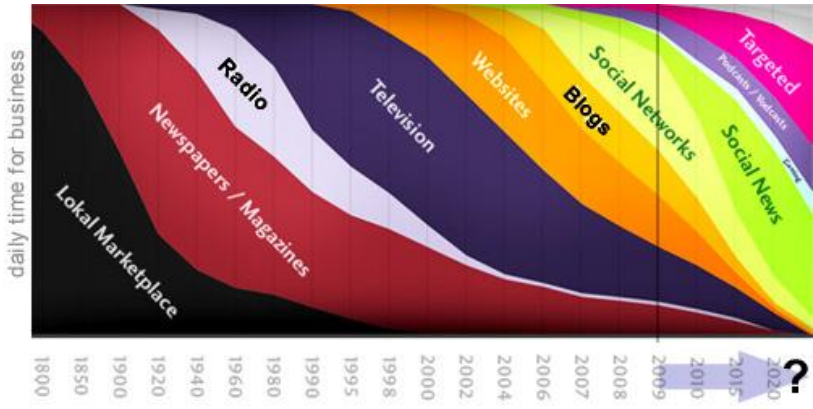
Smarter policy-making through improved collective cognition?

Anders Sandberg
Future of Humanity Institute
Oxford Martin School
Oxford University

Overview

- What is governance and policymaking, and why does it matter?
- How does technology change it?
- The problems of group decisionmaking
- New tools and architectures
- Where will we be in 2025?

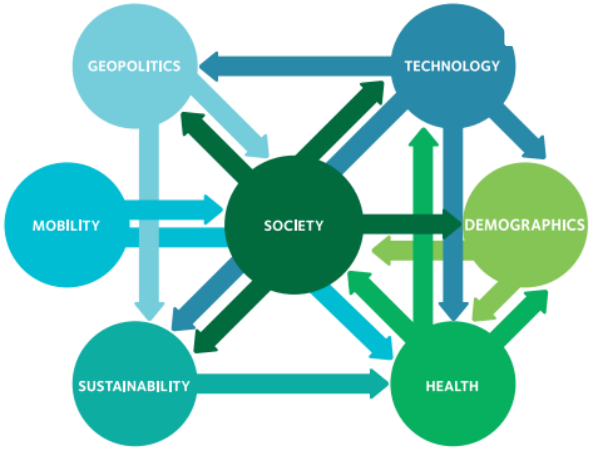


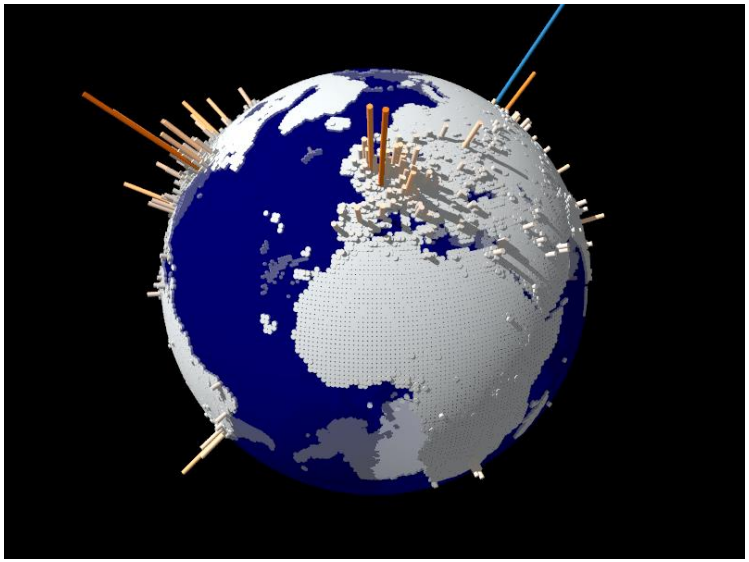
The policy cycle

- **Agenda setting**
  - Somebody needs to do something!
- **Policy Formulation**
  - What options do we have? (and who are *we*?)
- **Decision-making**
  - This is what we will do!
- **Implementation**
  - I am doing it.
- **Evaluation**
  - Did it work?

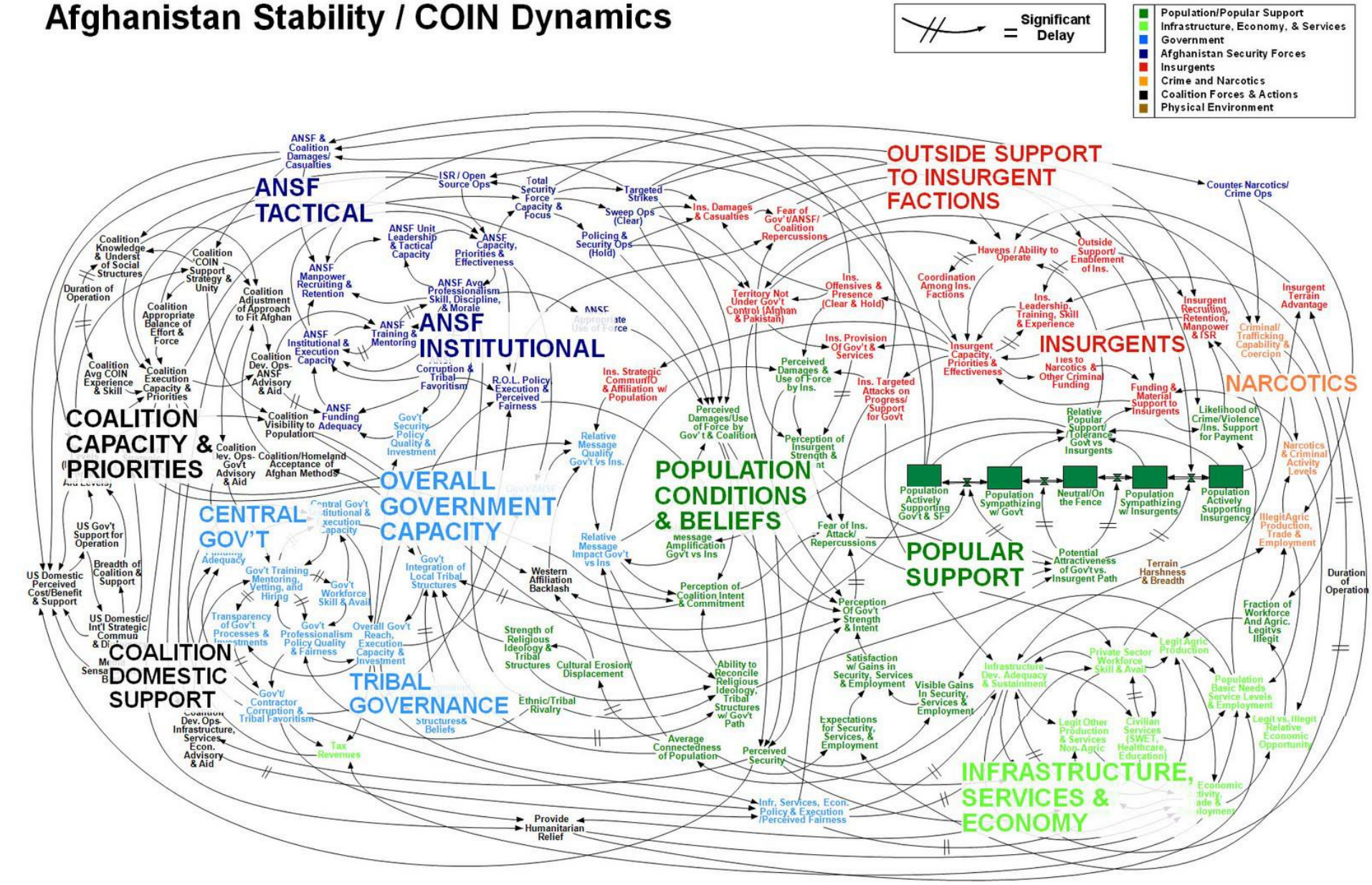
Challenges

- Diverse population
- Global, transnational
- Demographic change
- Complex threats
- Extreme priorities





