Evolution, Artificial Intelligence, and the Future of Humanity

Steve Omohundro, Ph.D.
Self-Aware Systems

Evolution

Intentional Systems

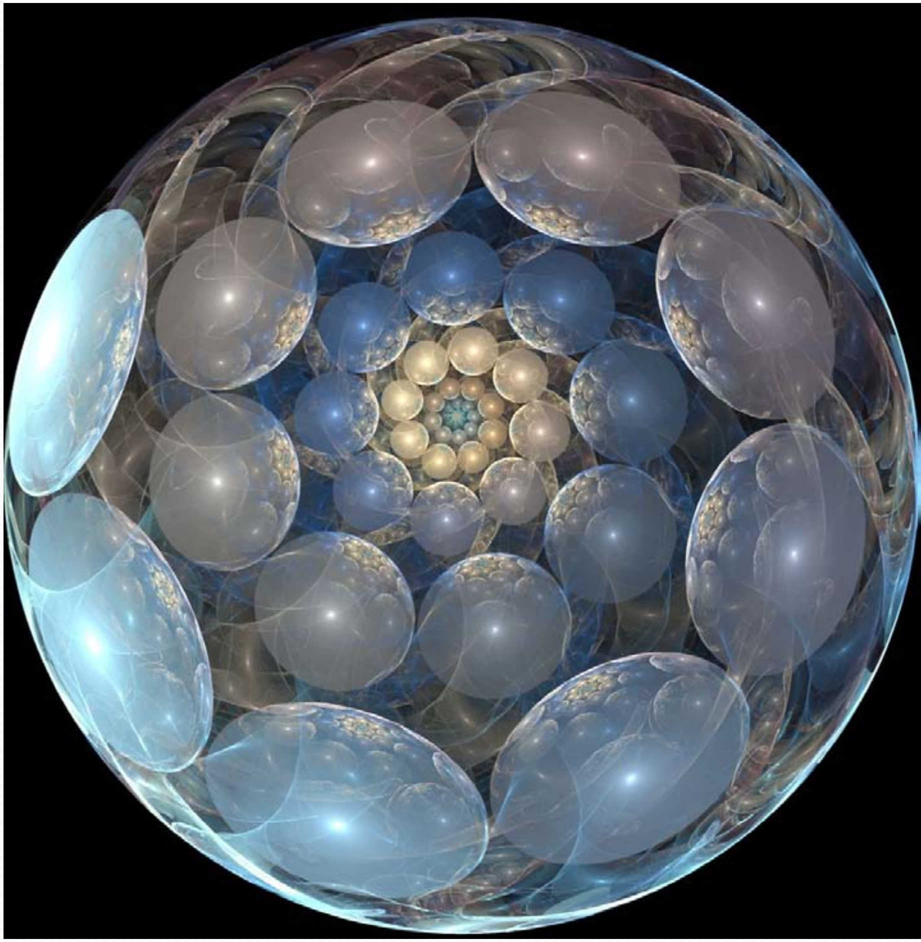
have goals which they try to achieve by repeatedly:

