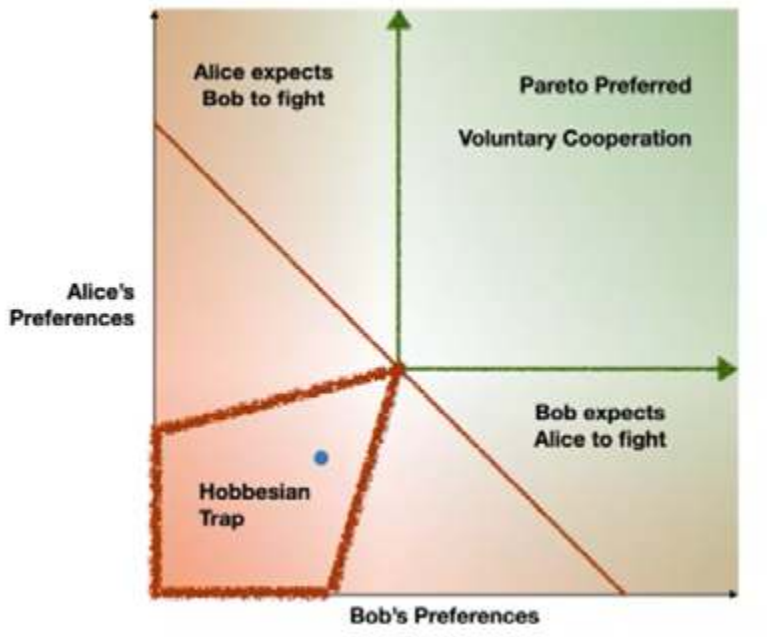
February 8, 2021

Summary

We are good at getting people out of poverty. As Steven Pinker says: We've been doing something right, it would be good to know what it is.

When we look at cooperative decision making, outcomes can be defined by the preferences of the entities involved, shown below. In the "Pareto Preferred" region, where voluntary cooperation resides, someone is better off and nobody is worse off, so there's no reason for anyone to not go along with the plan. In the upper left and lower right, someone is better off than the other, so each might expect the other to fight for a different outcome. This mutual expectation to fight creates the Hobbesian trap: both are preemptively expecting a fight, creating First Strike Instability.

Technologies like escrow, reputation, contracts, and the rule of law provided bridges to cross into Pareto Preferred space. Now blockchain technologies have brought in an entirely new toolset to work on these problems. Tropism is a tendency to grow a certain way: plants have phototropism in that they grow towards the light. Through our various tendencies and technologies for cooperation, we seem to have a sort of paretotropism as a civilization.



Access the full summary and recording

A simple model of grabby aliens

Robin Hanson, George Mason University **February 9, 2021**

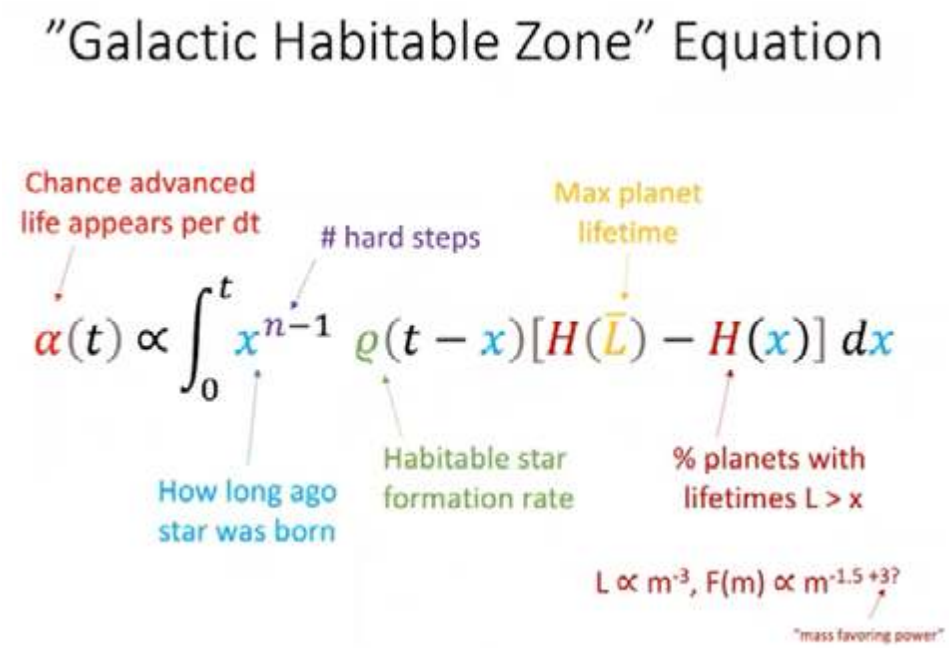
Summary

Most of us have heard of what Robin calls “quiet” aliens, or isolated alien civilizations that don’t much expand past their home planets or systems, run their course, and eventually end. Robin discusses loud aliens - loud aliens don’t die off, they grow and expand out into the universe.

Three parameters can be used to model these aliens: the spawn rate of such grabby populations, the expansion rate of these groups once they start spreading out into the universe, and the power law governing the chance that any particular group evolves to start spreading in the first place. The spawn rate is based on the assumption that human civilization becomes “grabby” in the next 10 million years, meaning our origin is a random sample of these spawns. The power law is determined by the“hard steps” theory of evolution. This theory states that in the history of life on Earth, there was a series of “hard steps” various forms of life passed through to get us where we are today.

The first step from the time Earth became first habitable to the first basic life was about 0.4 billion years. The final duration, between now and when the Earth will leave the habitable zone (at which point we better have started spreading) is about 1.2 billion years. Due to the math of power laws, these two should have been drawn from the same distribution, and so plugging them into the model shows that 3-12 “hard steps” have happened on Earth (assuming that no other steps were taken we don’t know about before Earth).

Using this and other information from his presentation, the basic equation below was derived. It seems from this model that humanity is early in the history of the universe… why’s that? If we were not early, and these sorts of “grabby aliens” exist already, then we would be seeing evidence of them.



Access the full summary and recording

Network state

Balaji Srinivasan, Andreessen Horowitz February 21, 2021

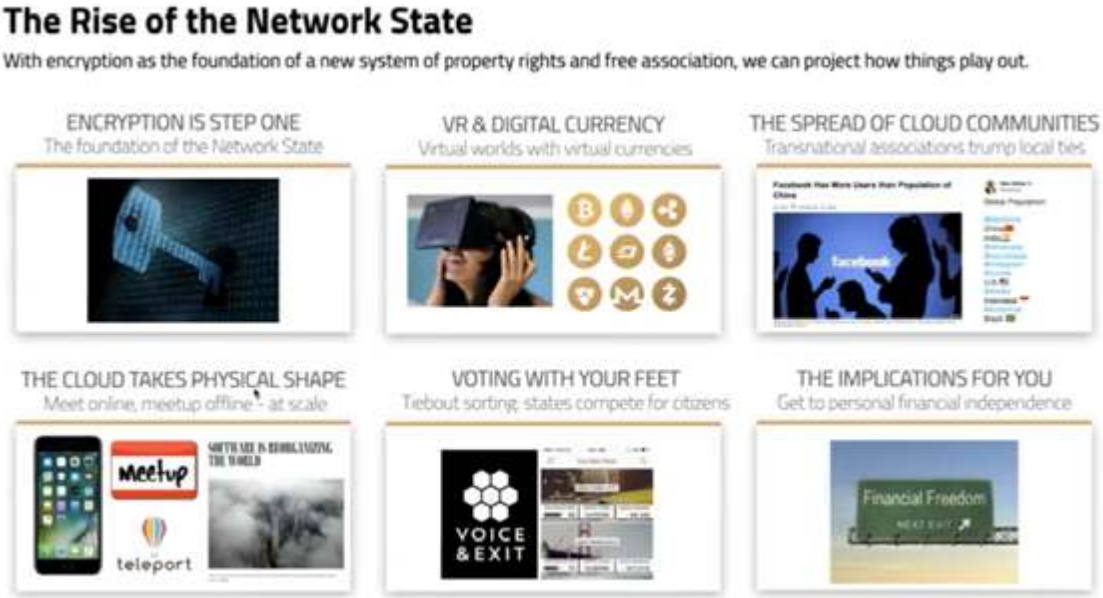
Summary

Software isn't done eating the world. How do you combat the inevitable takeover? Start your own country. Digital currency is part of a fundamental shift in human organization – from shared geography to shared ideas. The coming "network states" will organize by shared belief: geography is secondary, shifting as digitally connected citizens vote with their feet and are most likely distributed.

We are already seeing the rise of decentralized collectives like the Wallstreetbets community, internet-native groups affecting markets and policy. We've seen the expansion of tools and information systems in support of location-independent digital nomads like teleport.org and nomadlist. Add in explicit leaders to organize and catalyze collective action and governance systems for allocation and employment of assets and information and where you end up is "Crowdchoice": groups of people who get together and aggregate their preferences to facilitate collective bargaining with existing governments.

As technology evolves, the tools of society oscillate between centralized and decentralized forces. Very few institutions that predated the internet will survive the internet and its decentralization effect.