Afghanistan Stability / COIN Dynamics



WORKING DRAFT – V3

Setting priorities worth taking seriously

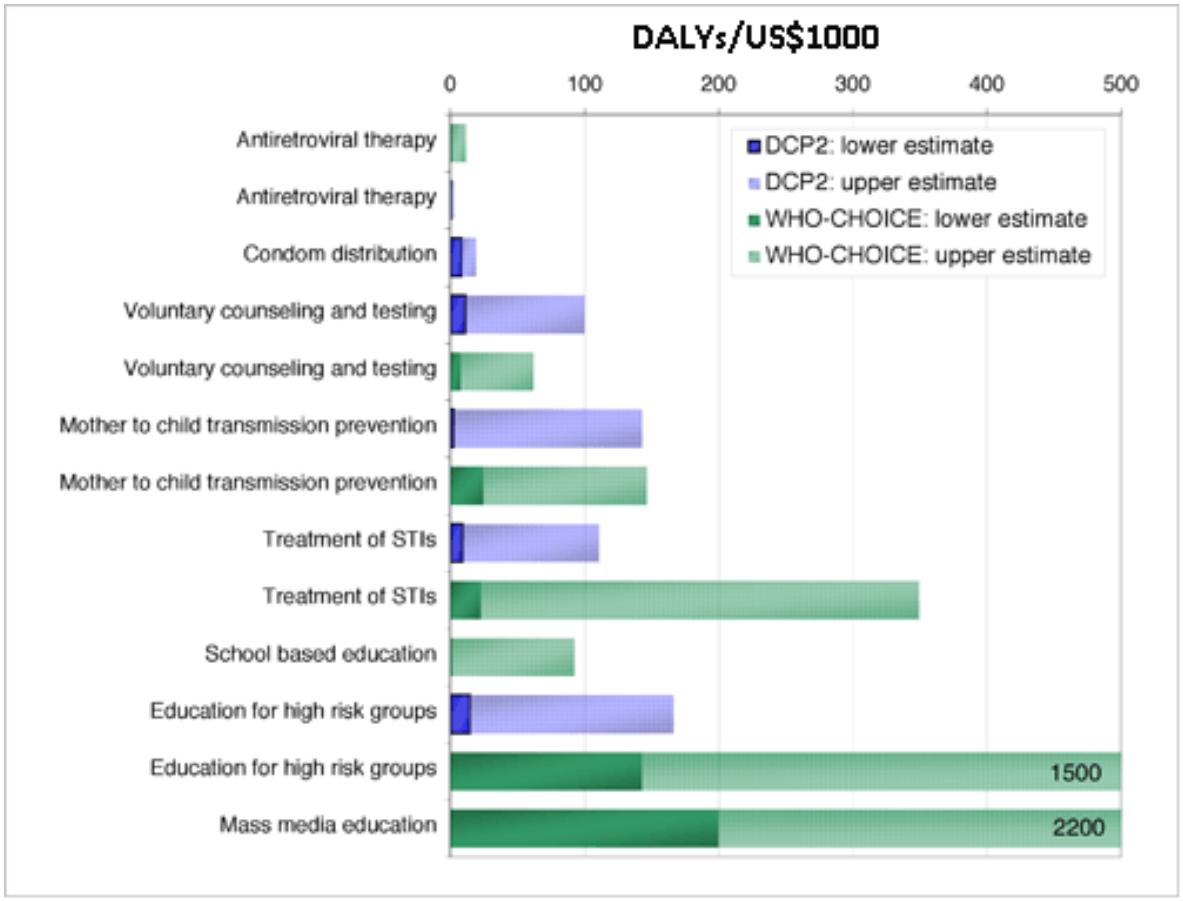
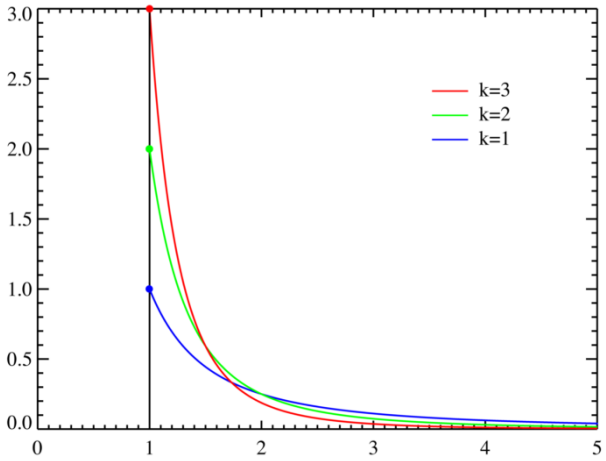
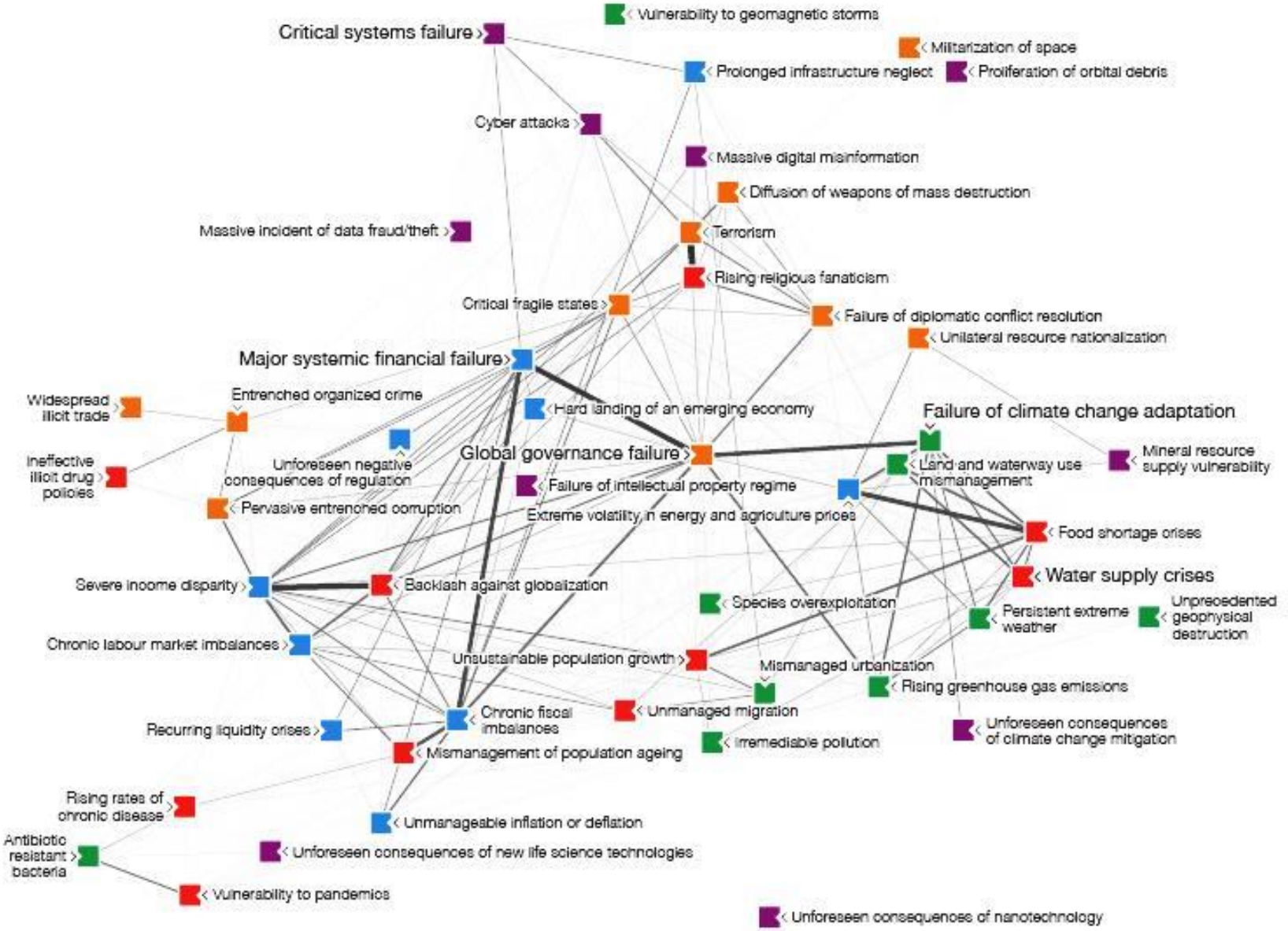


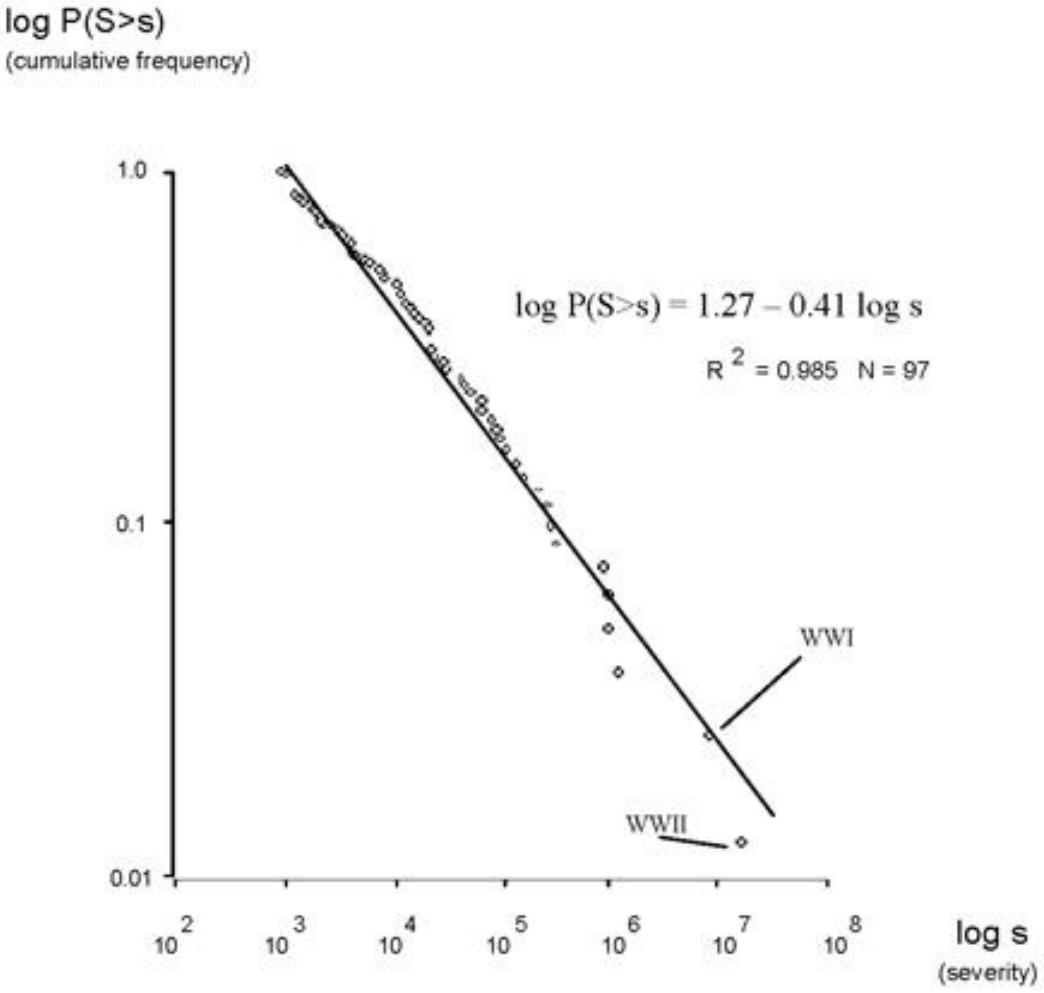
Figure 1: Cost-effectiveness (DALYs/US$1000) in fighting HIV/AIDS, http://www.givingwhatwecan.org



Optimal meta-level problem solving can at most give savings equal to half the expected utility difference between two alternatives – but if the alternatives matter *enormously*, spending effort is hugely valuable.

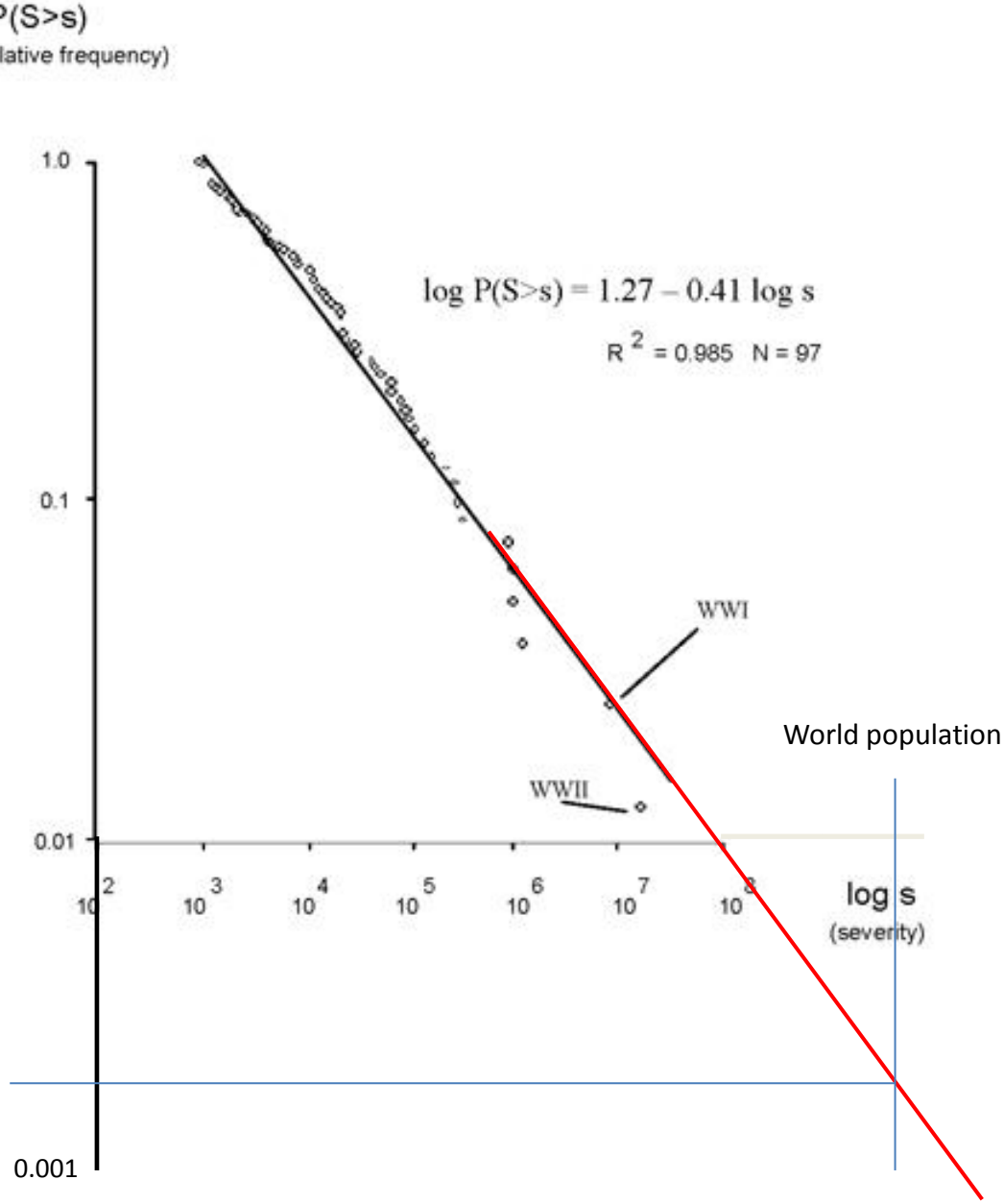

Vulnerability to geomagnetic storms
Critical systems failure
Militarization of space
Prolonged infrastructure neglect
Proliferation of orbital debris
Cyber attacks
Massive digital misinformation
Diffusion of weapons of mass destruction
Massive incident of data fraud/theft
Terrorism
Rising religious fanaticism
Critical fragile states
Failure of diplomatic conflict resolution
Unilateral resource nationalization
Major systemic financial failure
Widespread illicit trade
Entrenched organized crime
Hard landing of an emerging economy
Failure of climate change adaptation
Ineffective illicit drug policies
Unforeseen negative consequences of regulation
Global governance failure
Land and waterway use mismanagement
Mineral resource supply vulnerability
Failure of intellectual property regime
Pervasive entrenched corruption
Extreme volatility in energy and agriculture prices
Food shortage crises
Water supply crises
Severe income disparity
Backlash against globalization
Species overexploitation
Persistent extreme weather
Unprecedented geophysical destruction
Chronic labour market imbalances
Unsustainable population growth
Mismanaged urbanization
Rising greenhouse gas emissions
Unmanaged migration
Chronic fiscal imbalances
Recurring liquidity crises
Irremediable pollution
Unforeseen consequences of climate change mitigation
Mismanagement of population ageing
Rising rates of chronic disease
Unmanageable inflation or deflation
Antibiotic resistant bacteria
Unforeseen consequences of new life science technologies
Vulnerability to pandemics
Unforeseen consequences of nanotechnology