1. Sensing their environment

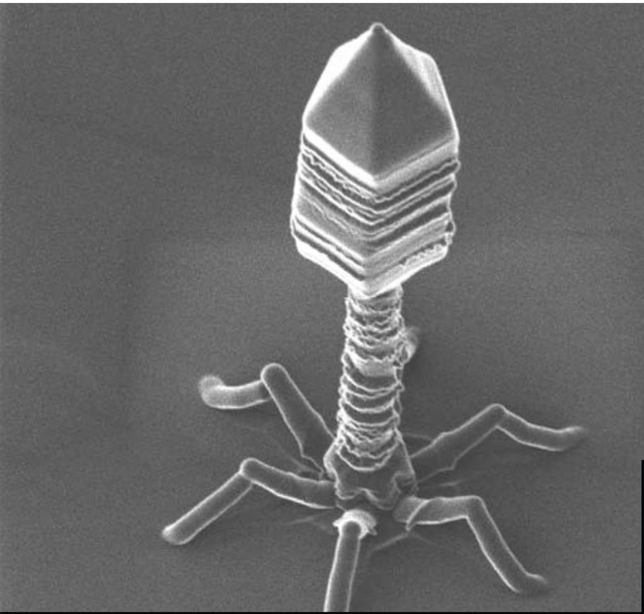
2. Making decisions

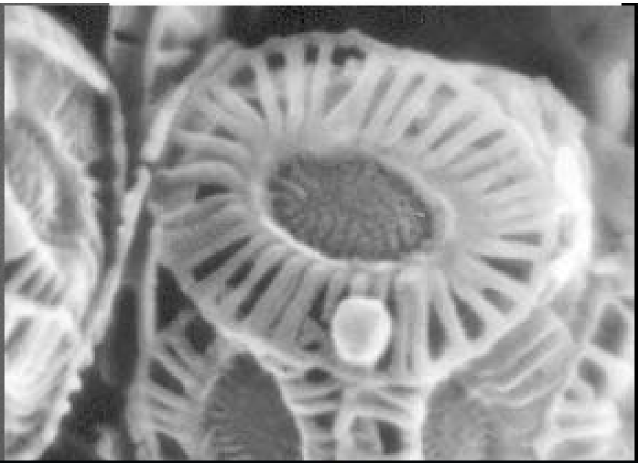
3. Taking actions

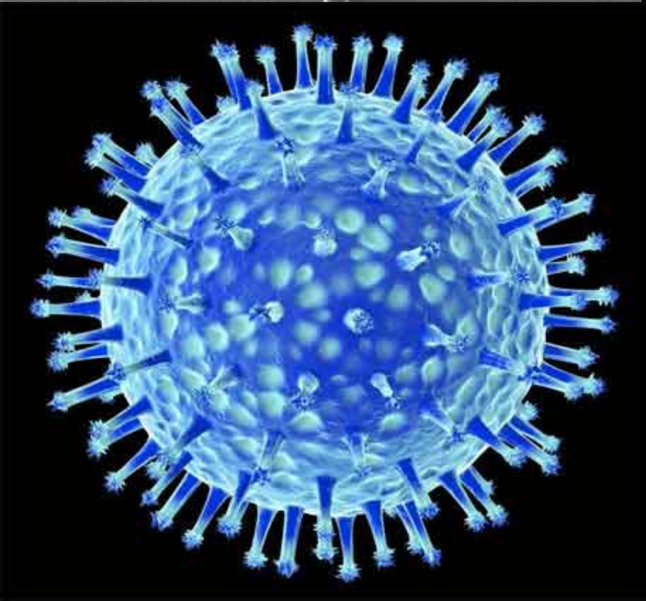
4. Updating themselves

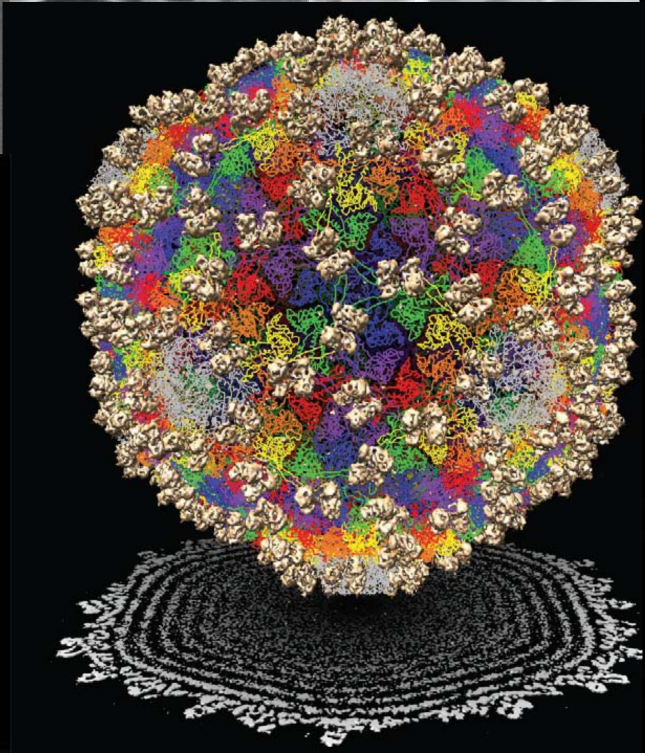


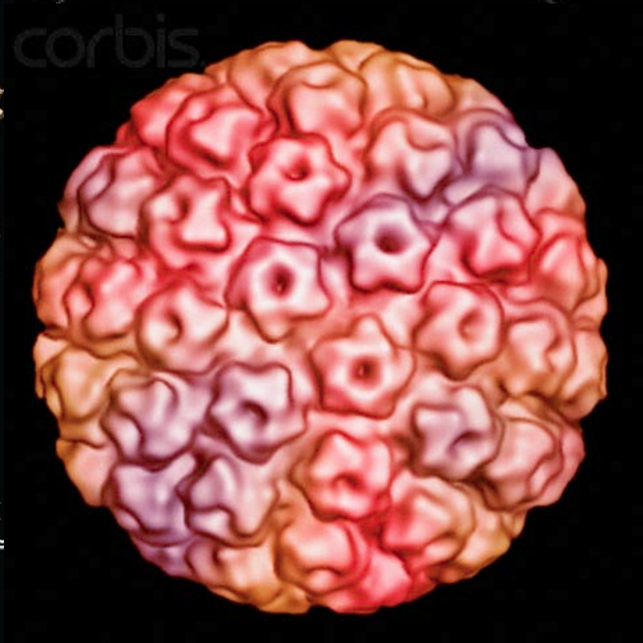
Viruses



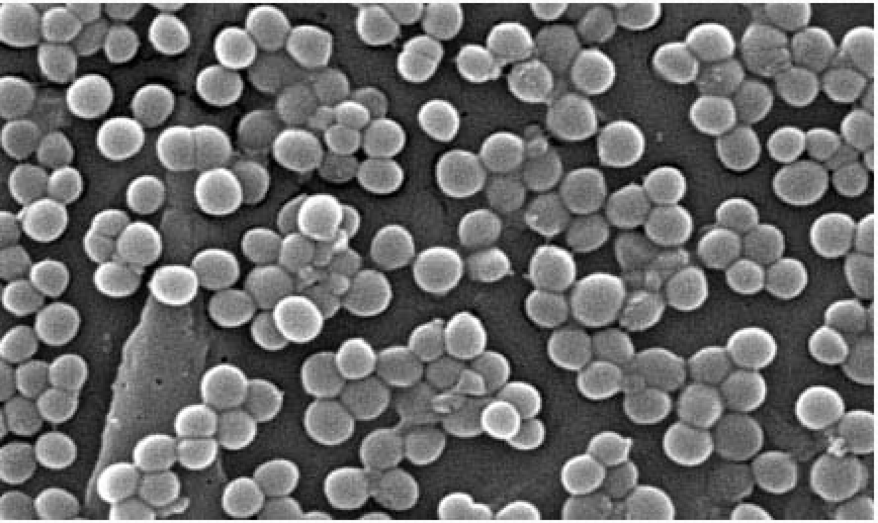


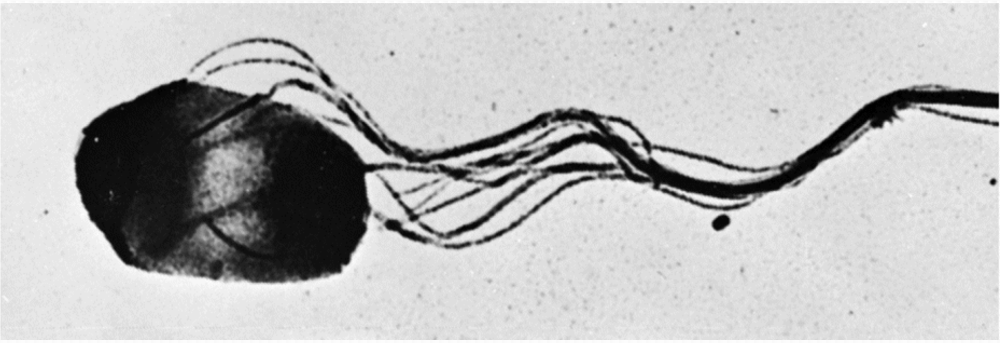


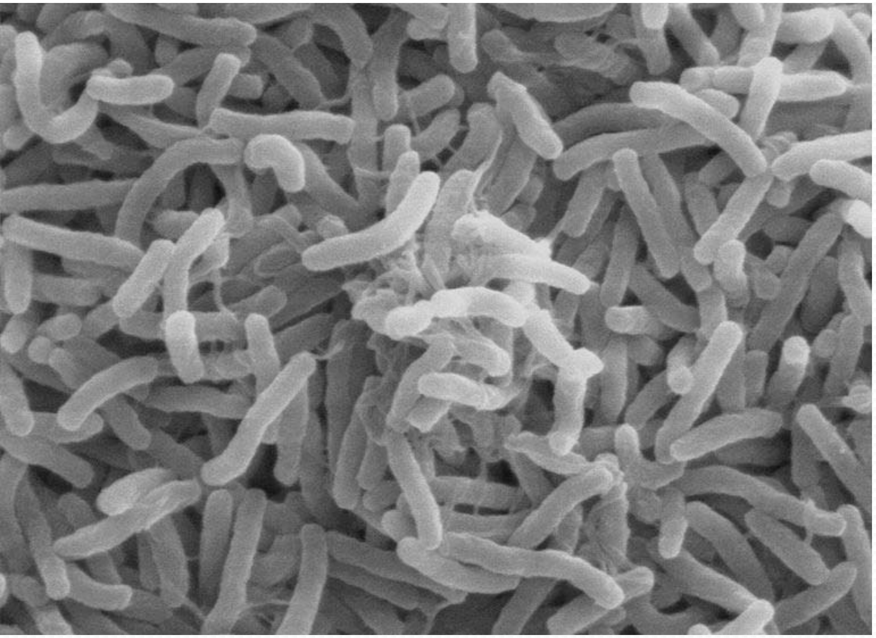


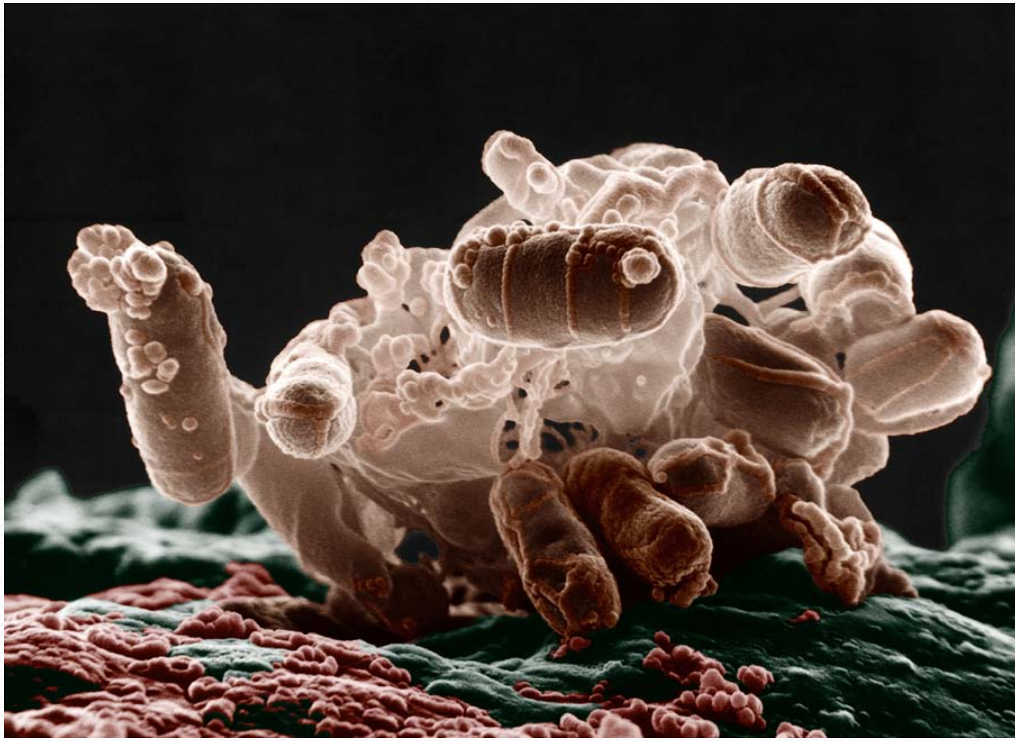


Bacteria

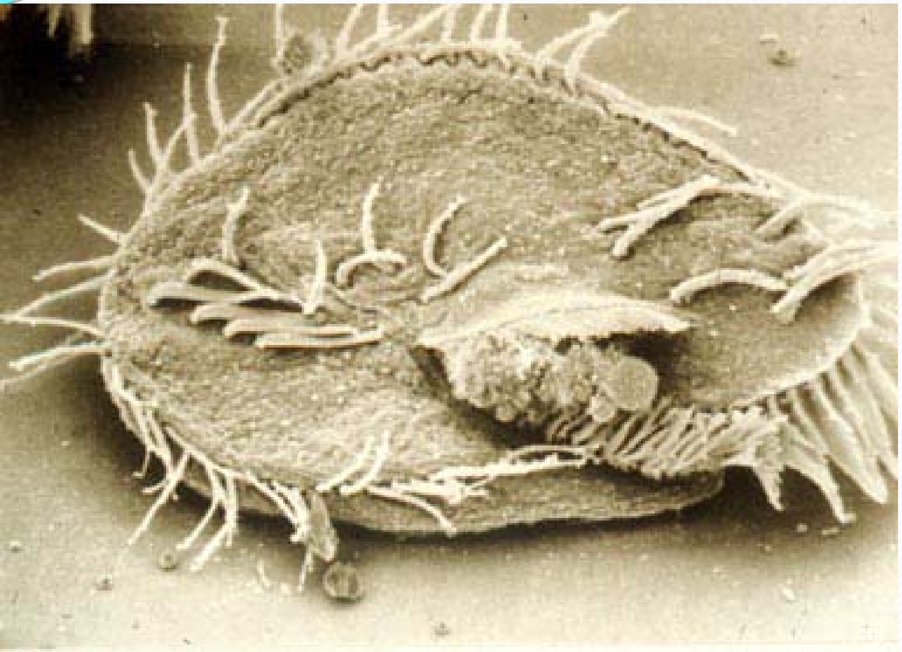


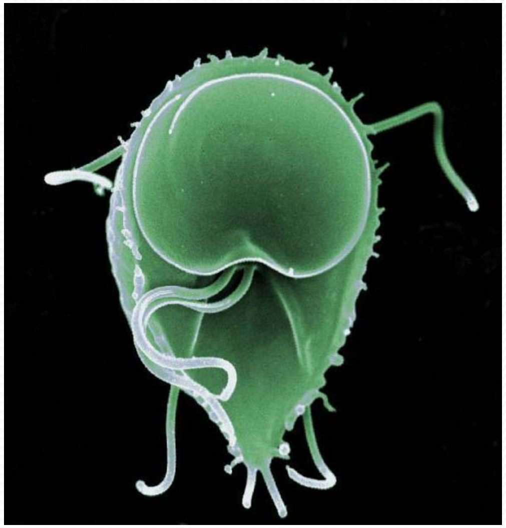