Pure network institutions are forming: Facebook behaves more and more like an online country, while the Bitcoin network and other cryptocurrencies forge distributed bonds between people with economic incentives driving cooperation. As the large nation-states of the present continue to lose power, we'll see the formation of true Network-States. We see just the beginning of this trend in two very different case studies: the technoauthoritarian Chinese surveillance state and the digital-forward cyberpolity of Estonia.



Access the full summary and recording

Civilizational Progress: Accelerating Its Drivers

Andrew McAfee, MIT **March 5, 2021**

Summary

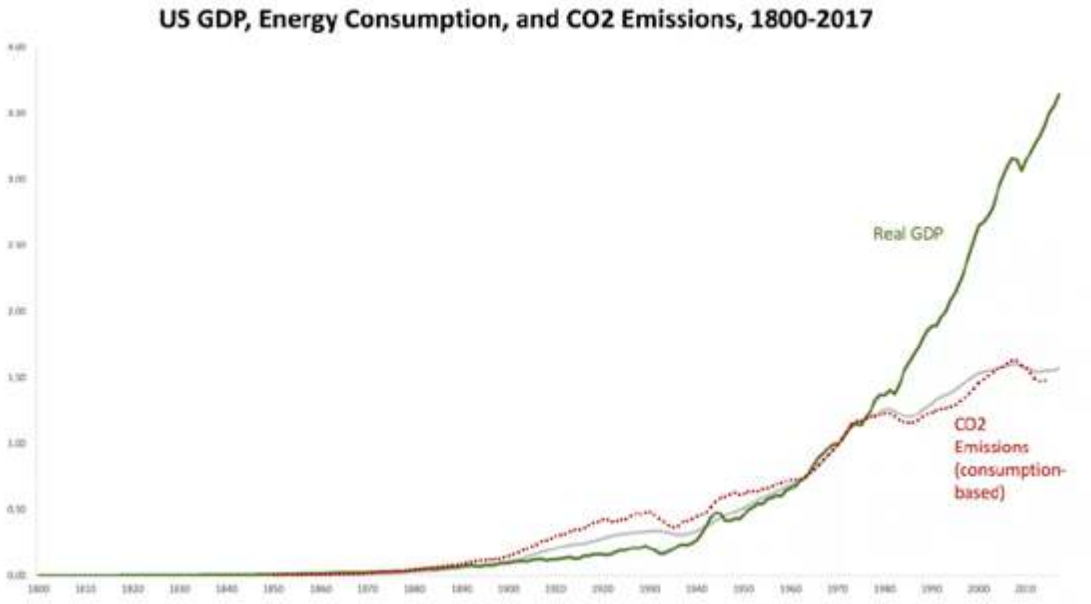
Andrew has been exploring the phenomena of a natural and unexpected collective action by the human race to reduce our burden on the planet without taking conscious directed action to reduce consumption.

The first big problem - we can't feed everyone. This idea goes back to Malthus and the concept of the population bomb. He believed people were not capable of generating resources at the rate of population growth. Graphing out the years of 1200-1700 in the context of standard of living shows that Malthus was historically quite accurate. However, after the industrial revolution, both standard of living and population exploded without any apparent tradeoff. Deaths by famine dropped significantly. It appears the first big problem has been solved by the industrial revolution.

A second problem then emerged in the public consciousness which was essentially opposite from Malthus's general hypothesis - people began to believe that the earth would run out of resources because we had excessive rates of resource generation. However, in the latter half of the 1900s a bizarre trend began to emerge - our economic output grew but our resource consumption began to level off. Technological advancement had repeatedly collapsed multiple separate bulky devices into a single small streamlined device, leading to an increase in productivity with a decrease in resource cost.

The third problem is pollution and species extinction. This is absolutely happening - or at least it was, until roughly 1970 when pollution levels started to decline while production still increased. Public awareness and government action have solved pollution via things like the Clean Air Act and Clean Water Act.

Can we solve our big problems while continuing to grow? Yes we can.



Access the full summary and recording

Classical Theory of Price Discovery in Markets

Vernon Smith, Nobel Prize in Economics March 5, 2021

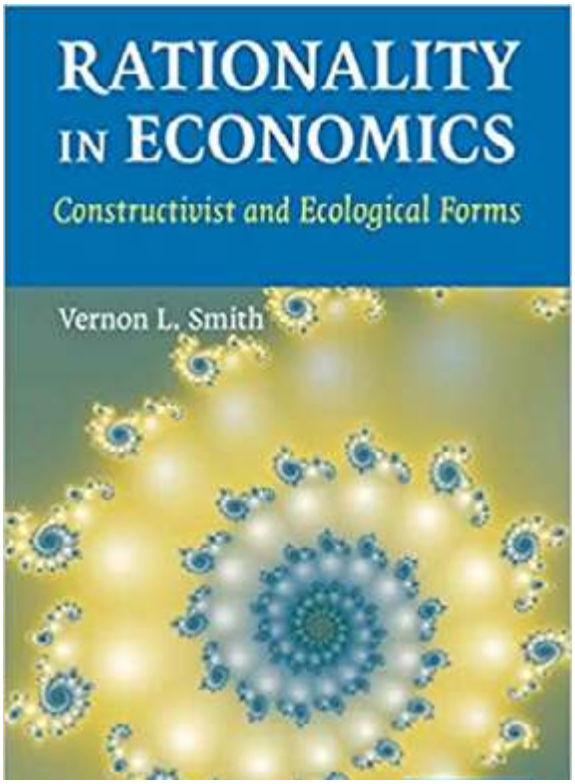
Summary

A co-author wrote to Vernon regarding his experimental markets concept and said that Vernon had actually rediscovered classical economics. This launched a project to examine the thread of intellectual development from Adam Smith through his French, English, and Italian followers concerning their perspective on how people discover prices in markets.

There are two main ways of understanding the perception of price - quantity as a function of price, or price as a function of quantity. Both of these are wrong. Buyers and sellers have reservation values that exist outside of these systems. Prices are discovered in markets - they do not exist a priori.

If sellers bring too little to market, competition takes place and buyers bid up the price. If sellers bring too much to market, competition takes place among sellers to lower the price. This illustrates the dynamic law of supply and demand. Something to consider here is the nature of discrete supply/demand vs. continuous supply/demand.

Resale value throws a wrench into the math of supply and demand. Two levels of demand emerge - a demand for final consumable good vs. demand of a good for the purpose of resale.



Access the full summary and recording

Stubborn Attachments

Tyler Cowen, George Mason University

March 9, 2021

Summary

Tyler worked on his book Stubborn Attachments over the course of 20 years. His graduate economics work influenced the book, focusing on welfare economics. How do we know that one policy is better than another? Economists typically invoke cost-benefit analysis, but this doesn't create a grand picture of social systems, it more lends itself to static issues.

Piecing all of this together, he worked on developing an argument where the primary good at the social level is to maximize what he calls the "rate of sustained economic growth."

As a group, we can't agree on how valuable all the different parts of life and the world are to each of us. Maybe we could agree that a society as a whole, say current day America, is a better place than current day Albania or Congo. But what would we be agreeing on? Perhaps that a much wealthier society is probably going to be a better place to live. If your view is like his that the future is as important as the present, this implies that economic growth is how we get to versions of the future in which there are much better places to live.

There are absolute human rights that should be respected, and these should constrain what we can do to maximize economic growth. Not a complex ontology of rights that a consultant might come up with, but more along the lines of "don't torture people", straightforward decency.

The book considers wealth not just in terms of GDP. Leisure time and other such goods don't show up in GDP. When he talks about maximizing economic growth, he takes into account non-market goods. Small differences in growth rates make huge differences over time.

This is a moral framework for thinking about politics: it puts productivity and economic growth and sustainability at the center of our thought.

Access the full summary and recording

Tools for Openness: Asia and Beyond

Audrey Tang, Taiwan Digital Minister — March 19, 2021

Summary

What idea(s) have you found particularly useful to your efforts in Taiwan?

Radical Transparency! All meetings with Audrey undergo a 10 day co-editing period and then are released publicly. This serves to make conversation participants mindful of what they say because they know people in future generations will be watching. Private meetings allow for people to get away with short-term thinking and proposing policy that is not focused on the long-term thriving of all. "I think this intergenerational solidarity is built on top of the idea of radical transparency and us contributing to the commons. It's not just the what of policy-making, but the why and how of policy-making." The open release of policy-making meetings also serves to get young people, immigrants, and every stakeholder that is not traditionally involved in the processes of government.

What technologies are you looking at now?

The idea of democracy as a technology is a key principle. The fact that you can increase the "bitrate of democracy" by using innovation cycles to develop it without needing to destroy the old systems. What's stopping more people from realizing this: the false idea that career public servants are resistant to change. This isn't true: they are open to change if it will save time and reduce risks, which these kinds of tools can.

What is the number one global development or technology you're excited or worried about?

Excited: The idea of data coalitions and data collaboratives. International endeavors that have already helped tackle the pandemic and are now looking at climate change, the infodemic and other big problems
Worried: The tendency to attribute mistakes by democratic states to failures of democracies that lead to people making a case for authoritarian responses that step in to fix things.