Cumulative war-size plot, 1820-1997



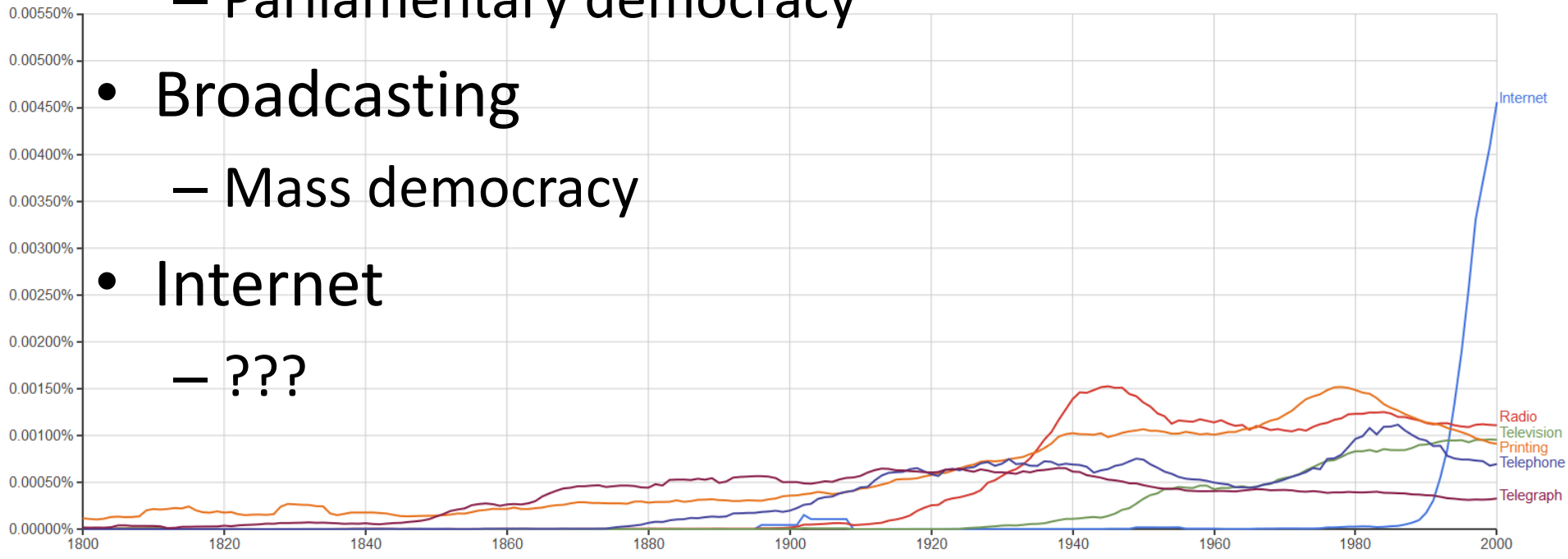
Data Source:
Correlates of War Project (COW)

Data Source:
Correlates
of War
Project (COW)



Social systems are information limited

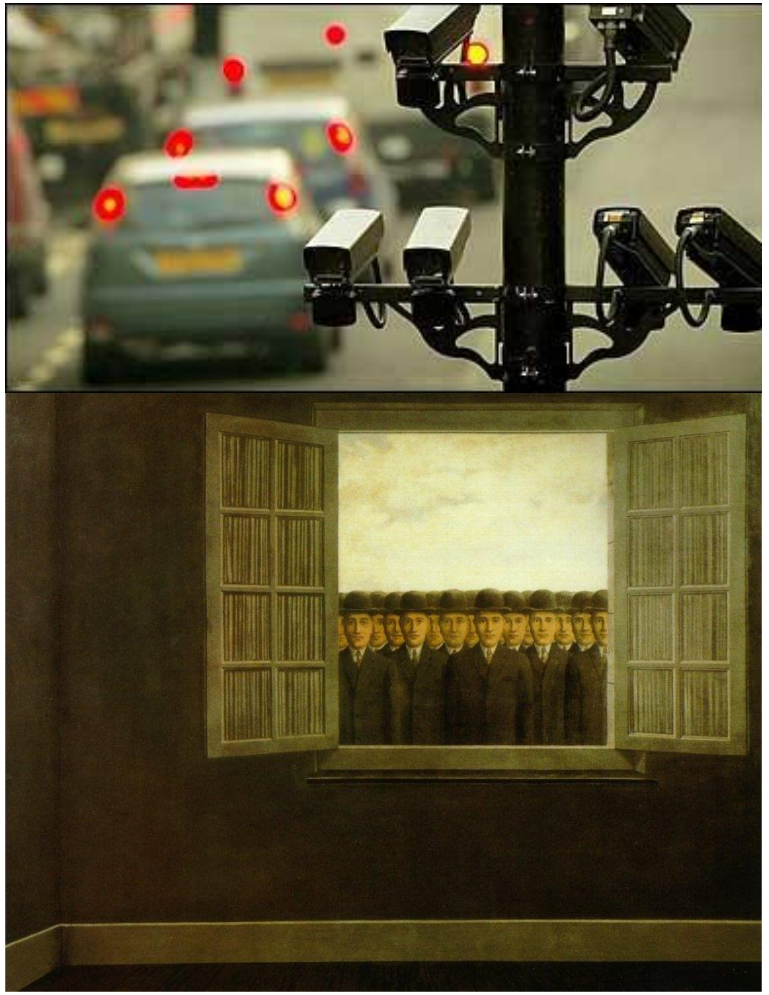
- Writing
  - Code of law
- Printing
  - Parliamentary democracy
- Broadcasting
  - Mass democracy
- Internet
  - ???



Governance innovations

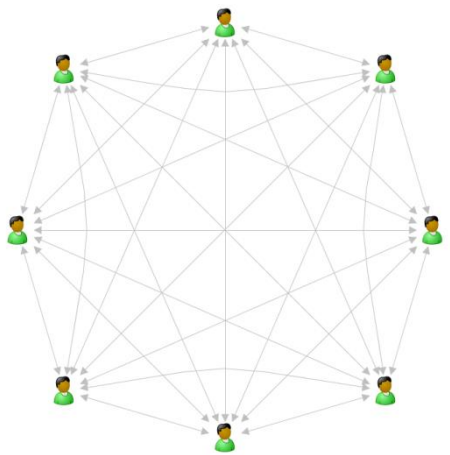
(after Moore & Hartley 2008)

- Not just better products and processes
- Changing the organisational boundaries (networked decisionmaking, financing, production)
- Tap new pools of resources
- Exploit government capacity to shape private rights and responsibilities
- Redistribute the right to define and judge value

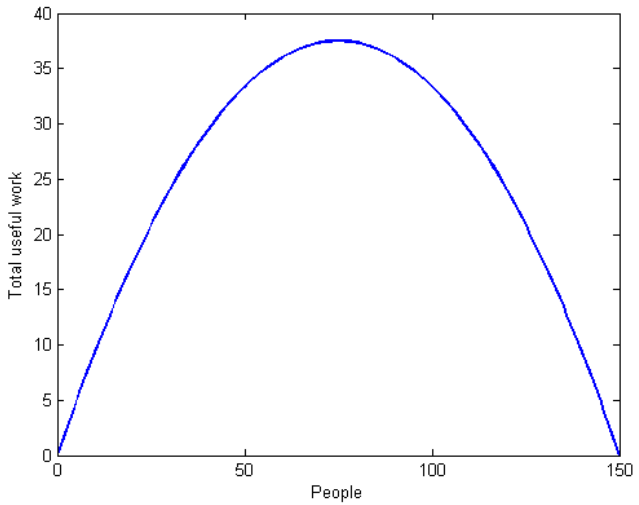


Group problem solving

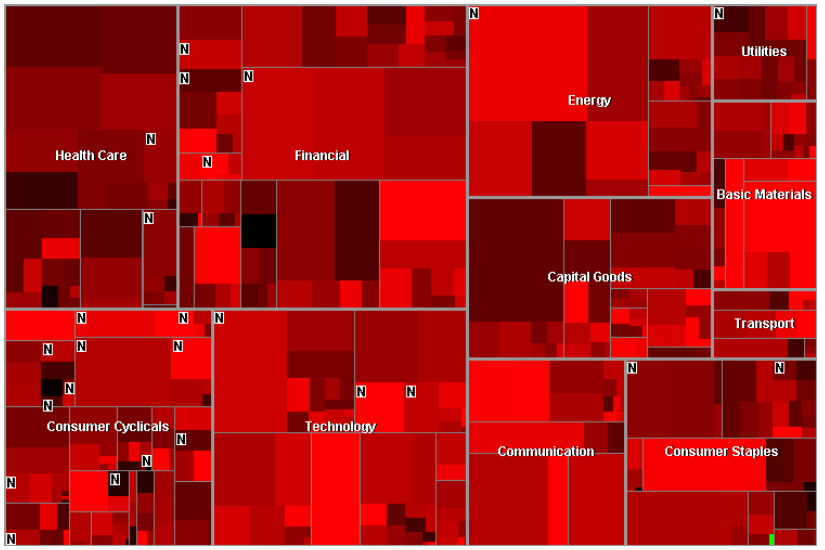
- Great for smoothing out individual bias
- Great for verifiable solutions
- Bad at “vexed problems”
- Scaling issues