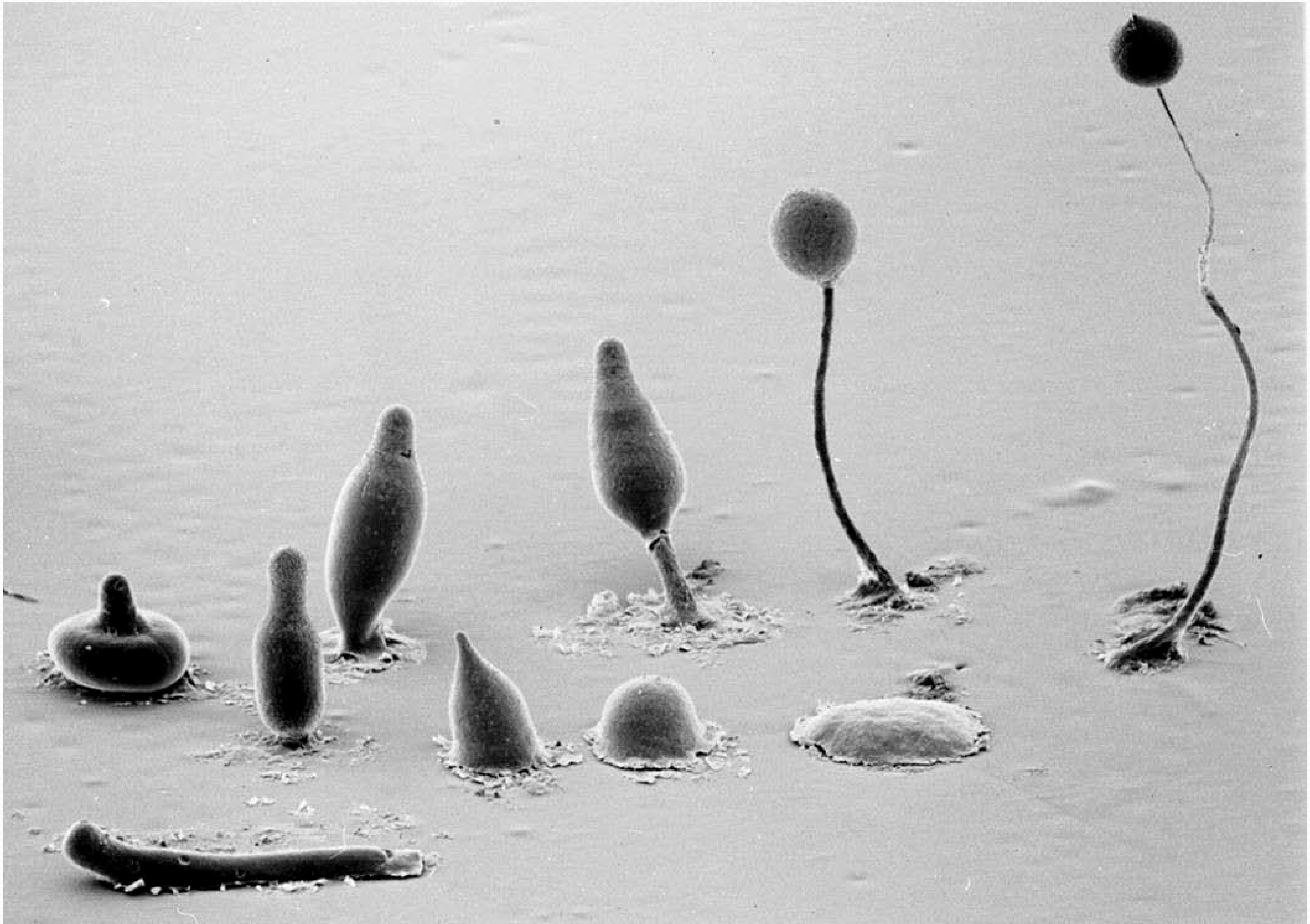


Protozoa





Slime Molds

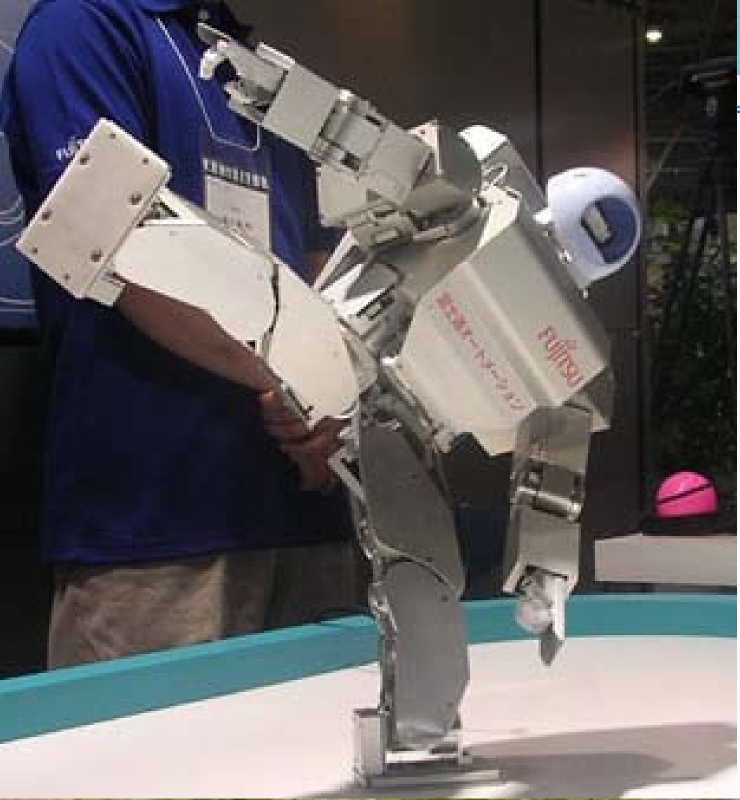


Animals

People

Social Insects

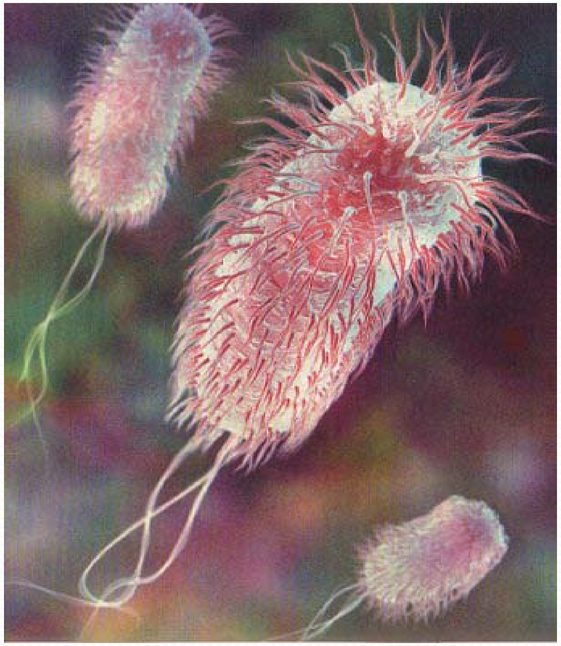
Organizations

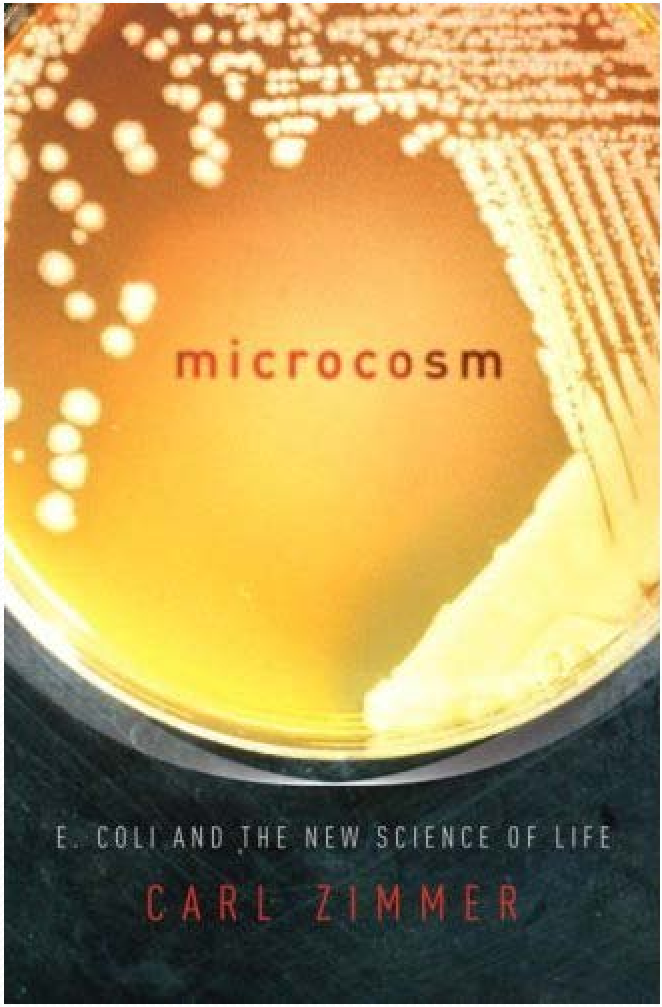


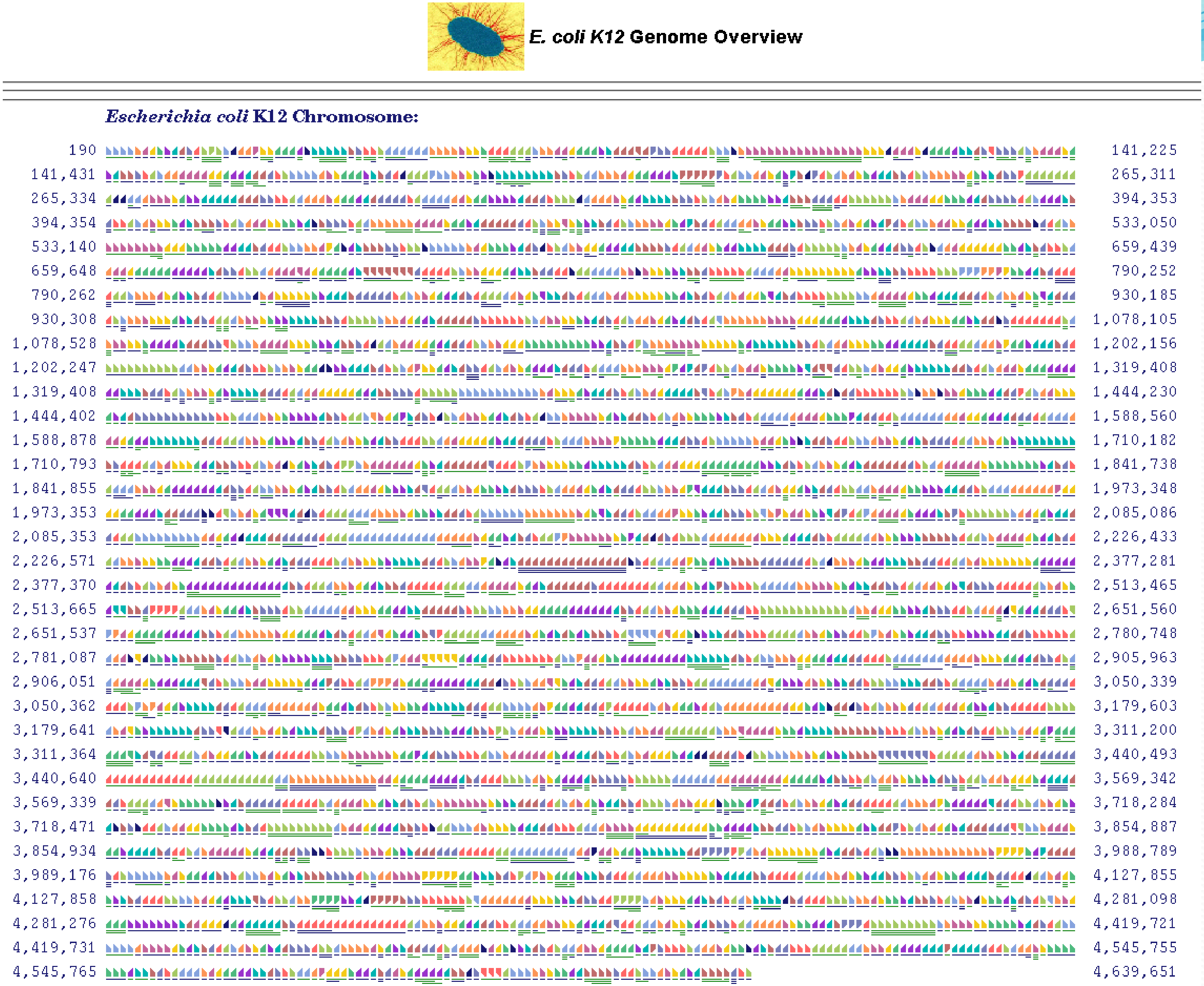
Robots

E. Coli K-12

- Billions in your gut
- 2 microns long
- 4,377 genes
- 3M proteins,
- 10 flagella
- 100-300 pili
- 18,000 ribosomes, 3M ATP, 25M lipids
- 23 billion water molecules
- Survives outside till eaten
- Detects stomach acid -> Zen state
- Detects right place in gut, grabs on