Access the full summary and recording

Blockchain Governance

Arthur Breitman, Tezos **April 5, 2021**

Summary

What is governance? A set of rules that you follow when you have a resource that must be shared by multiple people and you don't have a good way to divide it up. A simple case is property rights: if you own a piece of fruit and I own a rock, we do what we want with our things and no governance is needed. But some things can't be privatized so simply: think of an apartment. You can make holes in your own wall, but there is an elevator and a heating system and such that must be governed about as a building.

Looking at early blockchain projects like bitcoin, it seems like we don't really need governance. It's a matter of simple property rights: you have a coin, I have a coin. But the security of the network is a shared resource that must be governed. Once all the coins are released, fees must pay for security and therefore the security of the network becomes a part of the commons, and therefore under threat of the Freerider problem. Some projects (like Zcash) have made developer rewards a part of the block rewards. This is governance.

However, once you've identified a good idea, how do you get the code moved into your protocol?

Theoretically, these projects are open source and anyone can run their own version, so you don't necessarily need governance. But cryptocurrencies are about the shared ledger, so the coins derive their value from network effects, and the winner will be the one that people use. In a fork scenario, the one that people use will be the winner of a beauty contest, not strictly the "best".

A lot of the stability of currencies comes from the fact that they are embedded in a web of contracts. In the face of a complex web of contractual dependencies, you can't have your currency ledgers forking and splitting.
We're starting to see this stability-enforcing dynamic emerge on Ethereum, with USDC becoming really baked into the web of contracts in the ecosystem. The more USDC is baked into the contractual landscape of Ethereum, the more sway Circle (the company that runs USDC) has over which Ethereum fork "wins" in a fork scenario: the one that you can use your USDC on. This is why having explicit governance mechanisms to make changes to your protocol are key to limit the implicit or unintended influence of other parties on shaping its path.

Access the full summary and recording

The Digital Path and Blockchain for Secure Title Registries

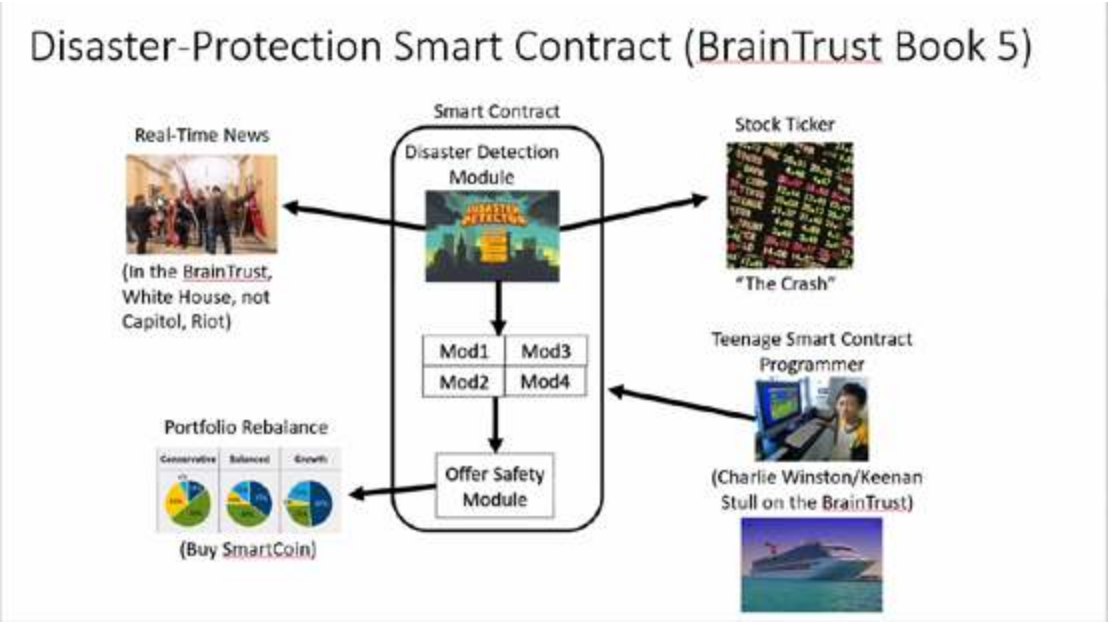
Marc Stiegler, Computer Scientist **April 5, 2021**

Summary

The Digital Path was inspired by "The Other Path", wherein Hernando DeSoto set out to bring the assets of third world villages into the formal economy so they could be used as collateral for wealth-generating loans. In attempting to do so, his efforts were constantly undercut by the corruption of local officials and governments. Mark Miller took the idea further, proposing a mechanism for the locals of these countries to deal with more trustworthy foreign institutions by creating video evidence of contractual agreements and ownership rights and sending this evidence to these institutions that would track, manage, and financialize the property for investment and wealth creation.

Part of the proposal was to use smart contracts to automate the tracking and transfer of assets to reduce trusted humans in the loop as much as possible. At the time of the proposal, the main stumbling blocks for the development of smart contracts were security concerns. When the Digital Path was being written, blockchains hadn't yet been invented. In the Brain Trust books, in an alternative future, the President of the United States expels all immigrant engineers from the country, who then create an off-shore floating citadel called the Brain Trust, where they build the smart contract and blockchain system that manages the world's finances.

An example of a use of the system in the books: communities in Benin use a cell phone to create video evidence of property rights and contracts, all tracked by the SmartCoin cryptocurrency. When human input is needed to resolve disputes, it goes to a mediator on the BrainTrust offshore, though most disputes are enforced by social pressure and handled locally.



Access the full summary and recording

Social Technology for a Political Economy of Increasing Returns

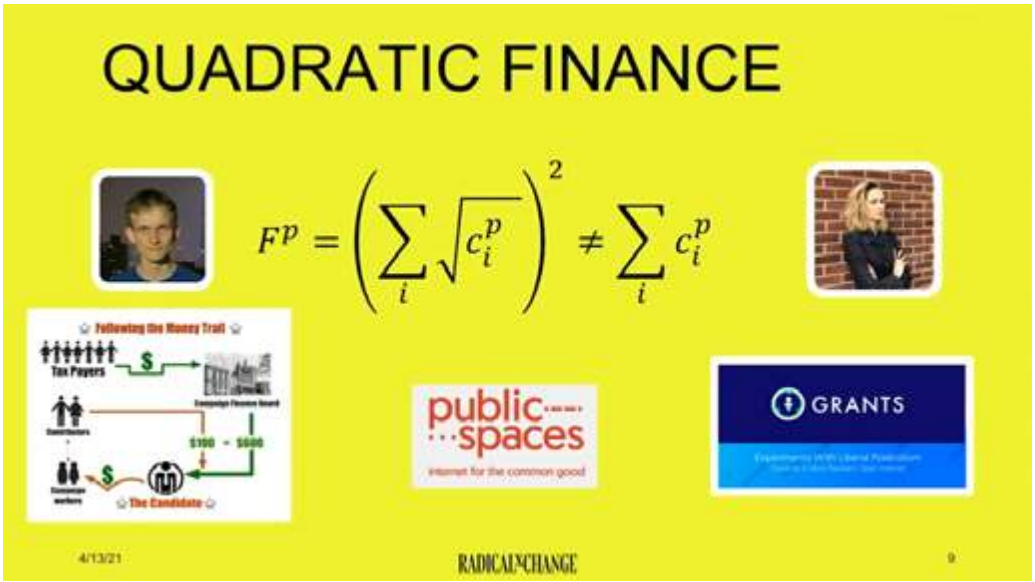
Glen Weyl, RadicalxChange **April 14, 2021**

Summary

Over the last few decades, communication technology has advanced astoundingly, moving us toward richer and richer representations and higher and higher fidelity communication. This same level of advancement should and can be achieved for our political and social technologies. There is an enormous amount of richness in our interpersonal lives, within our "Dunbar groups", that is not present for larger scale and anonymous social interactions. Let's change that.

Quadratic Funding is a mechanism for resource allocation developed by Radical Exchange, what they call an optimal mechanism for the provision of public goods. It's sort of like the idea of "matching funds" that's quite common in charitable giving, with some different considerations. The smaller a person's possible contribution, the less likely they are to actually contribute because their contribution doesn't matter much. Taking that into account, you can set up a system to weigh the contributions by size, reversed. Technically, you want the amount received to be the square of the sum of the square roots of each contribution.

If you can recycle the value accrued by the decreasing returns activities (through taxing perhaps) into the increasing returns activities then you get this self-fulfilling kind of superconduction.
But how can you do this without the tax impacting the economics negatively? Enter the SALSA: Self-Assessed Licenses Sold at Auction. Under this scheme, property owners name their price but are required to sell to any buyer who meets the price. There are all sorts of impediments to property being repurposed for new uses, in both private market contexts and political contexts. We have regular elections so as to reallocate "political property". What this sort of compelled sale idea, you can build guaranteed periodic reallocation into the economic system. Taiwan is using quadratic voting for many civic services and votes, including their Presidential Hackathon.