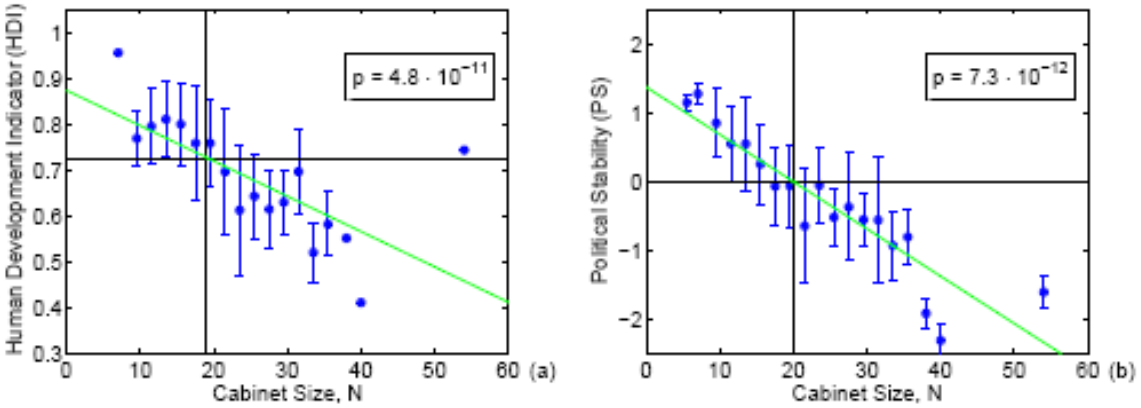


Total work = individual work – communications losses



Stupidity of crowds

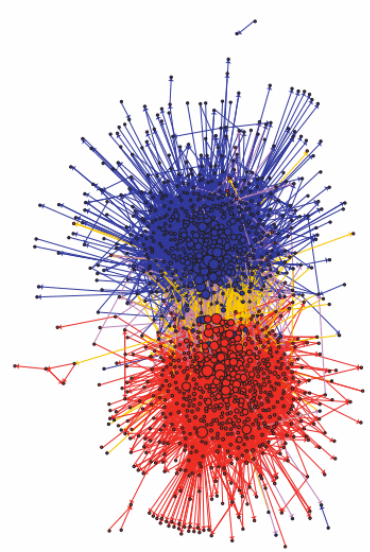


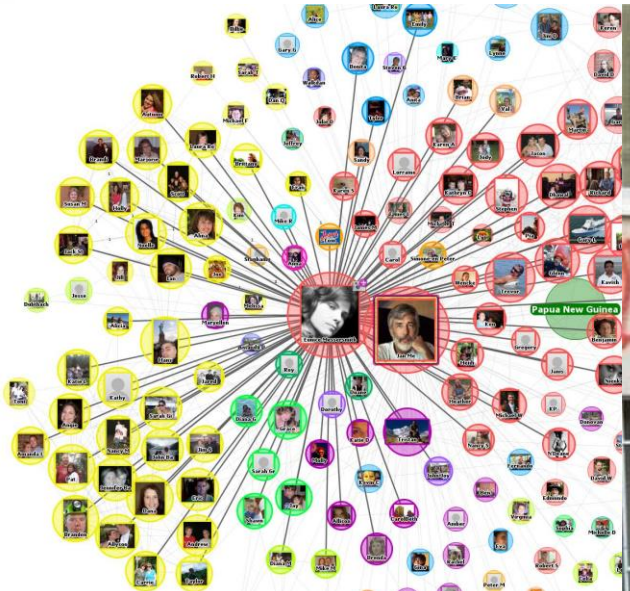


Peter Klimek, Rudolf Hanel and Stefan Thurner, Parkinson's Law quantified: three investigations on bureaucratic inefficiency, *J. Stat. Mech*. P03008 (2009)

Social media and information cascades

- Opinion formation, distribution
  - Twitter vs blogs vs television
- Microexpertise
- Microteaching