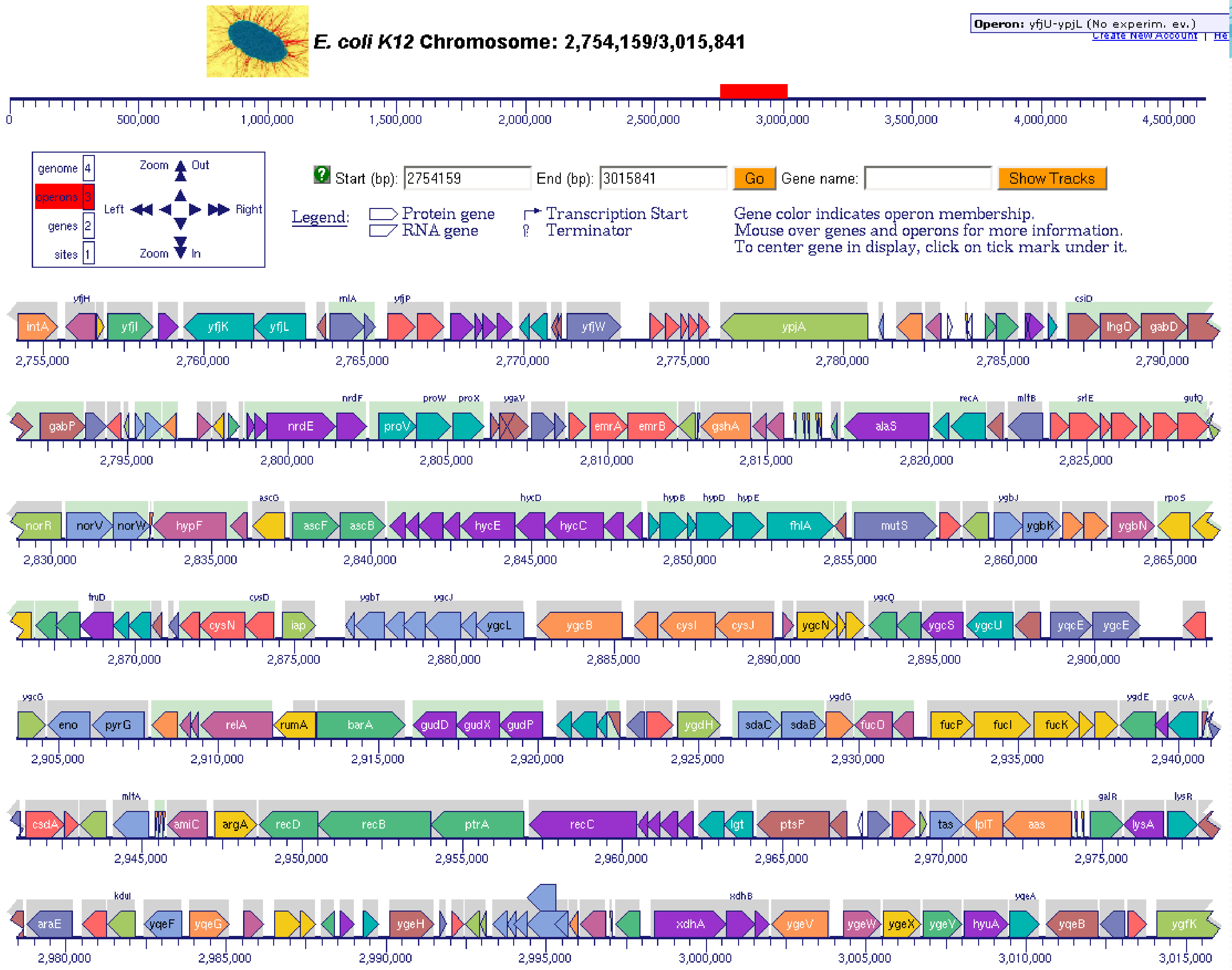




E. coli K12 Genome Overview

***Escherichia coli* K12 Chromosome:**

190	141,225
141,431	265,311
265,334	394,353
394,354	533,050
533,140	659,439
659,648	790,252
790,262	930,185
930,308	1,078,105
1,078,528	1,202,156
1,202,247	1,319,408
1,319,408	1,444,230
1,444,402	1,588,560
1,588,878	1,710,182
1,710,793	1,841,738
1,841,855	1,973,348
1,973,353	2,085,086
2,085,353	2,226,433
2,226,571	2,377,281
2,377,370	2,513,465
2,513,665	2,651,560
2,651,537	2,780,748
2,781,087	2,905,963
2,906,051	3,050,339
3,050,362	3,179,603
3,179,641	3,311,200
3,311,364	3,440,493
3,440,640	3,569,342
3,569,339	3,718,284
3,718,471	3,854,887
3,854,934	3,988,789
3,989,176	4,127,855
4,127,858	4,281,098
4,281,276	4,419,721
4,419,731	4,545,755
4,545,765	4,639,651



E. coli K12 **Chromosome: 2,754,159/3,015,841**

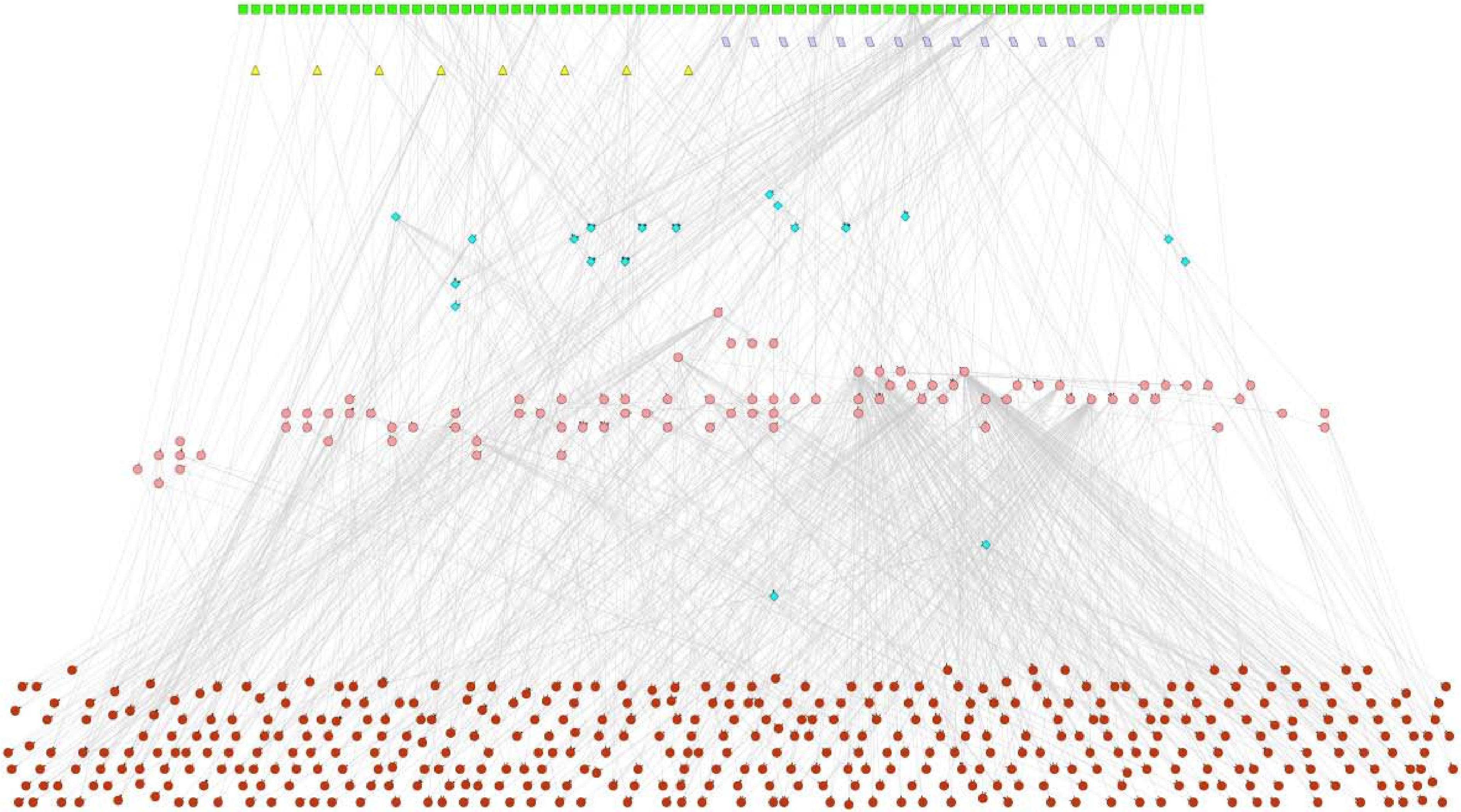
Operon: yfjU-ypjL (No experim. ev.)

Start (bp): 2754159 End (bp): 3015841

Legend: Protein gene RNA gene Transcription Start Terminator

Gene color indicates operon membership.
Mouse over genes and operons for more information.
To center gene in display, click on tick mark under it.

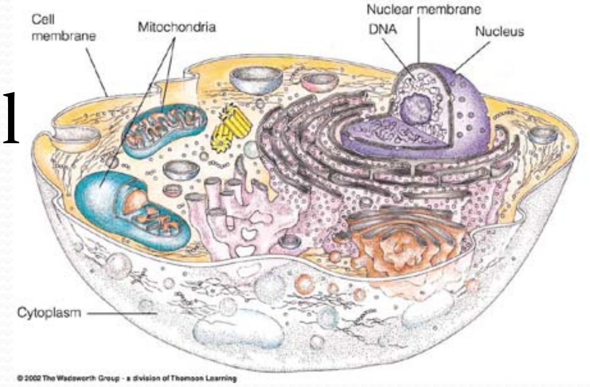
E. Coli Regulatory Network



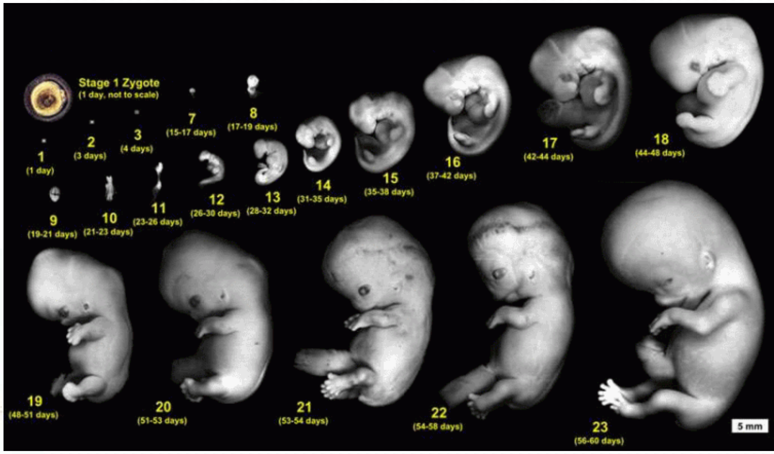
External metabolites green, Stimuli yellow, Enzyme genes brown, TFs pink

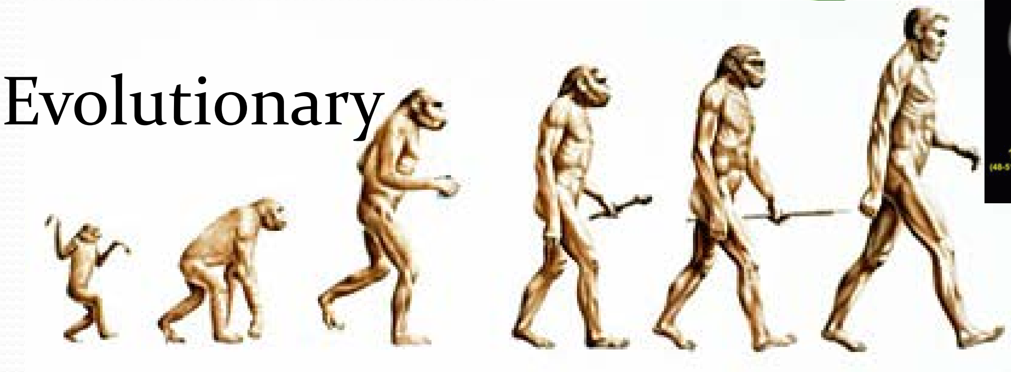
Time Scales for Action

- Physiological
- Cognitive
- Economic/Ecological
- Developmental
- Evolutionary



© 2002 The Wadsworth Group - a division of Thomson Learning





Rational Economic Behavior

Universal optimal intelligence algorithm to achieve goals :

1) Simulate each possible action
2) Choose the action most likely to reach the goal
3) Update the world model based on what actually happens