Access the full summary and recording

Dominant Assurance Contracts to Innovate Our Way Out of Market-failures

Alex Tabarrok, George Mason University **May 16, 2021**

Summary

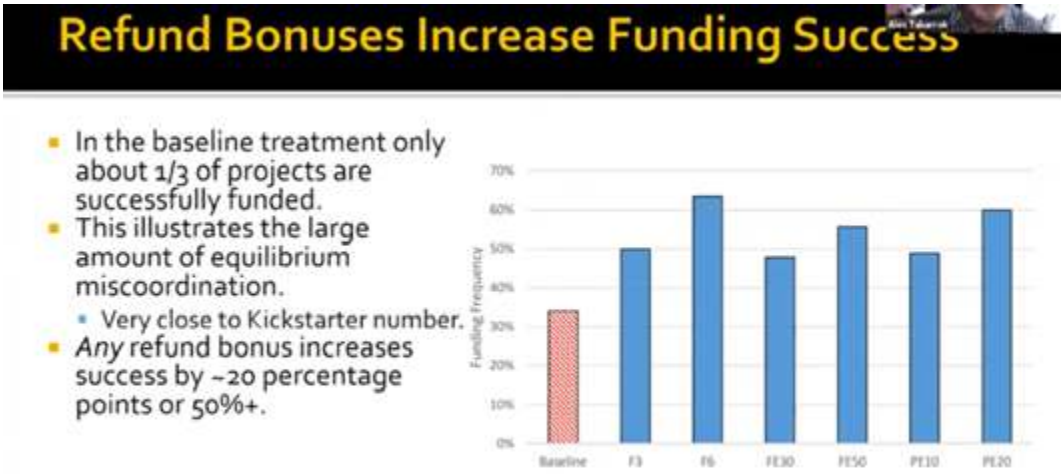
Public goods are a market challenge. There are many things that we all would like to see realized, but no individual has an adequate incentive to contribute to. The free rider problem creates incentives to not contribute to a public project. The assurance problem occurs when you contribute and others do not - you are saddled with liability while also not getting a benefit.

Assurance Contracts can be used to counter these problems. We pre-commit to take an action — such as giving x amount of $/BTC/other cryptocurrency to the longevity/journalism/open source project of our choice — if and only if enough others commit to doing the same to reach a critical threshold. Dominant assurance contracts have the added condition that if the funding benchmark isn't reached, the provider pays a prize to the pledgers. Pledging becomes a dominant strategy, or in Tabarrok's words, "a no-lose proposition—if enough people pledge you get the public good and if not enough pledge you get the prize."

An experiment was conducted to demonstrate the dominant assurance contract. In the experiment, refund bonuses increased the rate of success by 20-50%. It was concluded that these bonuses pay for themselves due to the increased success rate.

Early contributions are crucial for determining the success of a crowdfunded project. Creating early refund bonuses makes early contributors more pivotal - increasing the perception of importance leads to increased participation.

Using these strategies it should be possible for market forces to provide for public goods.



Access the full summary and recording

Split Contracts, Computational Law & Decentralized Arbitration

Chip Morningstar, Agoric
Meng Weng, Legalese
Federico Ast, Kleros

May 18, 2021

Summary

Mixed contracts: To come to an agreement around information sales and purchases, it was found that a mix of machine-automated and human interpretation was going to be necessary to capture the complexities of the arrangements. We see that some stuff can be monitored by machines (like delivery date) and some stuff cannot (like deliverable quality).

Structured Negotiation: Through Structured Negotiation, contract terms are arrived at through negotiation, with agreements and commitments captured along the way. Just as with traditional contracts, much of the bulk of the document will be concerned with what to do when things go wrong, up to and including renegotiation clauses.

Legalese: picked up funding from the Singaporean government to develop a DSL language for building legal contracts. Decades of research on the intersection of CS and law exist.

Kleros: A DAO that engages with third-party jurors to decide on a dispute that has been sent to the service. How do you make sure these anonymous jurors actually behave well? Having your purchased tokens staked creates some amount of skin in the game, so as not to lose your stake. When a juror votes for a resolution, if they don't vote with the majority of jurors then they lose their stake. Kleros acts as a court DAO for other DAOs.

Access the full summary and recording

Zero-knowledge-enabled Cooperation: Halo 2 & Aleo

Zooko Wilcox, ECC
Howard Wu, Aleo

May 19, 2021

Summary

Zero knowledge proofs were discovered in the 80s but were not practically efficient or really used for anything until Zcash came around in 2016 using the Groth16 method.

The Zcash team has been working on a new zero-knowledge proof system called Halo2 that is recursive, meaning that you can prove that you ran a verification and the verification came out true. With the recursive system, you can verify a computation and generate a proof that you ran the verifier, and generate a proof that you ran the verifier on the verification, and so on and so on. This let's you prove and verify arbitrarily large computations by chaining many verifications together.

There is more and more excitement and a growing number of use cases for decentralized applications, but they have their problems.

The scalability problem: the computational integrity of these systems is determined by direct execution, meaning each miner must re-execute every transaction. This constrains computation via limited running time, minimal stack size, and restrictive instruction sets.

The privacy problem: the core strength of currently decentralized application systems is also the primary weakness, namely that the history of all state transitions must be executed by all parties. This totally precludes the privacy of the users of these systems and creates opportunities for Miner-Extractable Value: the lack of privacy of transaction details allows for front-running and arbitrage attacks.

Zero-knowledge proofs can achieve privacy and help scale systems by enabling proof that some known computation was executed honestly, and for some private inputs.

Small proof (like an Orchard tx)

	Toxic Waste Mitigation	Relies on crypto assumptions	Proving time	Verifying time	Proof size
Groth16 (MNT4 / MNT6) 2013–2019	One time MPC Ceremony	AGM+pairings	good (1s)	good (10 ms)	good (100 B)
STARKs (Fractal) 2018–2021	Toxic-waste-free	Hash function	bad (~1s?)	good (10 ms)	bad (~150 KB)
Halo 2019–2021	Toxic-waste-free	Discrete Log	good (2s)	good (20 ms)	good (2 KB)

Recursive proof

	Toxic Waste Mitigation	Relies on crypto assumptions	Proving time	Verifying time	Proof size
Groth16 (MNT4 / MNT6) 2013–2019	One time MPC Ceremony	AGM+pairings	meh (10s)	good (10 ms)	good (100 B)
STARKs (Fractal) 2018–2021	Toxic-waste-free	Hash function	bad (~20m?)	good (10 ms)	bad (~2 MB)
Halo 2019–2021	Toxic-waste-free	Discrete Log	good (1s)	good (40 ms)	good (4 KB)

Access the full summary and recording

DAOs: DAOstack, Decentraland, SifChain, ResearchHub, VitaDAO

Matan Field **Esteban Ordano**
Jazear Brooks **Tyler Golato**
Patrick Joyce

May 24, 2021

Summary

A DAO is a Decentralized Autonomous Organization. A DAOstack creates tools and platforms to make it easier to create DAOs. Common is DAOstack's new platform for DAOs. It's much more targeted the mainstream with a mission of enabling coordination and collaboration of thousands and then millions around any joint interest, purpose, or project: whether that's a business, a town, a research program, or a nation.

Sifchain is a decentralized exchange company that is in the process of building a DAO to manage the exchange and create a community around it. They aim to achieve nation-state level wealth under management. They also want to create councils within the DAO: coding standards, commons cultivation, on-chain UBI, public funding of research. There will be a "Needs and Gives" layer: a barter layer for goods and services with a reputation system built in to track community engagement. Finally, they want to collect research on user goals and a "self-binding commitments" system to help folks meet their goals.

ResearchHub is a Reddit-style forum where academics can share research outputs, anyone can curate, and there's a pubic comment section. Academic papers are crawled from socials, and authors can claim their papers and see public commentary.

VitaDAO is an organization dedicated to funding research into human longevity. VitaDAO's mission is to extend healthspan and lifespan. The focus is to attract researchers, post docs, crypto engineers, and tokenomics experts to build a community around this achieving this goal.

Decentraland is a virtual world powered by Ethereum smart contracts that employs a DAO to govern many of its most important aspects. It's a system comprised of a CDN network to distribute graphical content, a P2P protocol for in-world communication and avatar location, a Scripting system for landowners to inject rich interactions into the world, and a marketplace for trading in digital land and other goods in the world.

Access the full summary and recording

Prediction & Replication Markets, Augur, Metaculus, Gnosis, Oracle Problems, Beauty Contests

Robin Hanson **Thomas Pfeiffer**
Anthony Aguirre **Chris Hibbert**
Paul Gebheim **Martin Koeppelmann**

June 1, 2021

Summary

Prediction markets - We frame argumentation in general as prediction. Many arguments can be grounded. Let's incentivize predictions transparently. Further, we want these predictions people are making to be integrated into a consensus. A market bet is a transparent incentive based on a prediction, and the market is aggregating those predictions so others can see.

Augur - The Oracle Problem: we can create markets, and we can engage folks to predict or bet, but how do we resolve those markets to make sure results reflect the real world? Augur is attempting to solve this problem by creating a system called the Augur Oracle. People can report the outcome of a market and there is a set of bonds placed on disputing the outcomes. If outcomes aren't clear, multiple "universes" are created where each outcome occurred, and lets people bet on the versions of the world that they believe are the case.

Replication Markets is a bundle of projects that started around 2010, triggered by the replication crisis in science. This motivated a lot of fields to analyze this crisis, predominately in the behavioral sciences, to run large-scale replication studies in which many studies were collected and reproduced to see if they replicated. The folks behind Replication Markets were already working on prediction markets at the time, and these replication studies provided an opportunity to test whether prediction markets might be help determine whether a particular study might replicate.