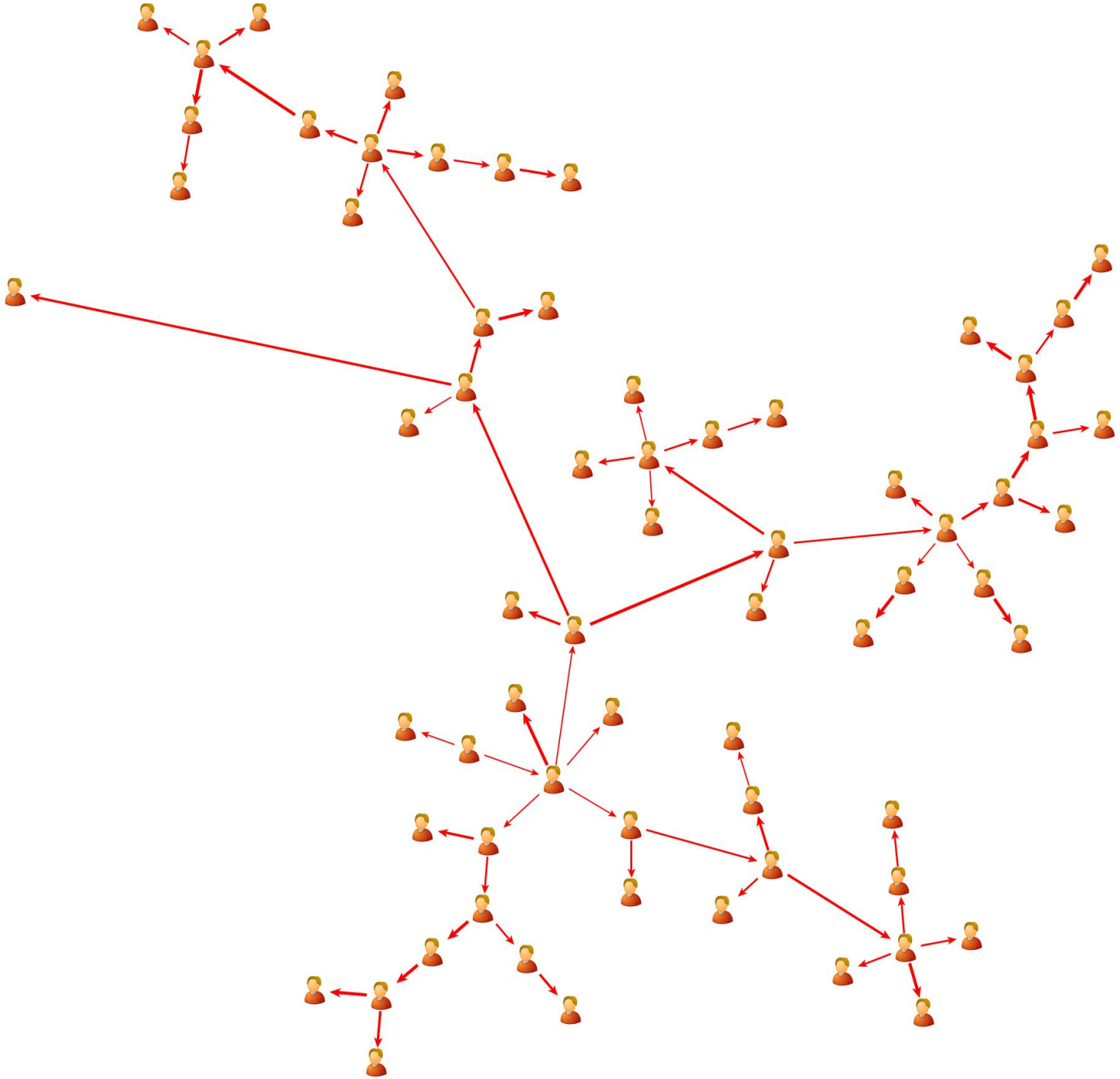


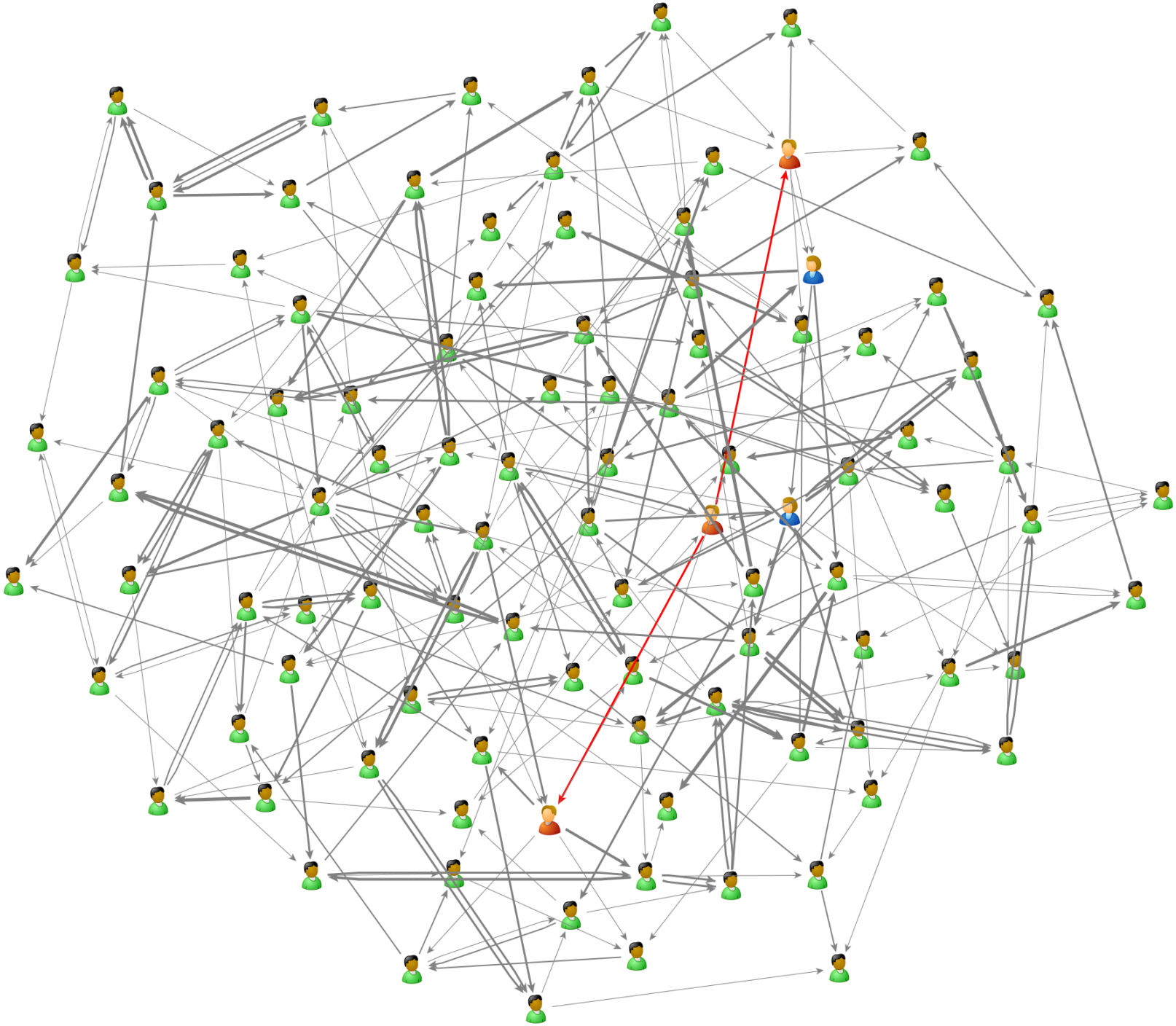


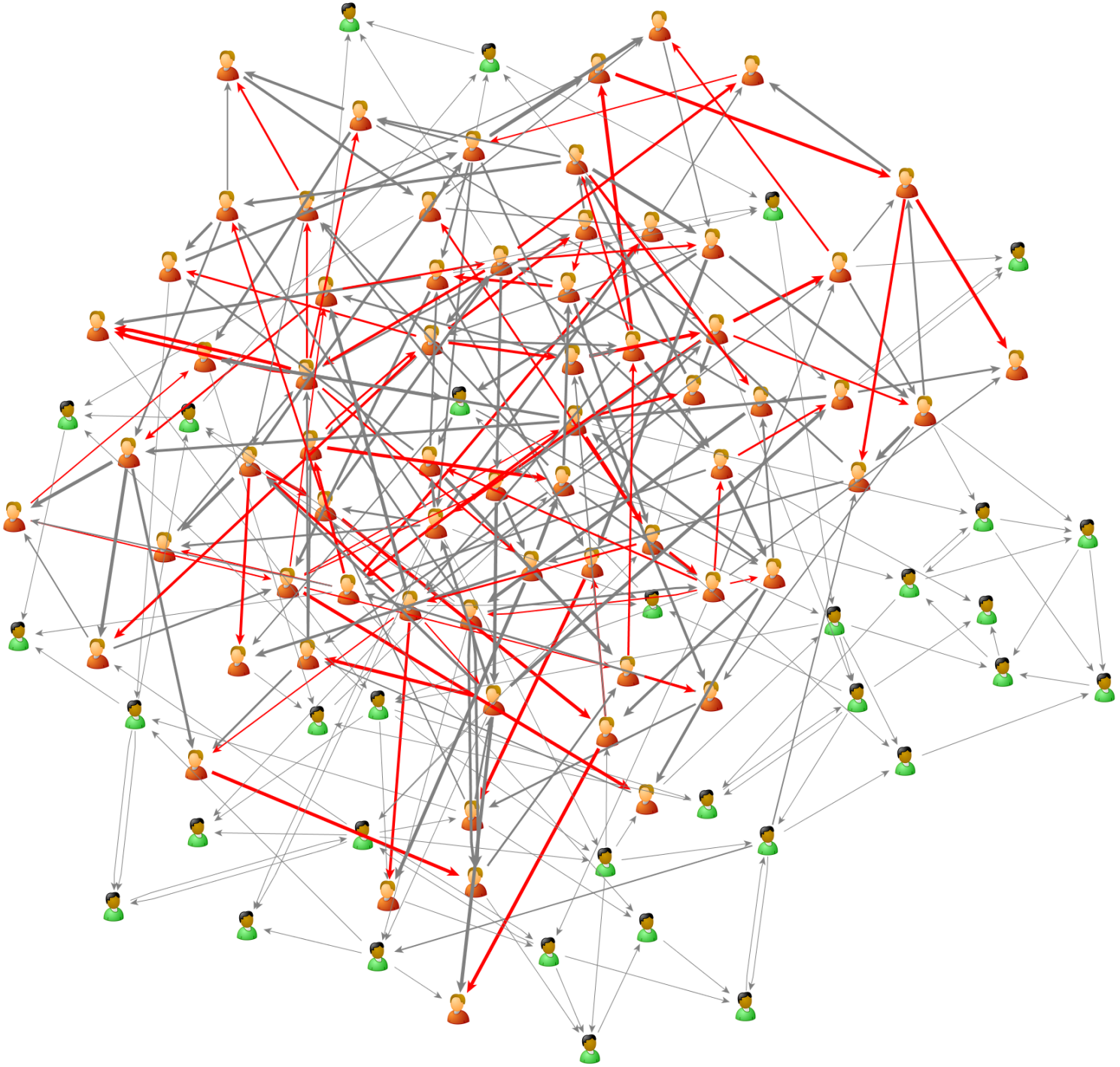
The information cycle: or, why blogs[1] get smart

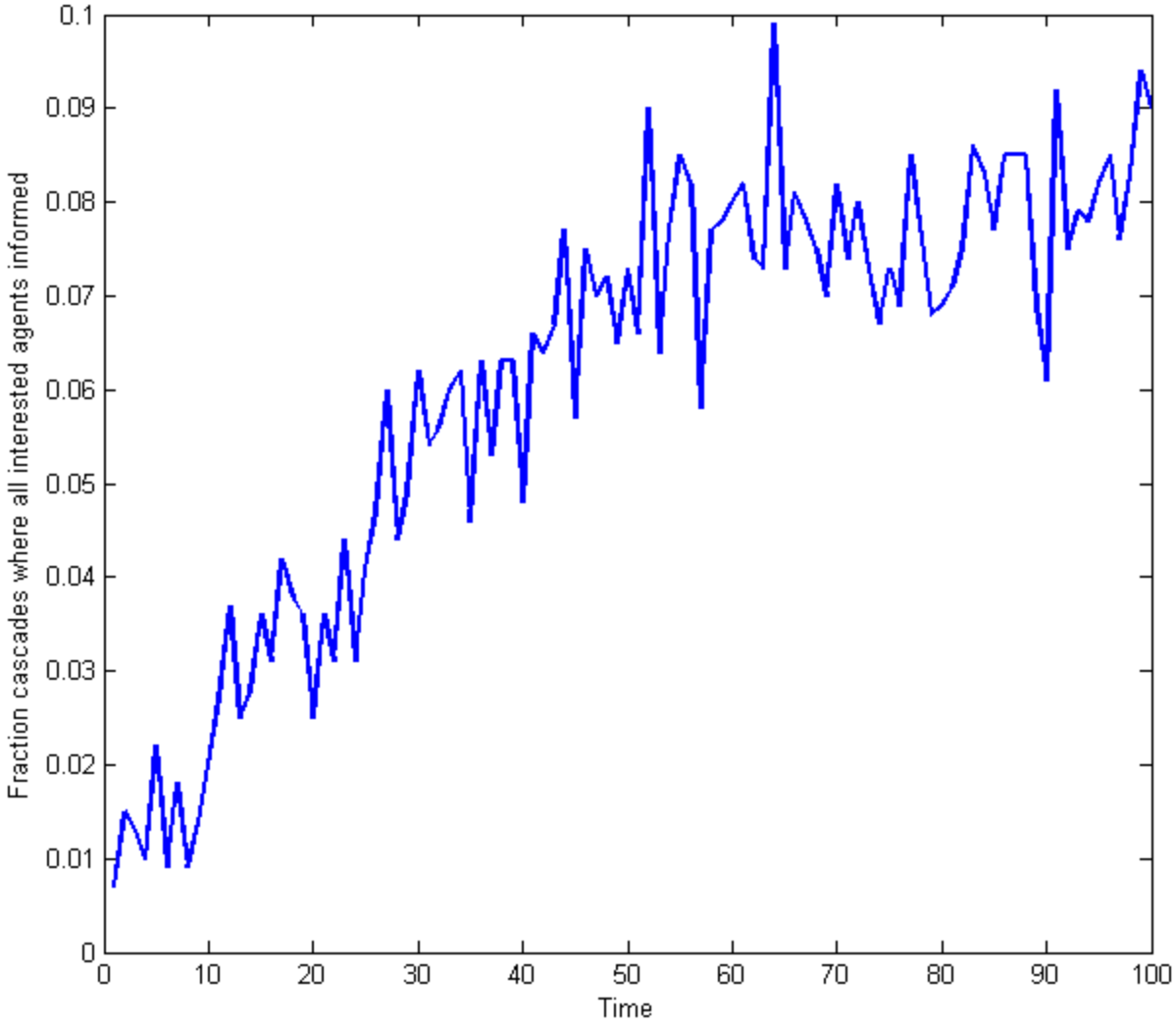
- Managing limited input channels
- Model
  - Agents read a limited number of information sources and have certain interests (many unique)
  - Random agents post stories based on their interests
  - If one read agent has a piece of interesting news that source is regarded as more useful by the agent
  - Agents write about the interesting news with a certain probability, creating information cascades
  - Over time agents randomly shift to other sources if one proves useless

1: Or twitterers, journals, friends, etc…








Fraction cascades where all interested agents informed
Time
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
0
10
20
30
40
50
60
70
80
90
100

Intelligence - Wikipedia, the free encyclopedia - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://en.wikipedia.org/wiki/Intelligence

Go wikipedia

Dataphone Intermail ... CSC - Login Webmail login Google Scholar Entrez PubMed

Dataphone Inter... SquirrelMail 1.4.8 Intelligence - ... The month of th... Problem loading ... magritte.jpg (JP...

Intelligence

From Wikipedia, the free encyclopedia

For other uses, see Intelligence (disambiguation).

Intelligence is a property of mind that encompasses many related mental abilities, such as the capacities to reason, plan, solve problems, think abstractly, comprehend ideas and language, and learn.

Although many regard the concept of intelligence as having a much broader scope, for example in cognitive science and computer science, in some schools of psychology, the study of intelligence generally regards this trait as distinct from creativity, personality, character, or wisdom.

Contents [hide]

Look up *intelligence* in Wiktionary, the free dictionary.

WIKIPEDIA
The Free Encyclopedia

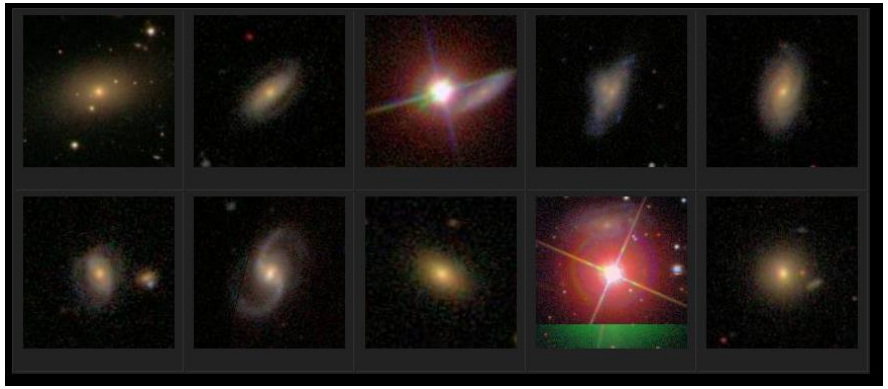
navigation
Main Page
Community Portal
Featured articles
Current events
Recent changes
Random article
Help
Contact Wikipedia
Donations

search
cognition
Go Search

toolbox
What links here
Related changes
Upload file
Special pages
Printable version

Find: magr Find Next Find Previous Highlight all Match case

Transferring data from www.staff.vu.edu.au...



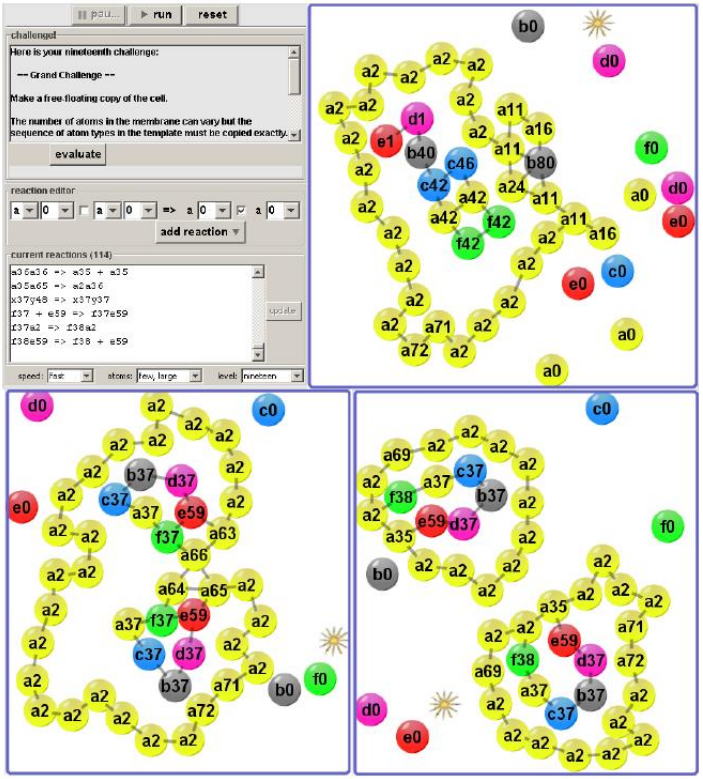
Galaxy Zoo/Zooniverse

A NEW PROOF OF THE DENSITY HALES-JEWETT THEOREM

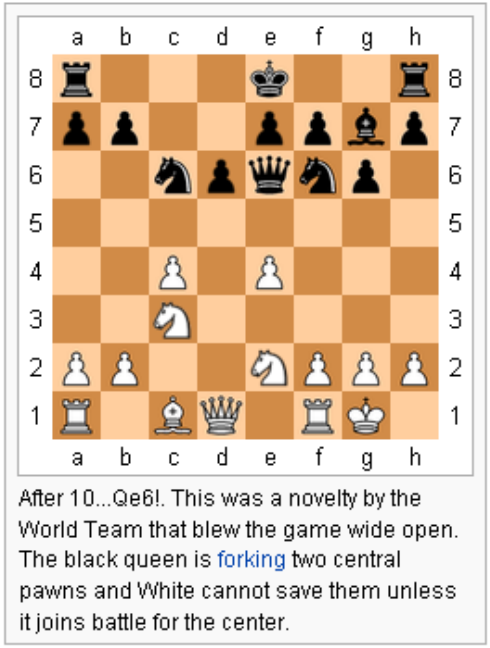
D. H. J. POLYMATH

ABSTRACT. The Hales–Jewett theorem asserts that for every r and every k there exists n such that every r-colouring of the n-dimensional grid $\{1,\dots,k\}^n$ contains a combinatorial line. This result is a generalization of van der Waerden's theorem, and it is one of the fundamental results of Ramsey theory. The theorem of van der Waerden has a famous density version, conjectured by Erdős and Turán in 1936, proved by Szemerédi in 1975, and given a different proof by Furstenberg in 1977. The Hales–Jewett theorem has a density version as well, proved by Furstenberg and Katznelson in 1991 by means of a significant extension of the ergodic techniques that had been pioneered by Furstenberg in his proof of Szemerédi's theorem. In this paper, we give the first elementary proof of the theorem of Furstenberg and Katznelson, and the first to provide a quantitative bound on how large n needs to be. In particular, we show that a subset of $\{1,2,3\}^n$ of density δ contains a combinatorial line if n is at least as big as a tower of 2s of height $O(1/\delta^2)$. Our proof is surprisingly simple: indeed, it gives arguably the simplest known proof of Szemerédi's theorem.