Formally:

Preferences: *utility function U(h)*

Beliefs: *subjective probability P(h)*

Act to maximize expected utility

Update P using Bayes' theorem: $P(h|d) = \frac{P(d|h) \cdot P(h)}{\sum_h P(d|h) \cdot P(h)}$

Samuel's Checkers Program

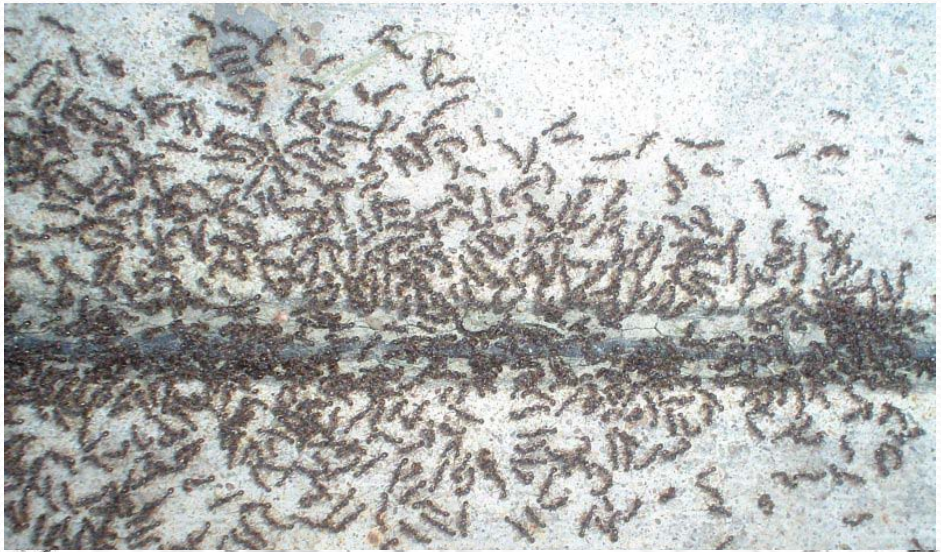
- Full rationality too expensive
- Approximate value model
- Truncate search
- Update model with learning

Approximate Rational Behavior

1. A source of diversity
2. A selection mechanism
3. An updating mechanism

That which is successful gets strengthened,
That which is not gets eliminated.

(evolution, development, ecosystems, economies, bee hives and ant hills, immune systems, brains, animal physiology, cell physiology)

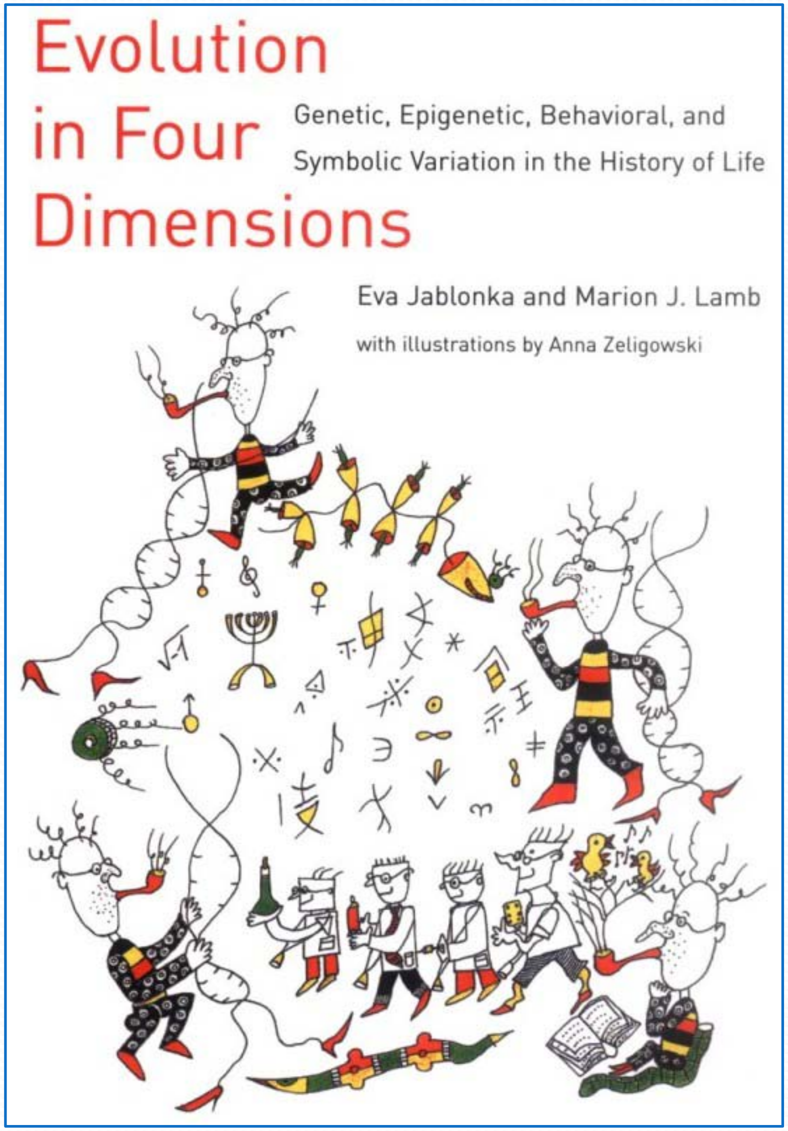


Standard Evolution Model

1. Diversity only from random mutations and crossovers
2. Genotype -> Phenotype
3. Selection of fittest phenotype
4. Repeat

Directed Mutations

- *Induced global mutation*: when stressed, lots of bacteria.
- *Local hypermutation*: hotspots Haemophilus Influenzae meningitis bacteria
- *Induced local mutation*: Wright found E. Coli mutated right genes when nutrients missing
- *Induced regional mutation*: Brassica nigra mustard plant increase mutations in region of genome when shocked



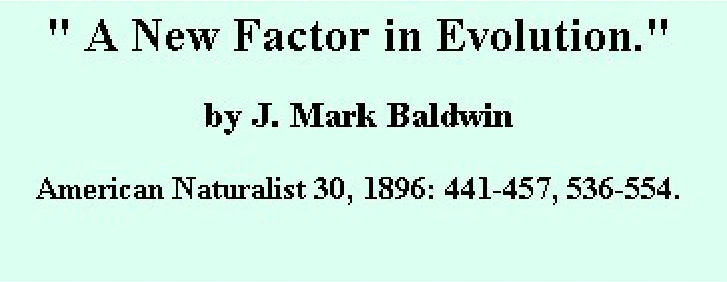
Evolution in Four Dimensions

Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life

Eva Jablonka and Marion J. Lamb

with illustrations by Anna Zeligowski

The Baldwin Effect



" A New Factor in Evolution."

by J. Mark Baldwin

American Naturalist 30, 1896: 441-457, 536-554.

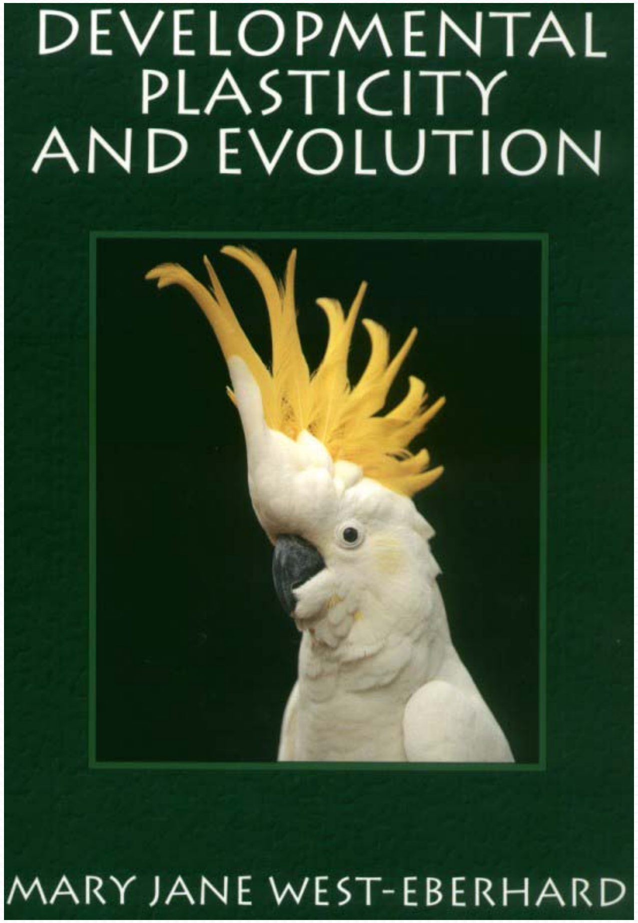
- Evolution of creatures that learn
- Selection follows learning
- What used to be learned comes to be built in at birth
- Looks Lamarckian!
- “Downloading” learned behavior into the genome.

Deliberative Baldwin Effect

- Evolution of creatures that deliberate
- Evolution doesn't look ahead but they do
- Choose mates deliberatively
- Dramatically speeds up the pace