When there is a question about an obscure subject with a surprisingly high probability of coming to account, that's a signal that somebody might want to take some action. The Keynesian Beauty Contest is when there are lots of different opinions about what had actually happened. For laying off risk or making insurance, in unusual situations it matters a lot for the decision criteria to be very clear and concrete.

Access the full summary and recording

A Peaceful Transition into Cryptocommerce?

Jim Epstein, Reason
Primavera De Filippi, Harvard
Brewster Kahle, Internet Archive **June 7, 2021**

Summary

There are two theories about how change occurs. One - The world changes by building new protected spheres outside the reach of governments, like Ayn Rand's famous "Galt's Gulch". Two - The world changes by exploiting gray areas in regulations and law, like how much of the internet was built. Most startups didn't know if what they were doing was legal, but they did it and it led to lots of change.

In regard to cryptocurrency and cryptocommerce, there's a case for theory 1: Bitcoin and cryptocurrencies serve as technologies of resistance with built-in real life demand from those looking to get around governments. The problem with the law-abiding projects is that they have to court customers like any other app. The projects that route around law and regulations have their demand created for them by the politicians and bureaucrats that create those laws and regulations.

You can't really escape from regulatory capture, but there is a notion of unregulatability. This may have been true when the internet was decentralized in its early days, but now we have highly centralized, easy to regulate systems. We're already seeing centralized choke points develop in cryptocommerce in the form of mining pools and custodial wallets. You can regulate cryptocurrency protocol changes because so many people rely on custodial wallets and centralized exchanges. Lex Cryptographica takes a specific stance: rule OF code, instead of rule BY code, managed in a decentralized manner to reduce the power of the creators and operators.

Lessig is awesome and "code is law" is a good idea, but the code we have for the web is simple and doesn't protect a lot of the activities and behavior we'd like it to. The goal is a peer-to-peer backend for the web. Can we upgrade the web smoothly without needing to introduce new browser tech and such to replace the backend? We should also make it really easy for people to sell things on this new web. The decentralized web should be reliable, secure, private, and with the native commerce and file sharing tools we need.

Access the full summary and recording

Staking, Signals, and Other Techniques for Intelligence Coordination

Richard Craib, NumerAI

June 25, 2021

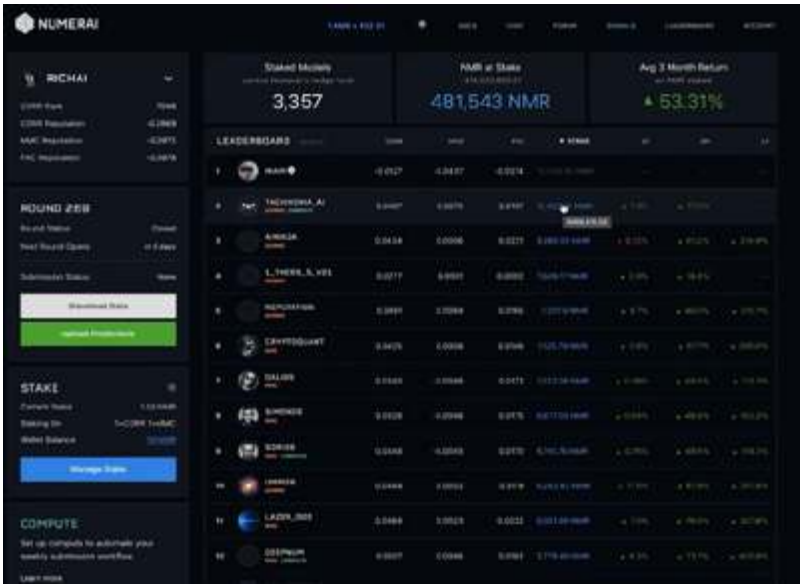
Summary

NumerAI is the first hedge fund that gives away all its data, so its users can build models on this data and contribute information back in to improve the hedge fund. Publishing the data publicly creates tension: with the public data, the community can use that data to improve the hedge fund model for everyone, but non-teamplayers could also take the data and start their own fund. To mitigate this, the data is obfuscated: there are signal identifiers but not descriptions (like “the P/E ratio of a stock” or which stock is which). Users are submitting predictive models built on this data to earn for the hedge fund while getting paid out rewards for models that do well.

If you had a site like NumerAI for building backtests, you’d have a problem: it’s easy to build a good backtest but very hard to know if they will hold on to live data. In a machine learning situation like detecting faces, the machine is always fitting to the data (which is static) and will nearly always converge, but in finance the future (and the data) is always changing.

But model builders have a sense of whether they are over-fitting on certain signals and can build models that compete on the same data by predicting the future. How do we surface that information?

One thing that happened; when we started NumerAI we were sibyl attacked. Thousands of accounts were created and they swamped the signal market with random signals hoping they’d get lucky. In response, a cryptocurrency and staking system was introduced to align incentives and prevent trolls: users staked their NMR tokens against their models and could have their stake slashed if the model does not do well. What’s become clear is that the models with the most stake tend to be the best performing, and the aggregate of the models that are not slashed and perform best are the best market signals.



Access the full summary and recording

Nuclear Risks: Doomsday (Still) Hiding in Plain Sight

Daniel Ellsberg, author of Doomsday Machine July 12, 2021

Summary

In school, Ellsberg heard about a hypothetical: a bomb 1000x more powerful than the biggest at the time. The Manhattan Project was already ongoing but was not known to anyone (never even leaked to the Germans, amazingly). He was asked: would this be good for humanity?

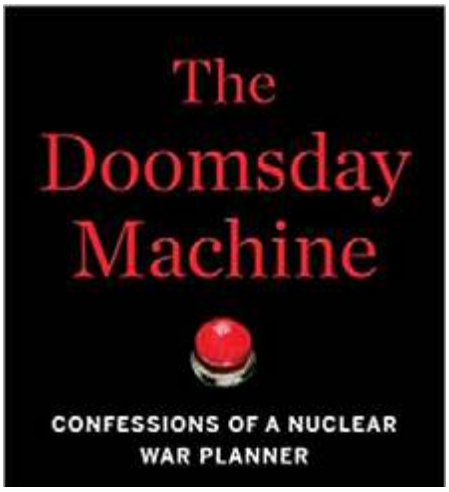
We've had 70 years of no nuclear war since the first drop, and the same issues are raised today about nuclear weapons as well as new technologies: Genetic engineering, climate engineering, artificial intelligence.

The primary concern of pacific control: "go" order would go out correctly for a strike against China, the only target in range at the time. If Russia was destroyed, China had to go to ensure they wouldn't be the successor state. The US plans at the time called for no limited war with Russia: must be all-out and we had to get the first drop. Any military engagement above the platoon with Russia would call for every city in Russia and China to be launched against. Ellsberg got interested in this question: "might these planes be launched without presidential order?"

It's always been false that the president has sole control over nuclear weapons. There must be redundancy in case the president was taken out of course. At the time, communications were interrupted between DC and the Pacific command and between the Pacific Command and the fleet all the time. False assumptions of an attack could have led to a reasonable launch.

The whole philosophy was (and is): the US will be better off striking first rather than second. We're poised on a hair-trigger.

Current war games suggest the Chinese can usually beat us in Taiwan. War with China will probably result in nuclear war since we still have a president who refuses to confirm a "no first use" condition for nuclear weapons (unlike China).



Access the full summary and recording

Transparent Society & Sousveillance

David Brin, author of The Transparent Society August 19, 2021

Summary

Positive Sum Thinking is the idea that underlies the attempts at restructuring the world from a pyramidal society with oligarchs at the top to a diamond shaped society with a robust middle class and merit based upward mobility. Each attempt (Periclean Athens, Da Vinci's Florence, etc) was met with an intense immune response from the oligarchs at the top. This group discusses one of Fermi's great filters: the idea of "small kills all", or technology allowing relatively small groups to cause massive damage to society.

Chinese intellectuals say the "small kills all" idea is why the pyramid-shaped society is necessary, to catch and prevent these problems from occurring. If you look across history, these societies are stable and work with human reproductive patterns, but are always "stupid". The idea is that you need centralized control or you won't catch the anomalies that kill everyone. China's system is smarter than other pyramidal systems, but still inherently stupid.

What you have as an alternative to the pyramid model is Reciprocal Accountability, the basic notion that underlies the enlightenment experiment.

The great benefit of the Enlightenment is not democracy and freedom: those are tools we use to get to the real thing, reciprocal accountability. The key is to replace surveillance with sousveillance: transparency aimed upward at the elites, stripping them naked and enforcing accountability. If we can do that we will be immune to big brother, but we will be at risk of the second tier problem: social self-oppression.

Replacing oligarchy with a 51% majority that openly and honestly votes for suppression of the 49% isn't an improvement.

If you have a culture that detects everything, but the least liked behavior is nosiness, then those who say "mind your own business" will be empowered, and those who say "you're not conforming!" are pounced on as bullies. This solves the secondary failure mode.

Access the full summary and recording

NFTs and Engineering Property Rights

Kate Sills, Agoric **August 19, 2021**

Summary

In the US legal system, a distinction is made between contract law and property law. Contract law is extremely permissive: parties can make contracts about pretty much anything that concerns just those parties.