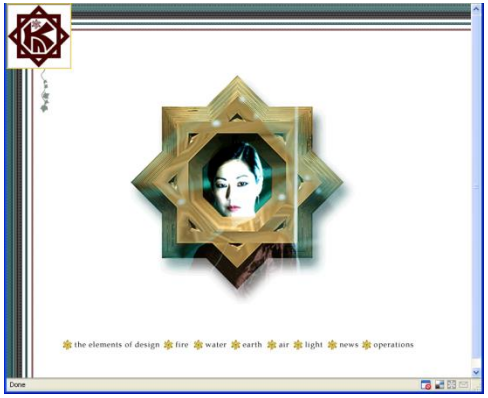
Polymath project, Timothy Gowers



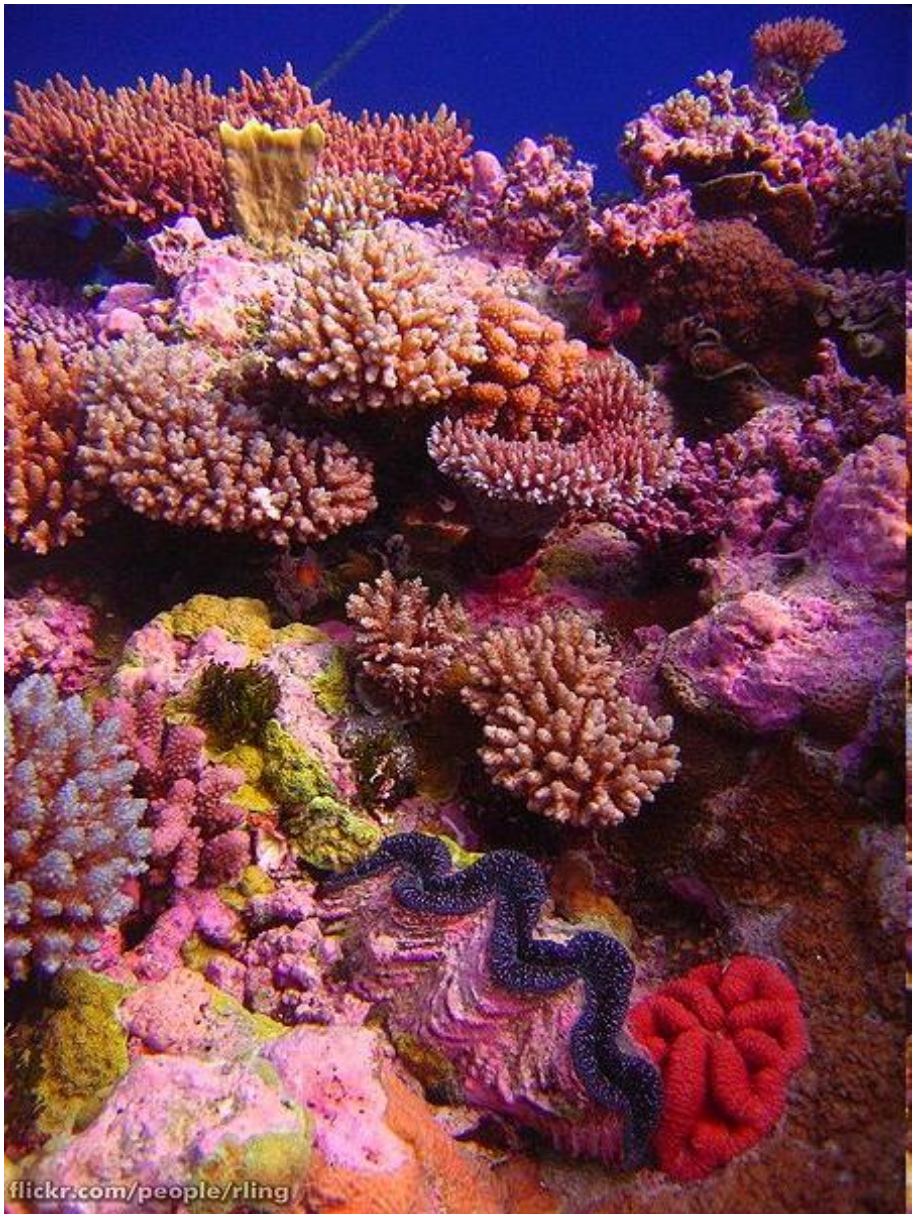
Organic Builder, Tim J. Hutton



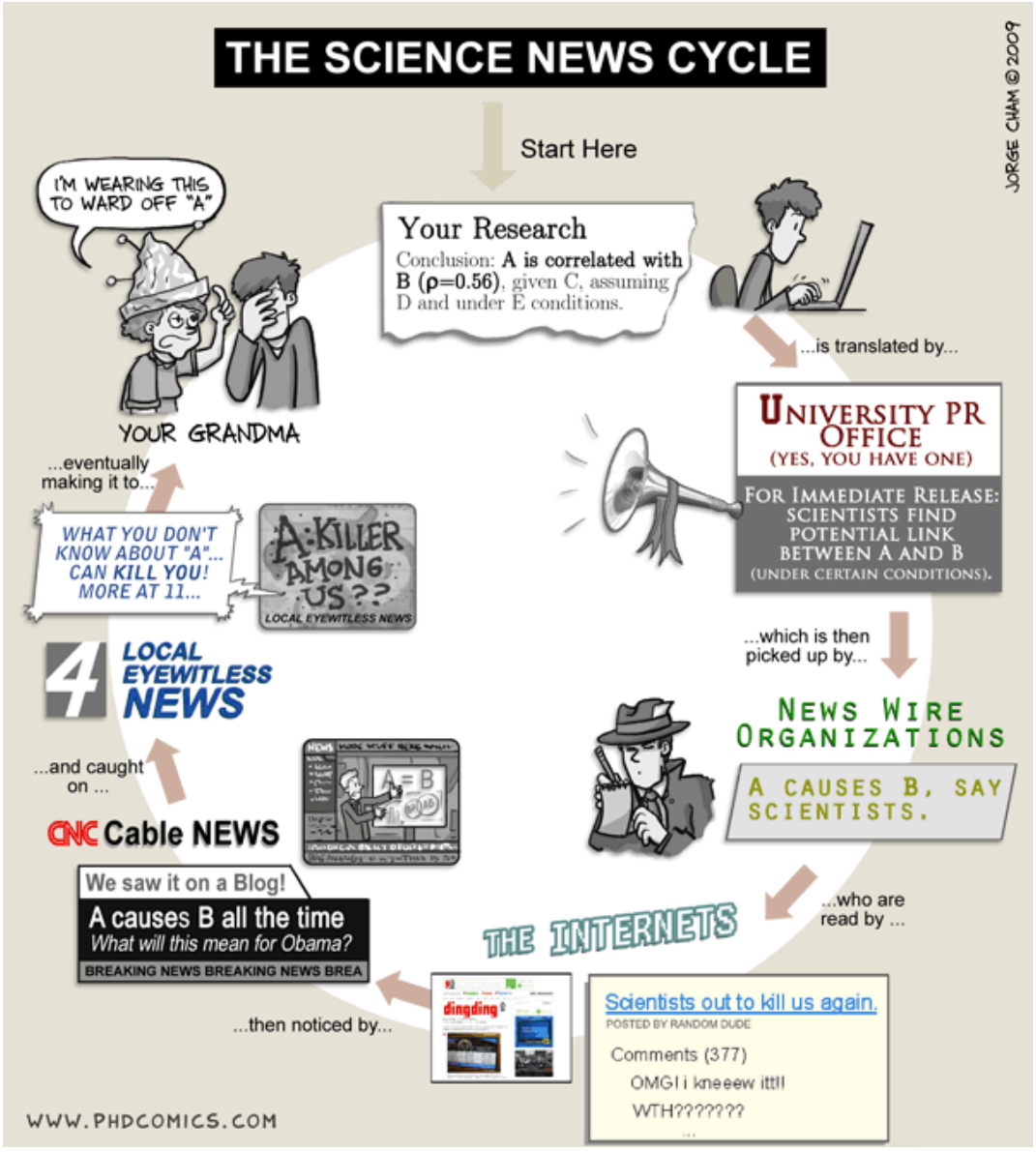
Kasparov vs. the world



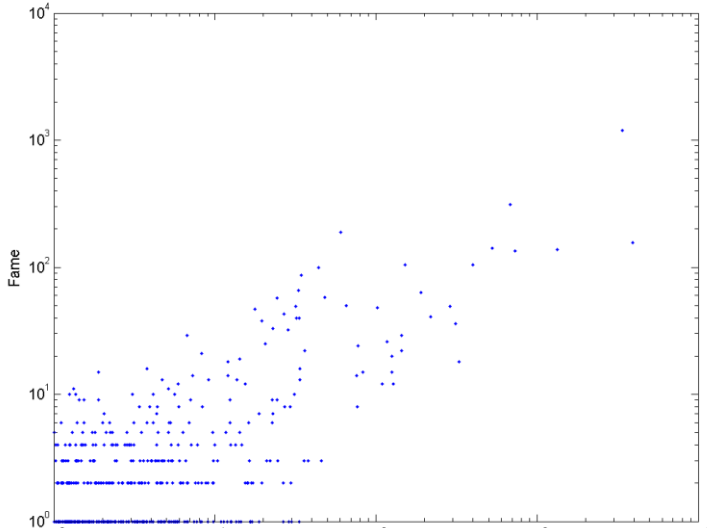
"The Beast", Microsoft Entertainment


flickr.com/people/rling

The disinformation cycle



Reputations



cousin_it

Wow, Yudkowsky's CEV runs into Yudkowsky's Löb problem! That's cute :-)

Stuart_Armstrong *06 June 2013 03:38:02PM* 0 points

Hadn't thought of it that way! :-)

But very cool!

hairyfigment *07 June 2013 12:59:17AM* 0 points

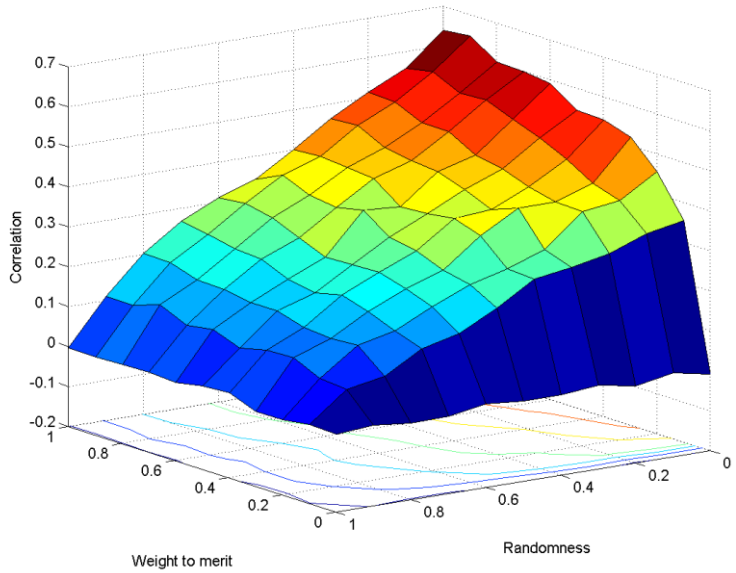
Can you say what made the connection non-obvious?