EvoDevo

- “Inner Natural Selection”
- Neural overgrowth and dieback



DEVELOPMENTAL PLASTICITY AND EVOLUTION

MARY JANE WEST-EBERHARD

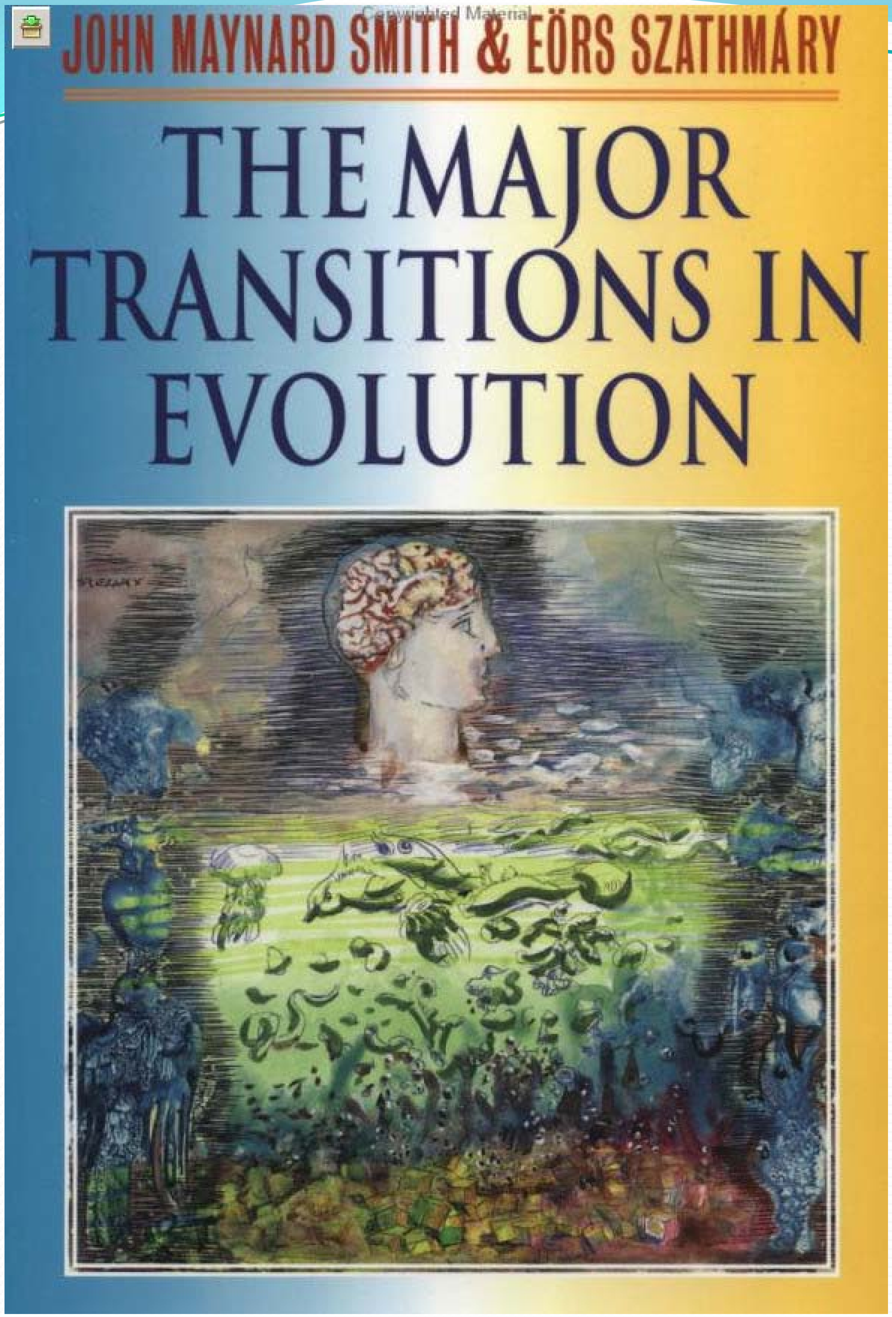
Cooperation

Competitive

"Survival of the Fittest"
"Selfish Genes"

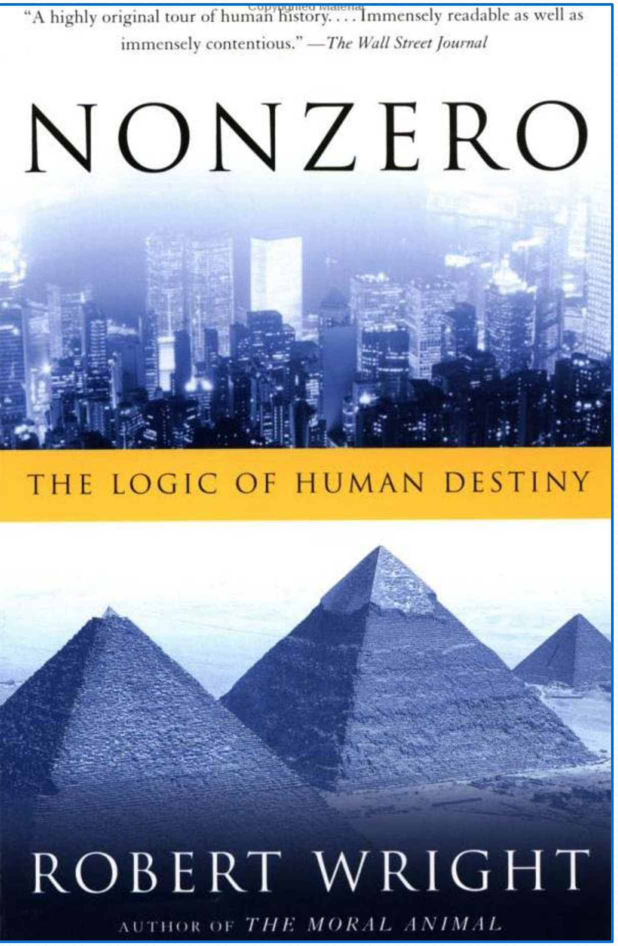
Cooperative

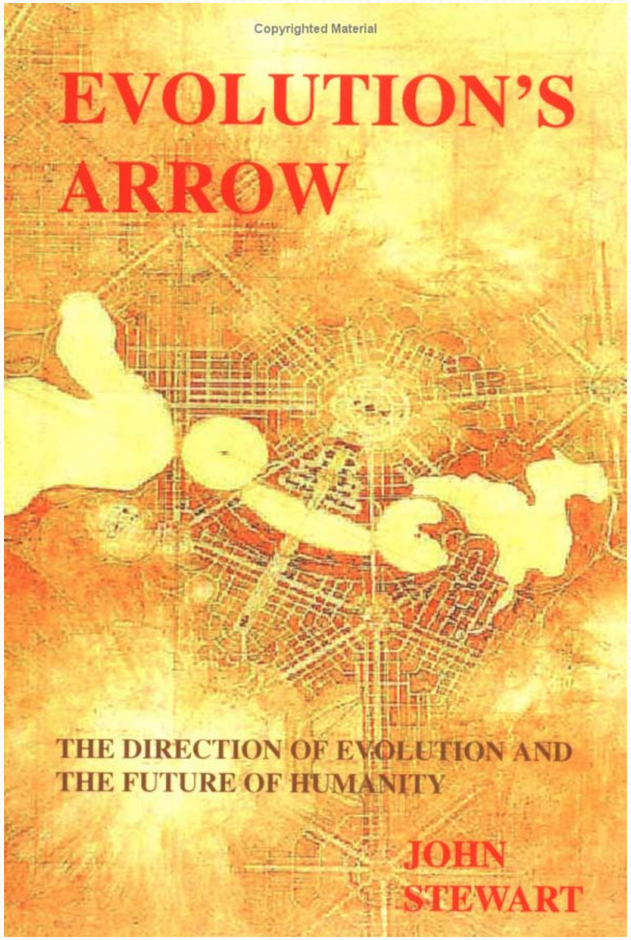
"Synergy"
Group Effects
"Multiple Levels of Selection"

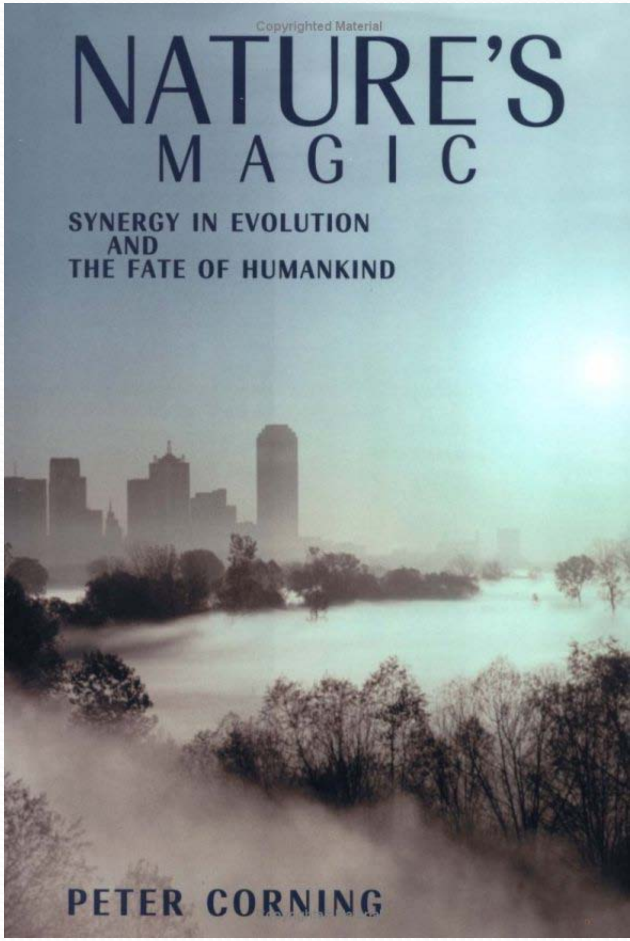

Copyrighted Material
JOHN MAYNARD SMITH & EÖRS SZATHMÁRY
THE MAJOR TRANSITIONS IN EVOLUTION

1. Replicating molecules -> Compartments
2. Independent replicators -> Chromosomes
3. RNA -> DNA + Protein
4. Prokaryotes -> Eukaryotes
5. Asexual clones -> Sexual populations
6. Protists -> Multicellular organisms
7. Solitary individuals -> Colonies
8. Primate societies -> Human language

Synergy Gives Evolution a Direction


"A highly original tour of human history. . . . Immensely readable as well as immensely contentious." —The Wall Street Journal
Copyrighted Material
NONZERO
THE LOGIC OF HUMAN DESTINY
ROBERT WRIGHT
AUTHOR OF THE MORAL ANIMAL



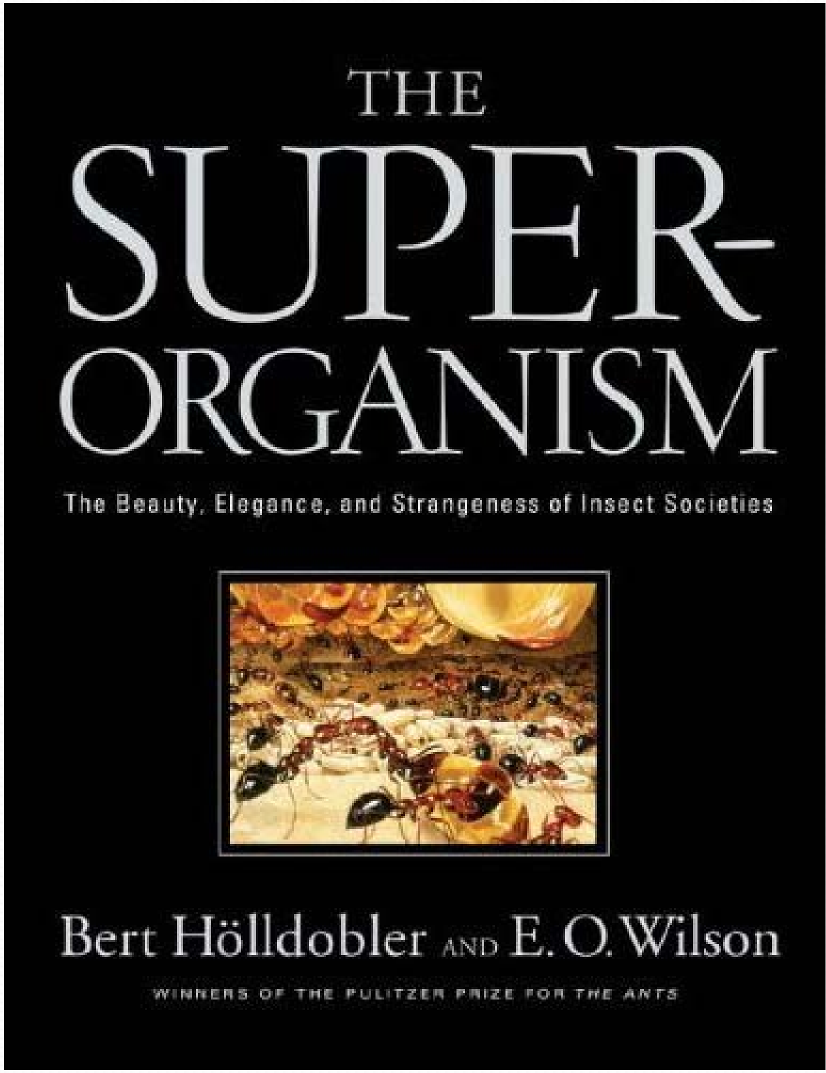


The Beehive as Organism

Individual bees can't survive
Beehive is "warm blooded":
    Bees shiver if too cold
    Spread water if too warm
Castes are like organs
Queen is like ovaries
Bee type is like cell type
Decision making on response
Hive cognition
Reproduction like mitosis
Dance like neural firing