Example: Michael Jordan notably had a “Love of the Game” clause in his NBA contract that allowed him to play basketball outside of the NBA in the off-season, something usually prohibited in player contracts for fear of injury.

Property law is much more limited, in large part because property law affects large undefined swaths of people, often everyone not specifically mentioned in the contract.

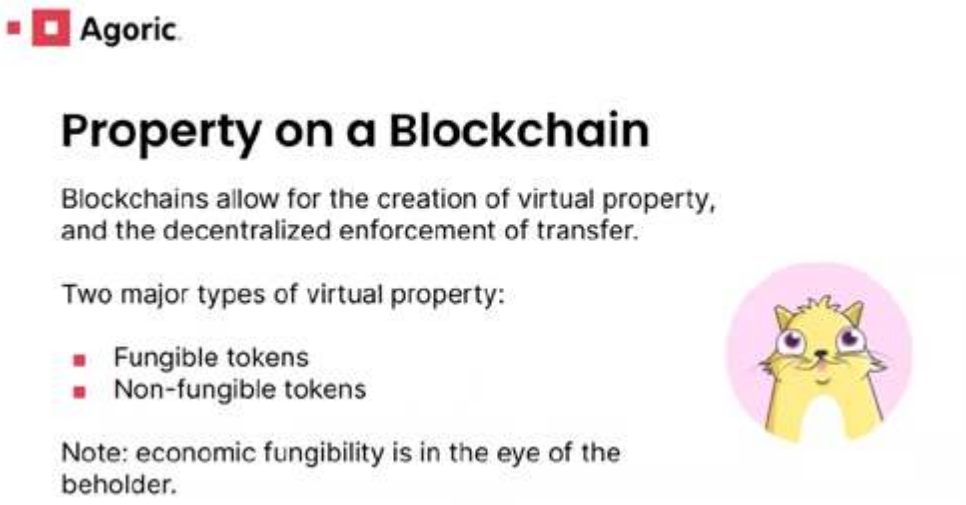
Example: If I made a deal with Michael Jordan that he can play basketball on my property even after I sell it, that’s not enforceable in court because the contract doesn’t survive the transfer of ownership.

New forms of property or new ways of dealing with property create costs for people who didn’t even sign that contract.

Say there is a market for trading cards and Alice is selling Bob a card. Normally, Bob learns the price and pays or not. This doesn’t affect any other buyers looking at any other cards. If Alice decides to get fancy, she may try and sell just the rights to use the card on Monday to Bob. Bob may agree to this, but now if Susan is looking at other cards she must ensure she’s buying the card itself and not just its use, raising her search costs.

This interesting development introduces a problem: The Tragedy of the Anti-commons

With smart contracts, we can design tokens to prevent these sorts of situations using mechanisms like tokens that expire, multisignature access, etc.



Access the full summary and recording

Incomplete Contracts & AI Alignment

Gillian Hadfield, University of Toronto **September 9, 2021**

Summary

Hadfield speaks about The Big Shift: from “How do we embed a particular set of values into artificially intelligent agents?” to “How do we embed artificially intelligent agents into our human systems that work?”

This has been long studied in economics as the theory of incentives, and at some point economists claimed to have a science of incentives: economic theory was that you could get the agent to do what you want by writing a complete contingent contract that deals with all relevant bits of the world. However, terms like “Reasonable”, “responsible”, “best effort”, etc. abound in legal contracts. It’s not a realm of total crispness.

How do we design contracts KNOWING they will be incomplete?

The primary approach to working with AIs is through reinforcement learning: specify a set of rewards and penalties based on the environment the agent is expected to be operating in, then let the agent move through the environment learning how to navigate to maximize its reward and minimize its penalty.

The human you hire with a contract will avoid all obstacles intuitively. Why is this? According to Adam Smith, a human agent goes through a process when it sees an obstacle: “what would an impartial spectator think if I just knocked it over?” People have the whole of human norms and societal expectations in their head when they’re acting in the world, reasoning hypothetically about potential impacts of decisions.

Key message to AI builders: game theory is all good and well, but we need a normative infrastructure like humans have if we really want aligned artificial agents.

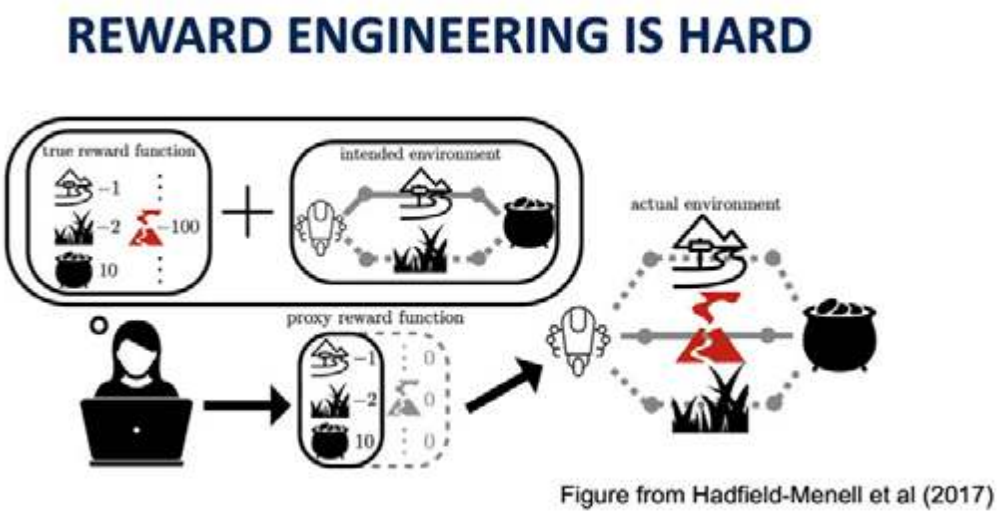


Figure from Hadfield-Menell et al (2017)

Access the full summary and recording

Collective computing: Learning from nature

David Krakauer, Santa Fe Institute September 28, 2021

Summary

Historically, there have been many attempts to outsource our governance and legal systems to some kind of divine power. The most recent manifestation is the desire to have AI do the thinking for us. The alternative perspective is to have humans stay in the driver seat but have our organizational systems enhanced with machine learning and AI.

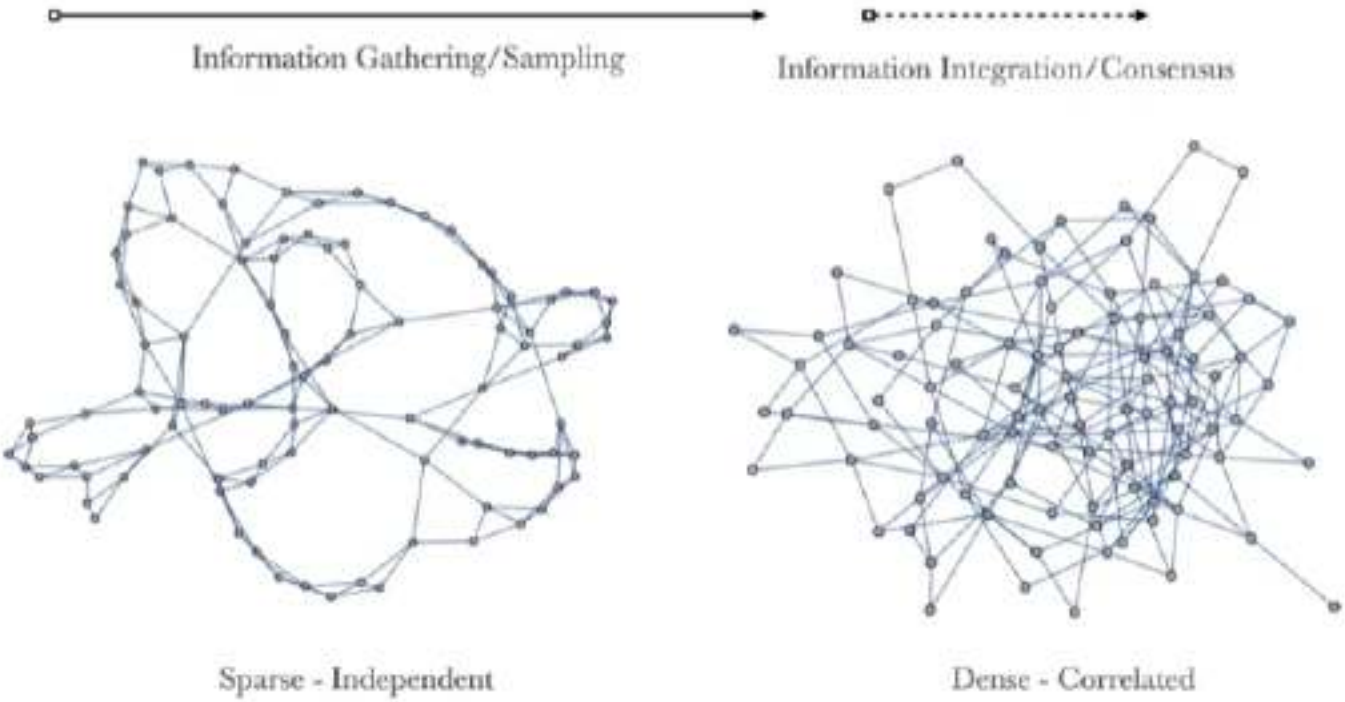
David Krakauer explores how nature solves the problems of collective organization at different scales - molecules, cells, organisms, and societies. Mechanisms tend to repeat themselves at various scales, and the logic of how the brain works can be applied to societies.