What do you see as Eliezer's main research objective?

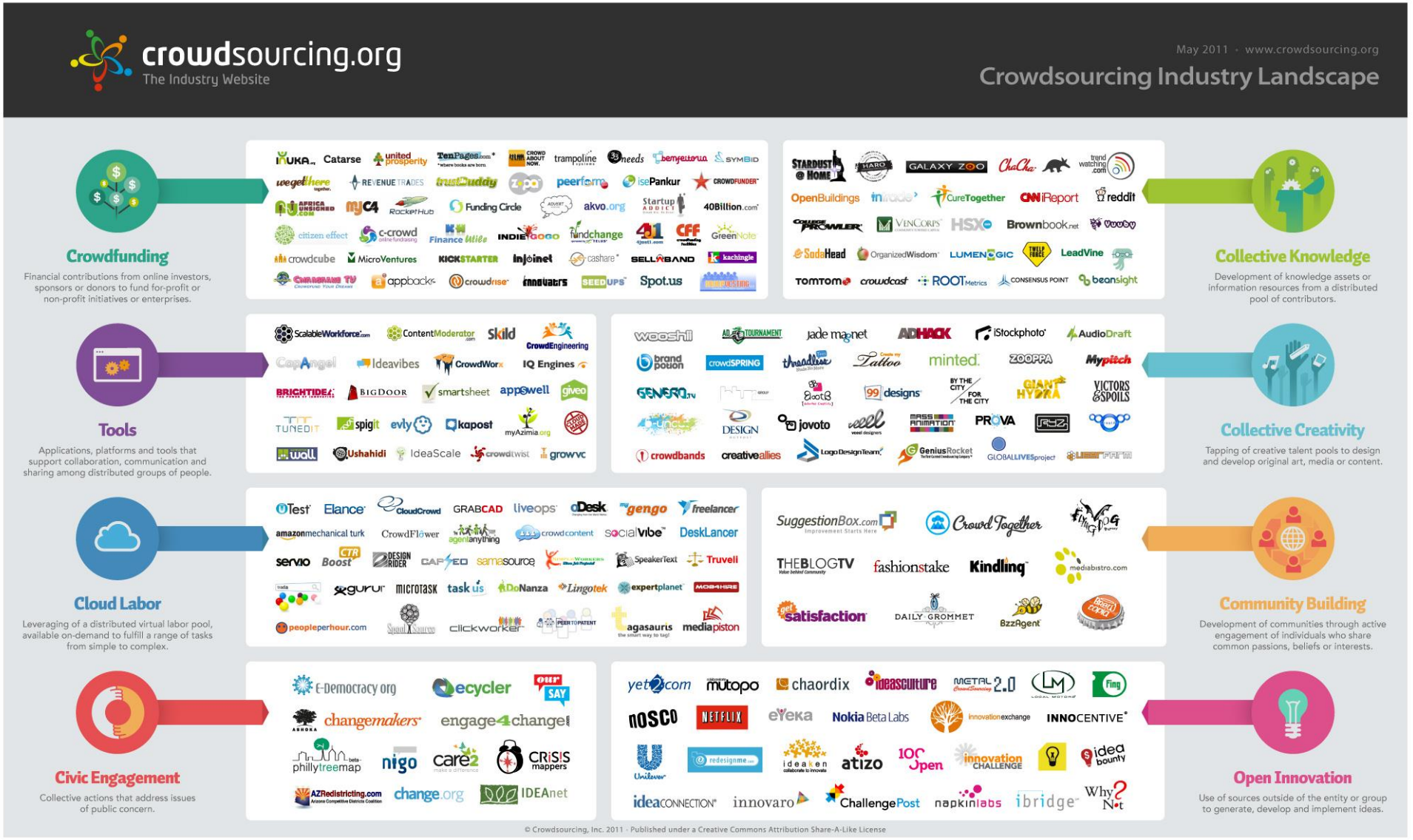
Stuart_Armstrong *07 June 2013 07:13:22AM* 4 points

> Can you say what made the connection non-obvious?

The fact I had come up with this idea well before hearing of Loeb's theorem.

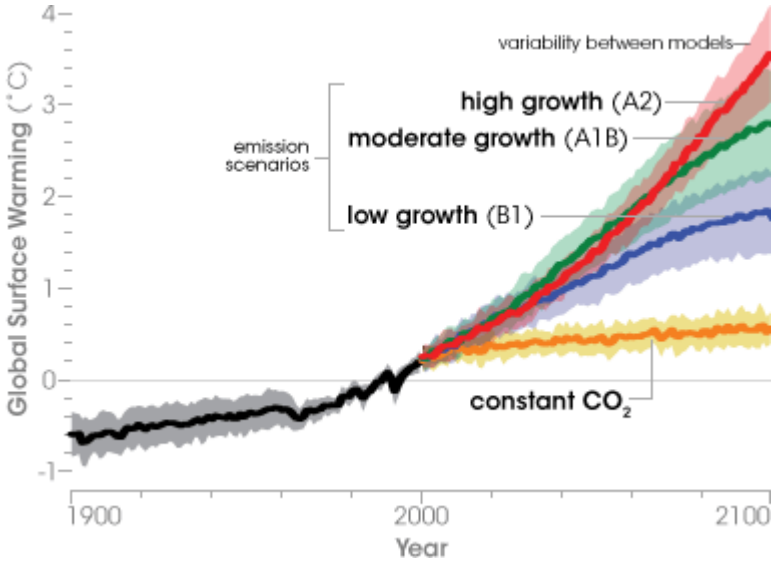


Crowdsourcing



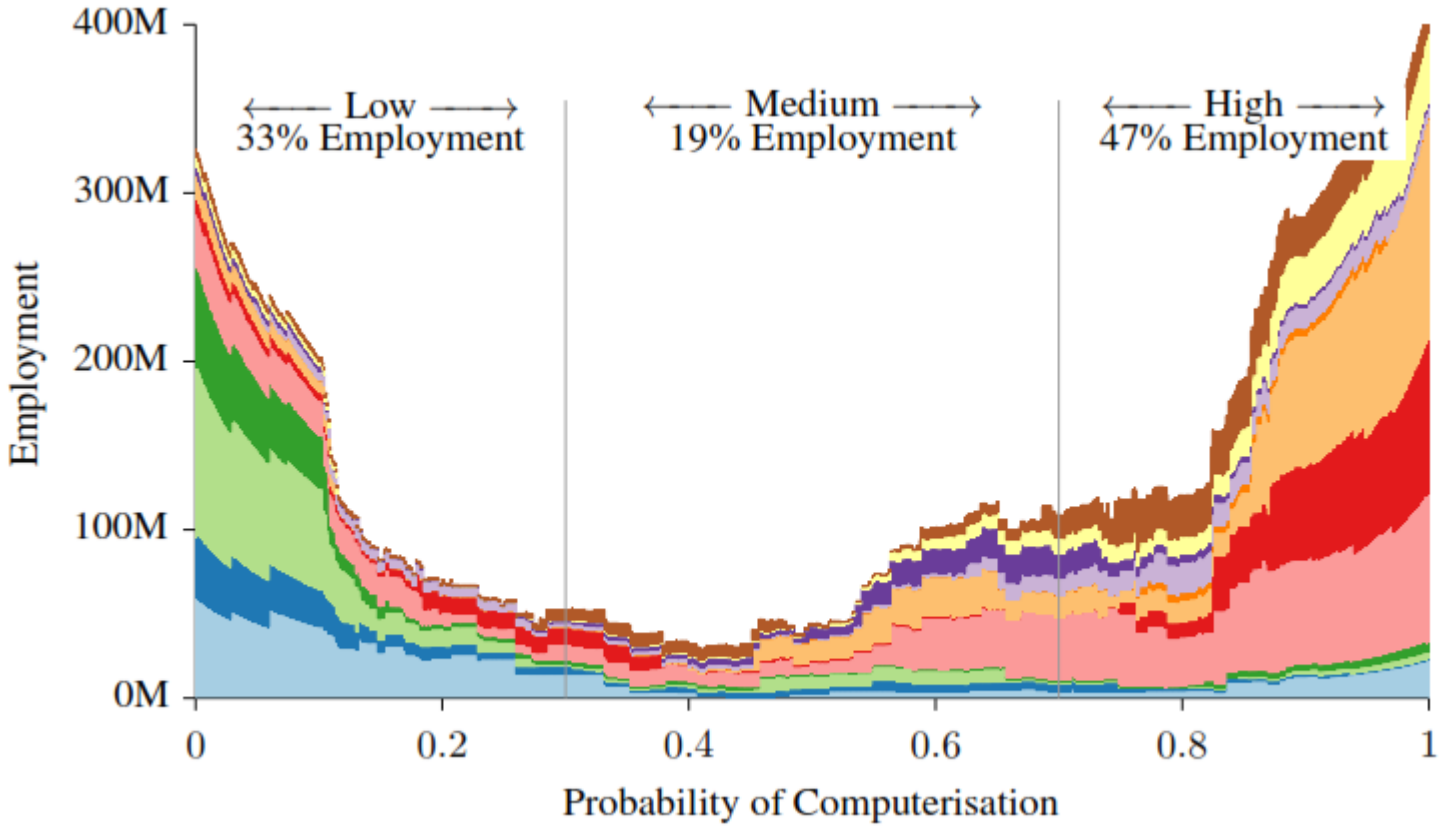
Modeling

- “All models are wrong, but some are useful”
- Get scenarios based on data
- But which data? Which kind of scenarios?
- One-to-one total modelling to the rescue?



Big Data

- Big Input
  - Default mode
  - Surveillance
  - Internet of things
  - Identity technology
- Big Data = Big Centralisation
- The government cloud
- Superknowledge
- Civil service Watson?