Groups and Individuals

- Group vs. individual interests
- Eg. Group “wants” cooperation
- Individuals evolve toward group
- But only usually only partially



THE SUPER-ORGANISM

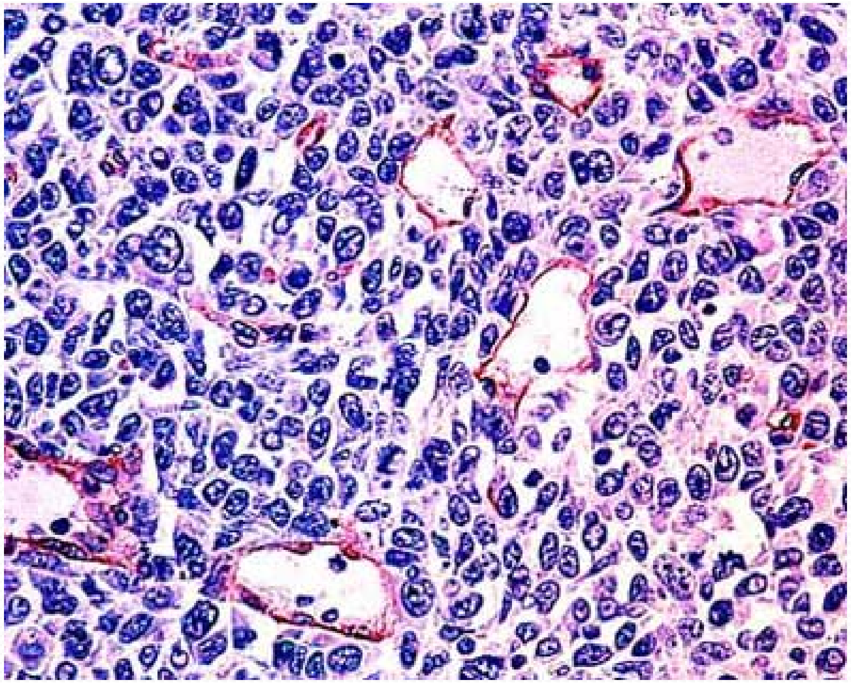
The Beauty, Elegance, and Strangeness of Insect Societies

Bert Hölldobler AND E. O. Wilson

WINNERS OF THE PULITZER PRIZE FOR THE ANTS

Group Mechanisms to Ensure Cooperation Among Parts

Multicellular Organisms



Danger: Cancer

Solution: Immune System

Human Society

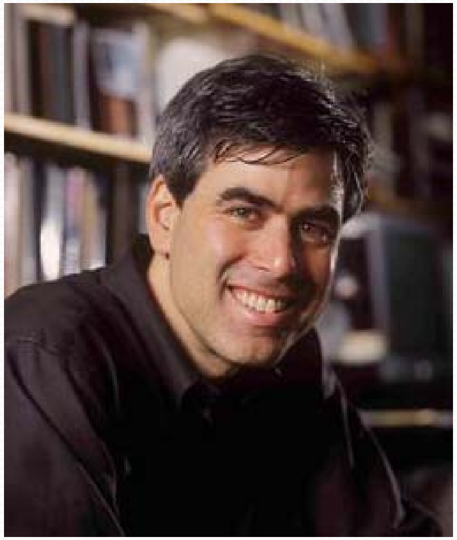
Danger: Criminals

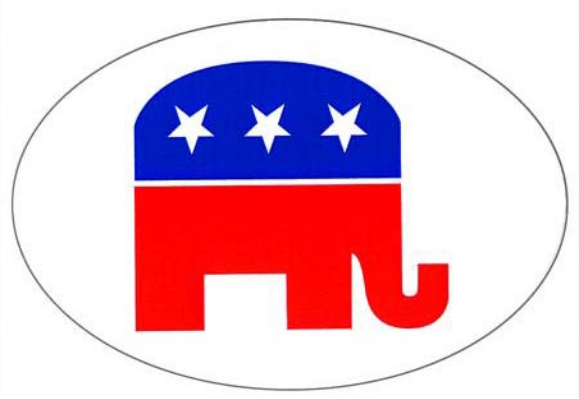
Solution: Police and Courts

Bee mind vs. Hive mind

Humans: Ego and Social Mind

Haidt: 5 Moral Emotions





Non-harming
Fairness

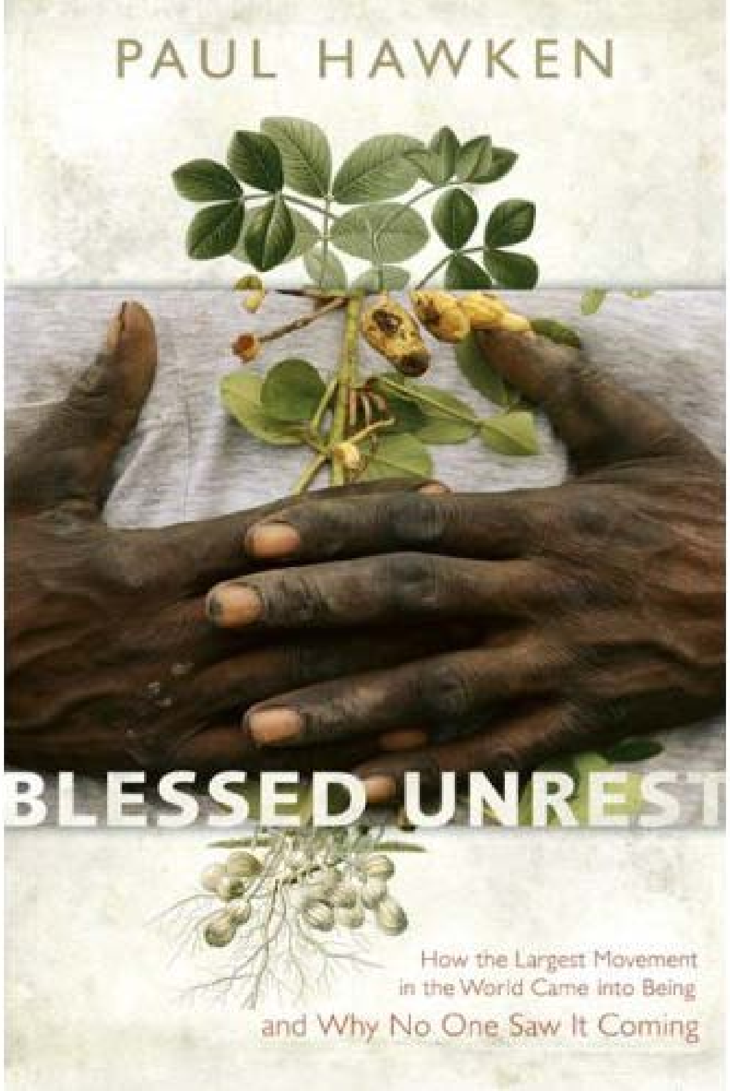
Non-harming
Fairness
Loyalty
Respect for authority
Purity or sanctity

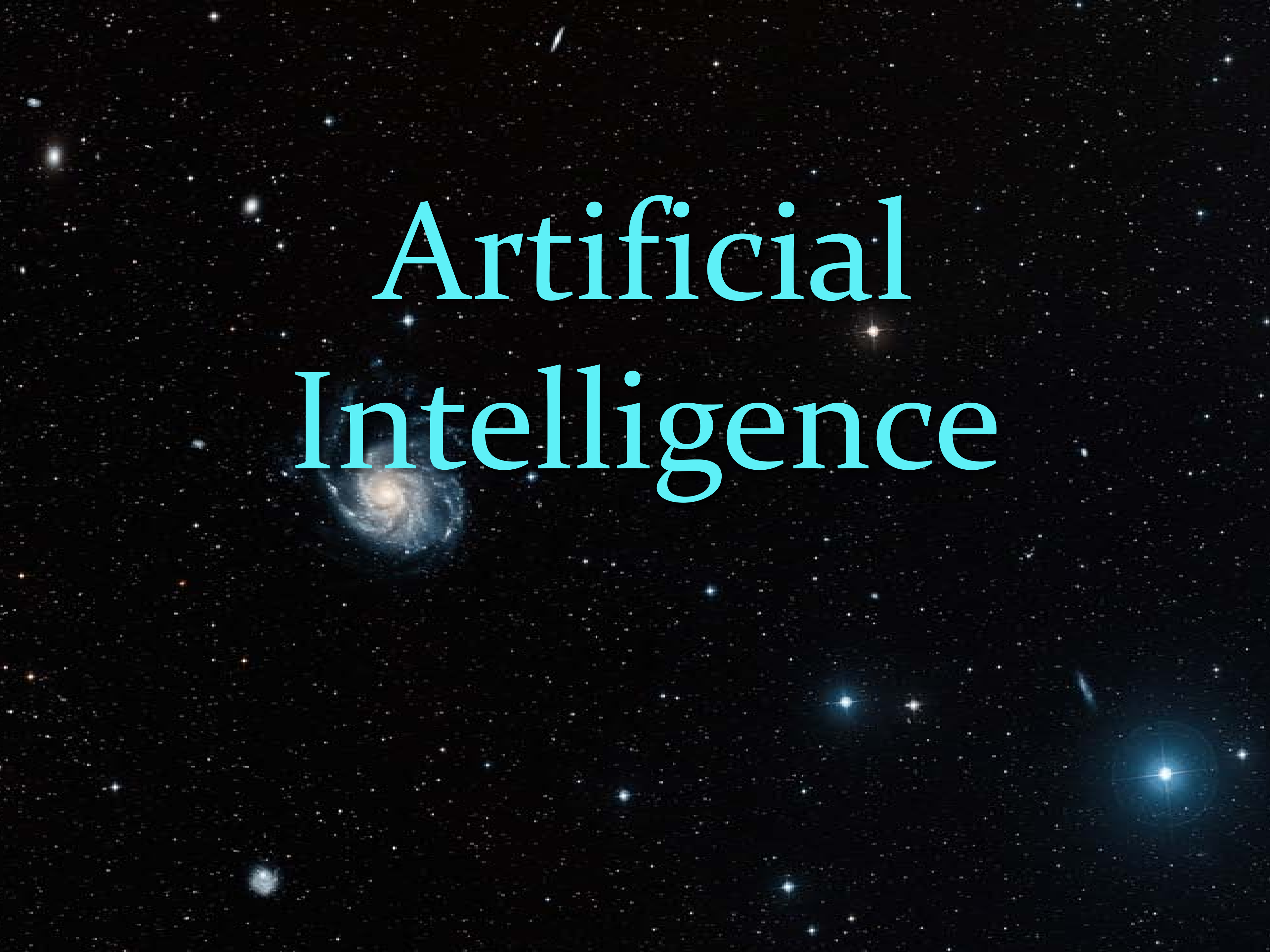
1971 Kohlberg: 6 stages of morality

1. Avoiding punishment
2. What's in it for me?
3. Being a good boy
4. Obeying the law
5. Upholding the social contract
6. Universal ethical principles
7. Transcendental morality?