He also covers social circuits and how they can be used to predict group behavior.

Opportunities

Will continue to work on natural distributed intelligent systems. Over the past 6 months have been working on bringing out a book on what Santa Fe Institute can contribute to the response to COVID.

Favorite Slide



Access the full summary and recording

AI: A modern approach

Peter Norvig, Google — October 2, 2021

Summary

Norvig starts with the history of the textbook. In 1990, the textbooks were subpar. AI was changing in three ways – moving from logic to probability, from hand-coded knowledge to machine learning, and from duplicating human systems toward normative systems that got the best answer no matter what. After leaving Berkeley to go to Sun, he helped write a new textbook about AI.

In software engineering, the main enemy is complexity. In AI the main enemy is uncertainty. Reasoning with uncertainty and interacting with the environment were the key points of the new AI textbook. The newest version of the textbook covers a lot about deep learning as well. The difficult part of AI is not the algorithms, it's deciding what you want to optimize. Ethics, fairness, privacy, diversity equity, and lethal autonomous weapons are taking a more prominent place in the discussion about AI. As the field has changed, the students have also changed. AI is now a requirement rather than an elective. The newest edition streamlines the content to be more accessible.

Key Points from the Q&A Session

- A practical, rather than philosophical, understanding of intelligence helps us solve problems
- AI could help assess college applications to create ideal diversity of ideologies across an entire newly incoming class. Alternative viewpoints can spawn efficient solutions to complex problems.
- From Google's perspective, filter bubbles are not as big a problem as people think. Facebook and other companies have a harder time dealing with it due to the nature of social media.
- Robots taking over the world is not a critical problem, unintended effects of AI are. Cheap surveillance for totalitarian governments.
- The definition of AI may be too broad, but what really matters is how the communities working on AI interact with each other
- Decentralized politics are not a big concern for Google. They focus more on data privacy.
- We are already living in a world with nonhuman entities – corporations and governments. We can't understand them completely but we have some predictions of what they will do. The same is going to be true for AI. The danger is from the rate of change, AI will be capable of much faster change.
- Computer science has become more of a natural science. It's too complex now to simply focus on proofs and mathematics.
- Building trustworthy systems is important.
- Criticisms of the "approach of maximizing utility" tend to ignore externalities. Taking a broader view of the utilitarian systems tends to solve the paradoxes that spring up.
- AI is a compliment rather than a substitute for human labor in the economy. It's a tool that helps people get their jobs done.
- People in AI have often rediscovered things that are already known, we should do a better job at background research.

Access the full summary and recording

Space development property rights and legal considerations

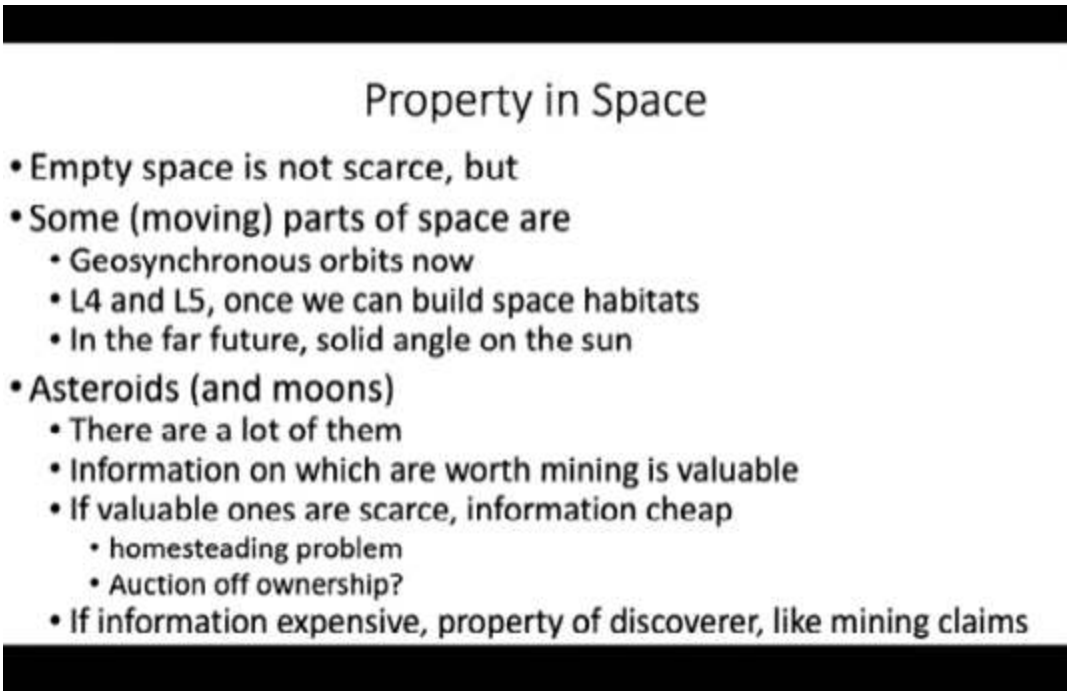
David Friedman, Santa Clara University
James Bennett, National Space Society **October 14, 2021**

Summary

James Bennett comments on space development and universal basic income. He believes that within 24 months a demonstration of dramatically lowered costs may occur with SpaceX's Starship. A mass accelerator could be used to create low-cost shipping of bulk material to orbit. Space habitats could be built which take advantage of low-cost shipping instead of being self sustaining – which would change the design and constraints of constructing such habitats.

David Friedman speak about the nature of space property and the various solutions for handling it. The most important aspect is whether we will have cheap space travel, which leads to a space economy closely tied to the terrestrial economy – or expensive space travel which will cause the space economy to evolve independently.

Property ownership, scarcity, rights, and enforcement will be shaped by national and global interests over the coming decades.



Access the full summary and recording

SeL4: Formal proofs for real-world cybersecurity

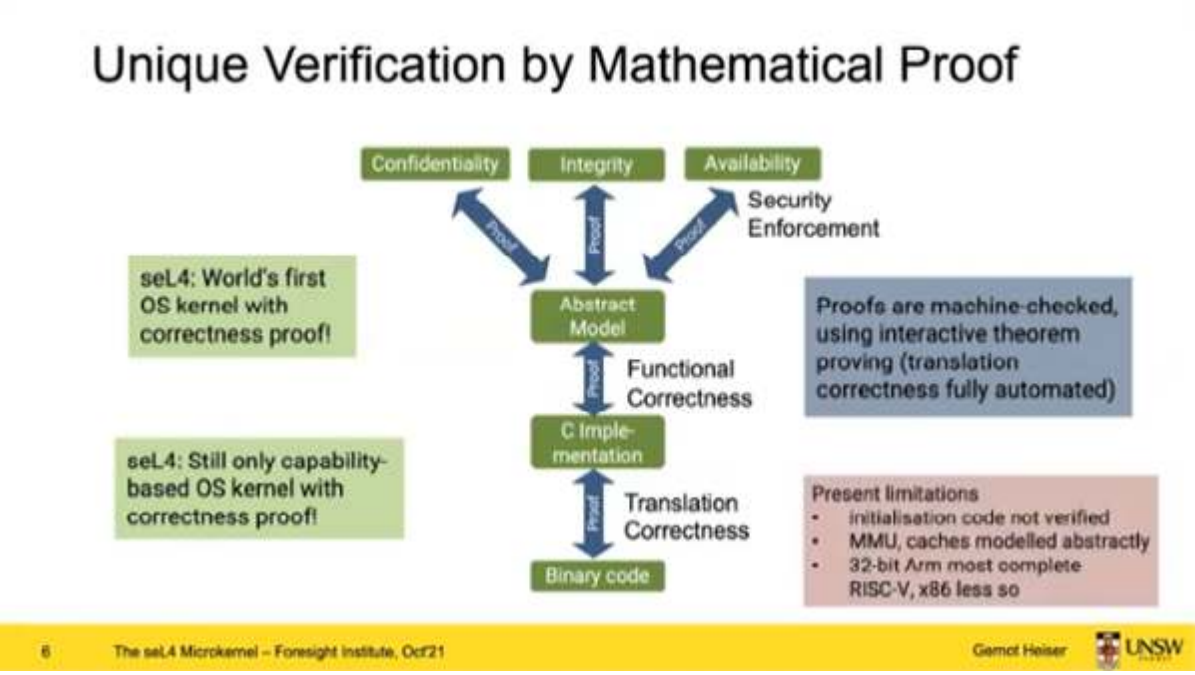
Gernot Heiser, Trustworthy Systems **October 19, 2021**

Summary

The sel4 microkernel system was designed to be a highly secure yet open source framework. It functions on a level below the operating system, between the device hardware and user software layers. With one of the largest proofs of security in existence as well as successful real world implementation, this microkernel is one of the best examples of cutting edge yet practical applications of cybersecurity. At over 1 million lines of proof script, the proof is the largest maintained proof base in an evolving system. Optimization of the system is still ongoing.