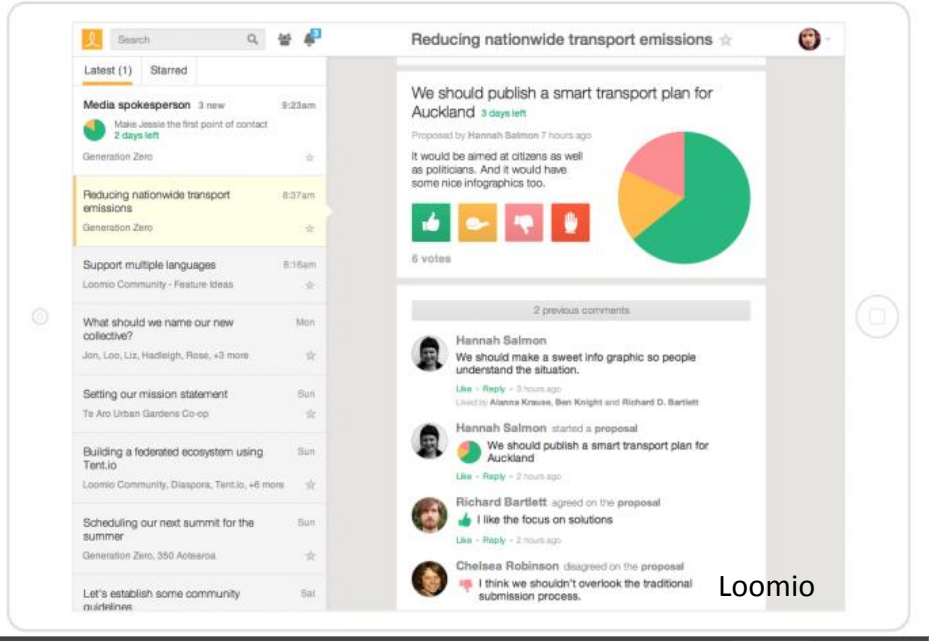
- Management, Business, and Financial
- Computer, Engineering, and Science
- Education, Legal, Community Service, Arts, and Media
- Healthcare Practitioners and Technical
- Service
- Sales and Related
- Office and Administrative Support
- Farming, Fishing, and Forestry
- Construction and Extraction
- Installation, Maintenance, and Repair
- Production
- Transportation and Material Moving

Variable	Probability of Computerisation		
	Low	Medium	High
Assisting and caring for others	48±20	41±17	34±10
Persuasion	48±7.1	35±9.8	32±7.8
Negotiation	44±7.6	33±9.3	30±8.9
Social perceptiveness	51±7.9	41±7.4	37±5.5
Fine arts	12±20	3.5±12	1.3±5.5
Originality	51±6.5	35±12	32±5.6
Manual dexterity	22±18	34±15	36±14
Finger dexterity	36±10	39±10	40±10
Cramped work space	19±15	37±26	31±20

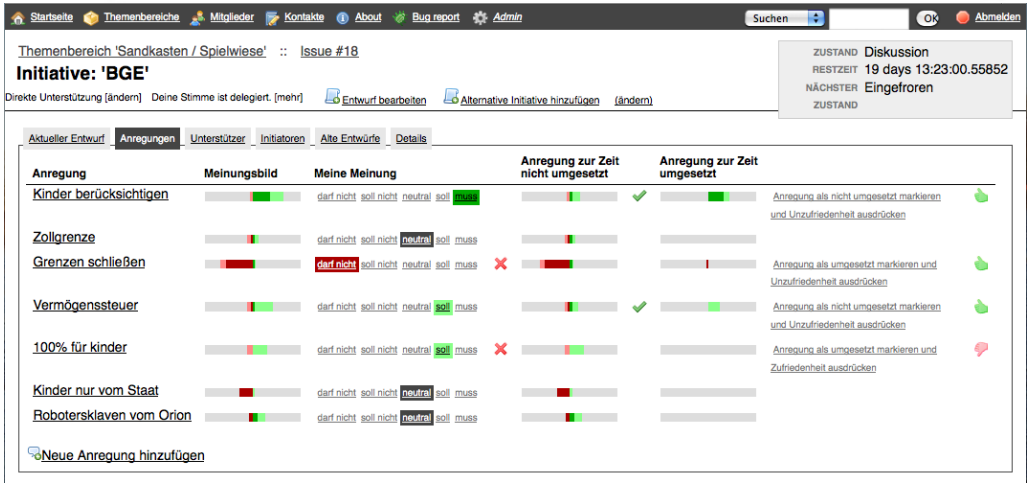
The future of employment: how susceptible are jobs to computerisation? Carl Benedikt Frey, Michael A. Osborne, 2013

New organization forms

- Adhocracy
- Wikicracy
- Delegative democracy
- Stigmergic technology
- Outsourced cognition
- User interfaces as political tools

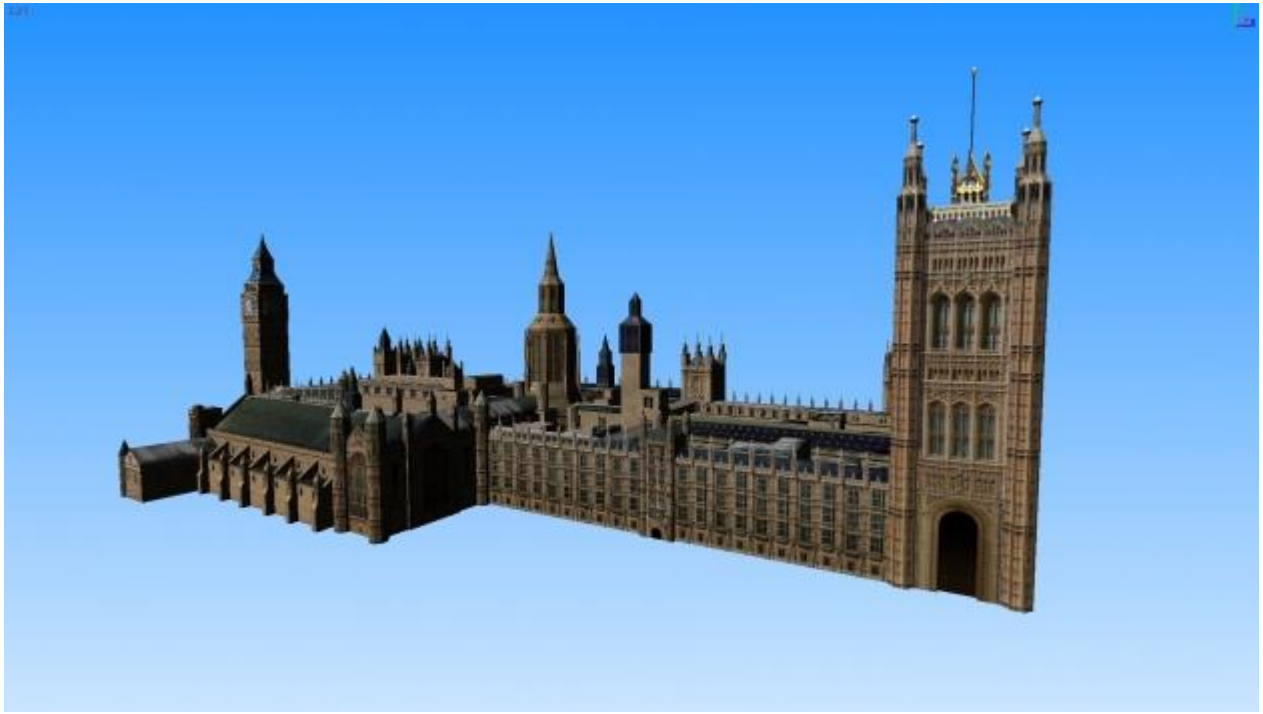


Loomio



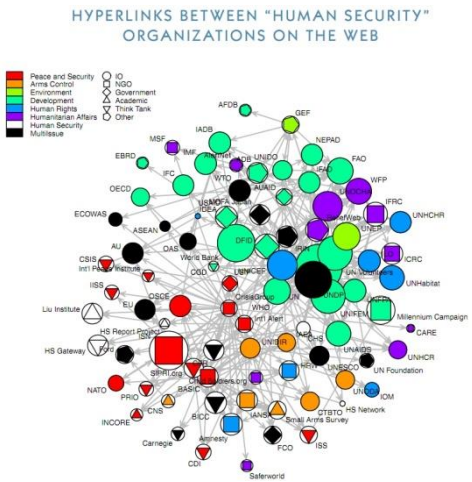
Liquidfeedback

Where will we be in 2025?



Better agenda setting

- Big data: good for finding answers, bad at asking questions (and who has access?)
- Simulation: show that there is an issue
- Large diverse groups: more chance of noticing issues, more competition in agenda setting



Better policy formulation

- Simulation: test the policy options
- Large diverse groups: more policy options
- New tools for decisionmaking
- Wikidemocracy

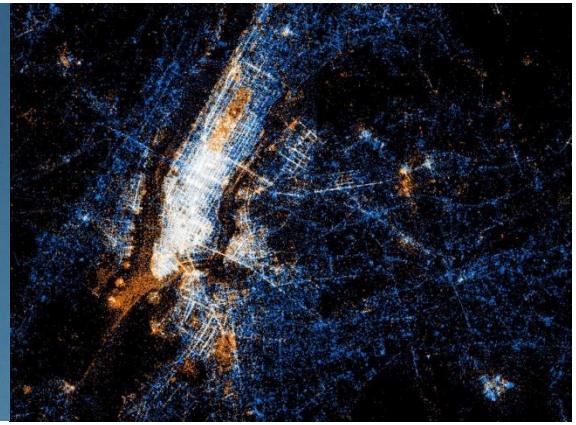
Better decisionmaking

- Large diverse groups: big challenge to legitimacy and tracking responsibity
- Coalitions of the working
- Decision tools?
  - Debiasing
  - Combining preferences
  - Rights implemented as code?

Better implementation

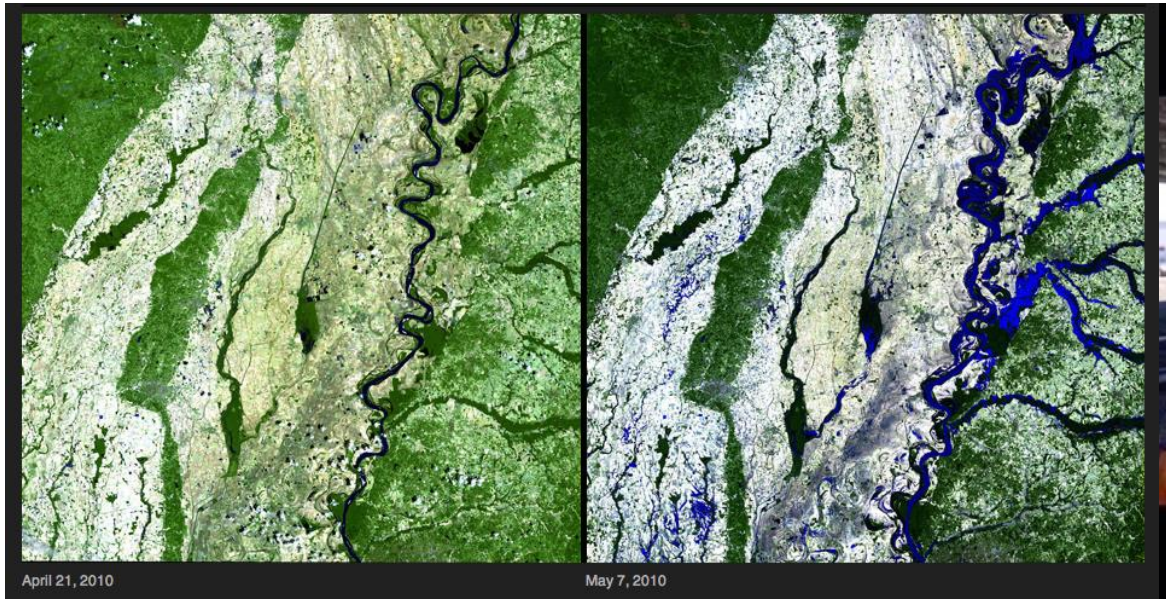
- Outsourcing
- Futarchy (Hanson): outsource implementation completely, leave policymakers to set values and goals
- Automated policy implementation





Better evaluation

- “Evidence based policy”
- Randomized controlled trials
- Surveillance/big data/sousveilance
- Automatic auditing



Debugging policymaking

- Detecting problems with policymaking is a policymaking problem
- We can not see rare problems unless we test longer than the MTBF (Mean Time Between Failures)
  - Hard to parallelize for society-scale innovations
- Handle malicious users
- Signal-to-noise ratios matter
- Policies have half-lives!
  - Set learning rate carefully