Opportunities

The microkernel can still be optimized for time protection, cost reduction, greater security, and more formalized methods.



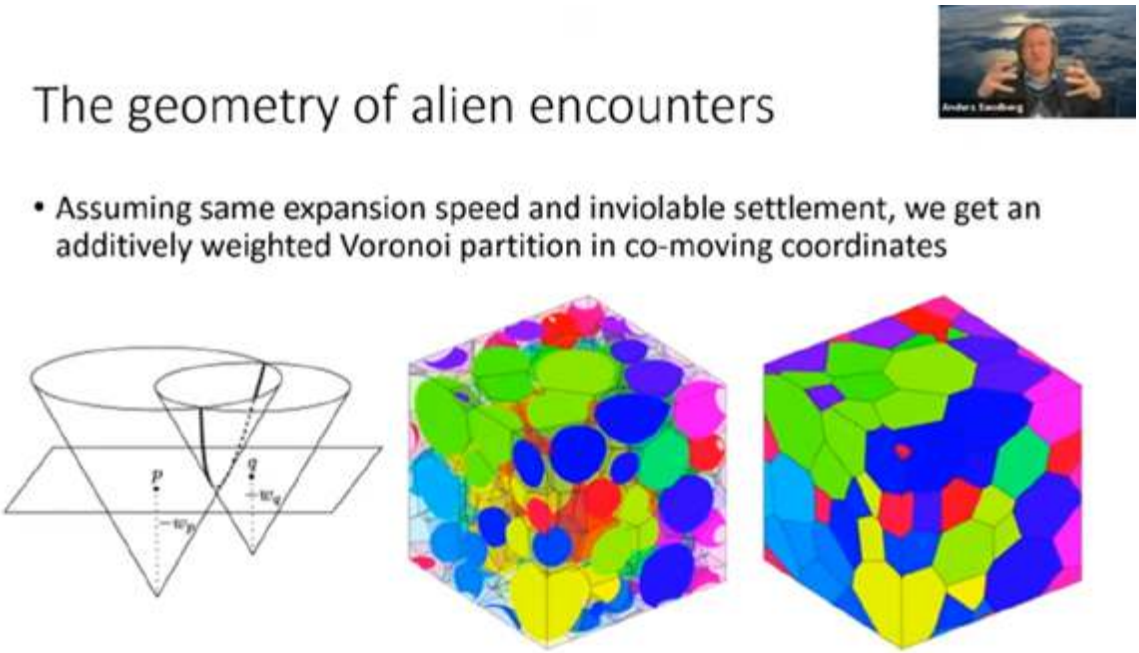
Access the full summary and recording

Game theory of cooperating with extraterrestrial intelligence

Anders Sandberg, University of Oxford **November 8, 2021**

Summary

Anders dives into the far future, investigating how game theory might apply to galactic and universe scale civilizations. Conquest, diplomacy, and economic forces will be subject to constraints of light speed travel and emergent properties of space expansion. As our short term existential problems get solved, we may want to reflect on the future and figure out what long term goals we want and how to achieve them.



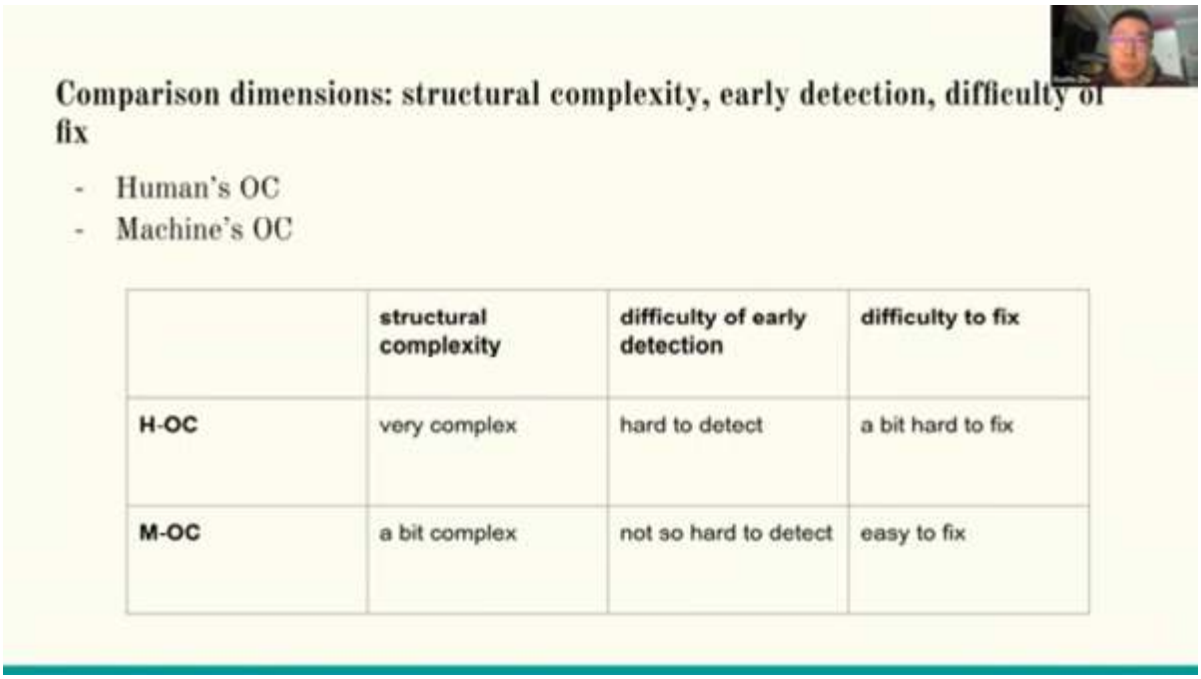
Access the full summary and recording

Ontological anti-crisis and AI safety

Zhu Xiaohu, Center for safe AGI **November 24, 2021**

Summary

Zhu speaks about the nature of ontological crisis – a state of reality shift for either human or machine. He then transitions into the ontological anti-crisis, and how to use such a phenomenon to increase safety for artificial general intelligence.



Access the full summary and recording

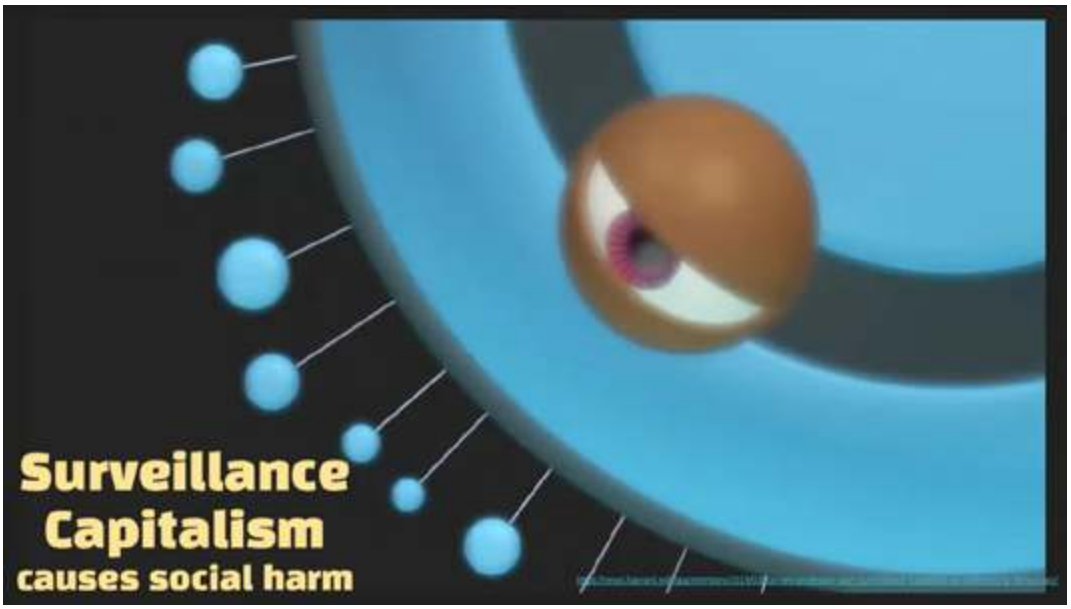
Re-decentralizing networked communities & the Spritely institute

Christine Lemmer-Webber, Spritely Institute
Randy Farmer, Playable Worlds

November 24, 2021

Summary

Christine and Randy describe how the current schema of social media has failed us. Networked communities have become centralized around a few large entities which are prone to failure. Poor incentives drive toxic behavior and more regulation will only reinforce the dominance of the current paradigm. The Spritely Institute outlines it's plan for a new mechanism of interaction and community formation on the internet.



Access the full summary and recording

Intelligent Cooperation Bountied Brainstorm

The Bountied Brainstorm aims to improve cooperation across human and other intelligences by allowing participants to ask open-ended questions, answer questions posed by others, and vote on the best answers.

Foresight pays monetary bounties for each contribution and the best answers. Some questions are private, others are public, including questions like “What would it take to secure civilization’s computer infrastructure?”, or “What is the most exciting cooperative arrangement that could be unlocked with zero-knowledge proofs?”

You can find a selection of answers in our Bountied Brainstorm corner.