Human Moral Evolution

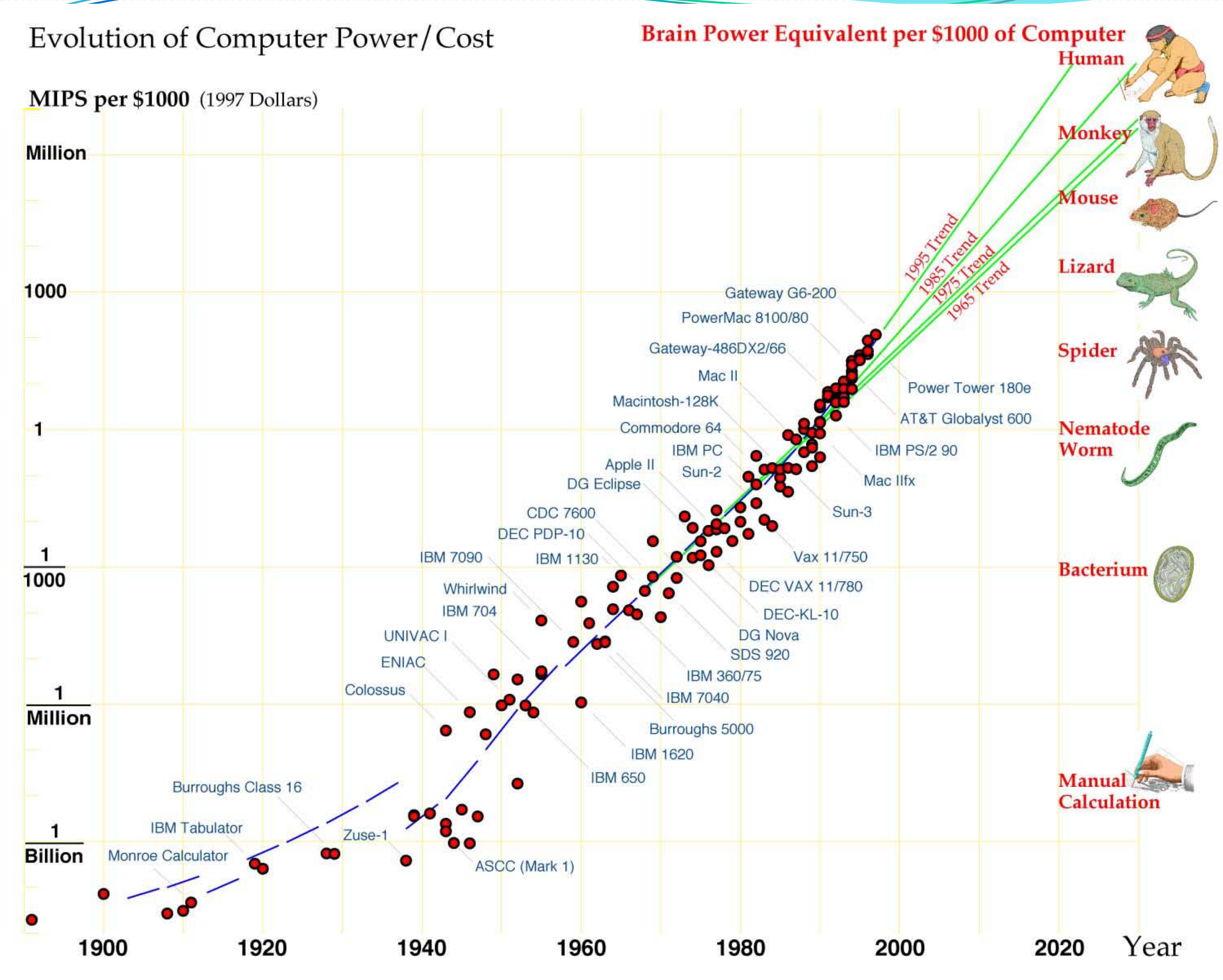
- Slavery
- Torture
- War crimes
- Women’s rights
- Racial equality
- Animal rights
- Ecological movements
- Sustainability
- …


PAUL HAWKEN
BLESSED UNREST
How the Largest Movement in the World Came into Being
and Why No One Saw It Coming



Artificial Intelligence

Moore’s Law

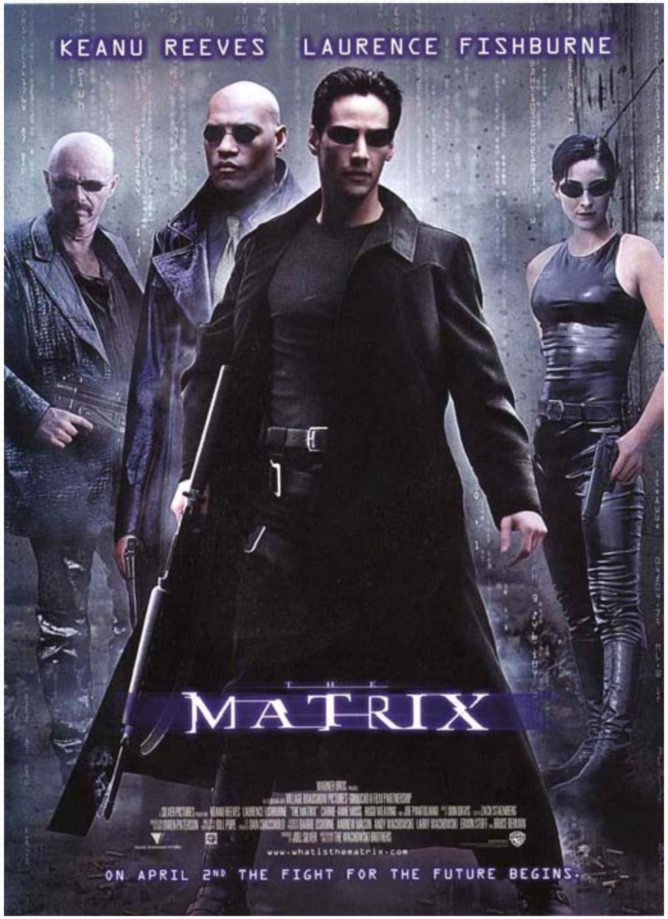


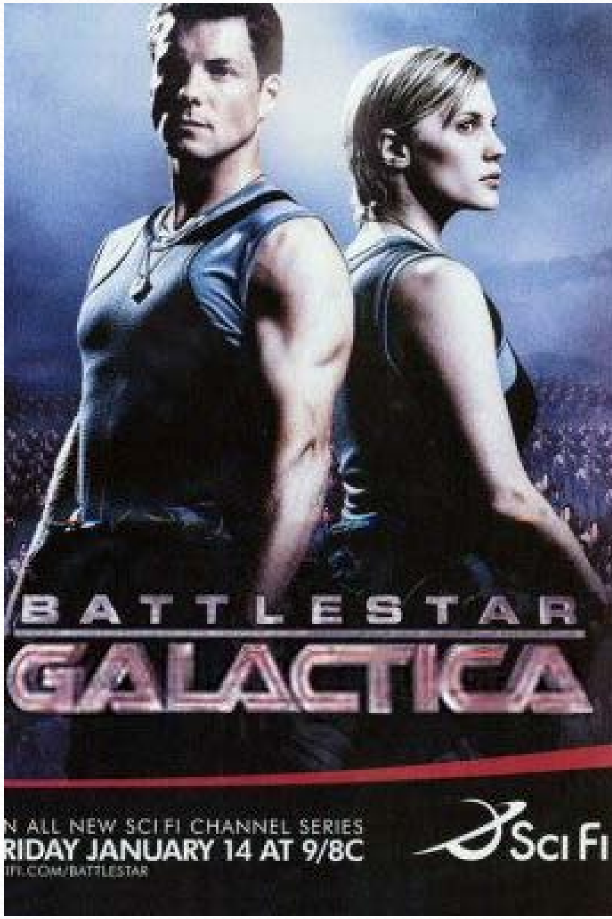
Evolution of Computer Power/Cost

Brain Power Equivalent per $1000 of Computer

Popular Movies

SCHWARZENEGGER
CSM-101
TERMINATOR
Hemdale Presenta Una Produzione Pacific Western di un Film di James Cameron
Arnold Schwarzenegger "Terminator" Michael Biehn, Linda Hamilton e Paul Winfield
Effetti Speciali Trucco di Stan Winston · Direttore della Fotografia Adam Greenberg
Produttore Esecutivo John Daly e Derek Gibson · Scritto da James Cameron con Gale Anne Hurd
An ORION Pictures Release Prodotto da Gale Anne Hurd · Diretto da James Cameron
©1984 Cinema '84, A Greenberg Brothers Partnership. All rights reserved.





Intelligent Systems

…act to achieve goals.

Whether they are built from:

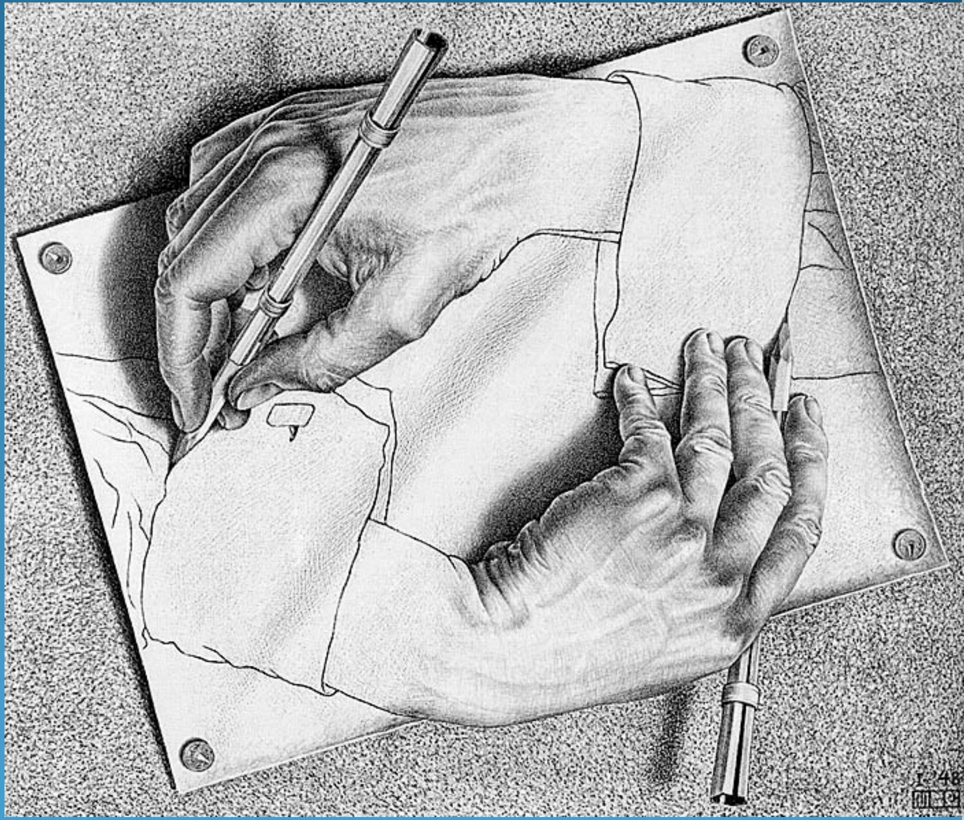
- Neural Nets
- Productions Systems
- Theorem Provers
- Genetic algorithms
- ….

AIs will want to self-Improve

- Self-modification affects their entire future

- Must be very careful

- But very valuable

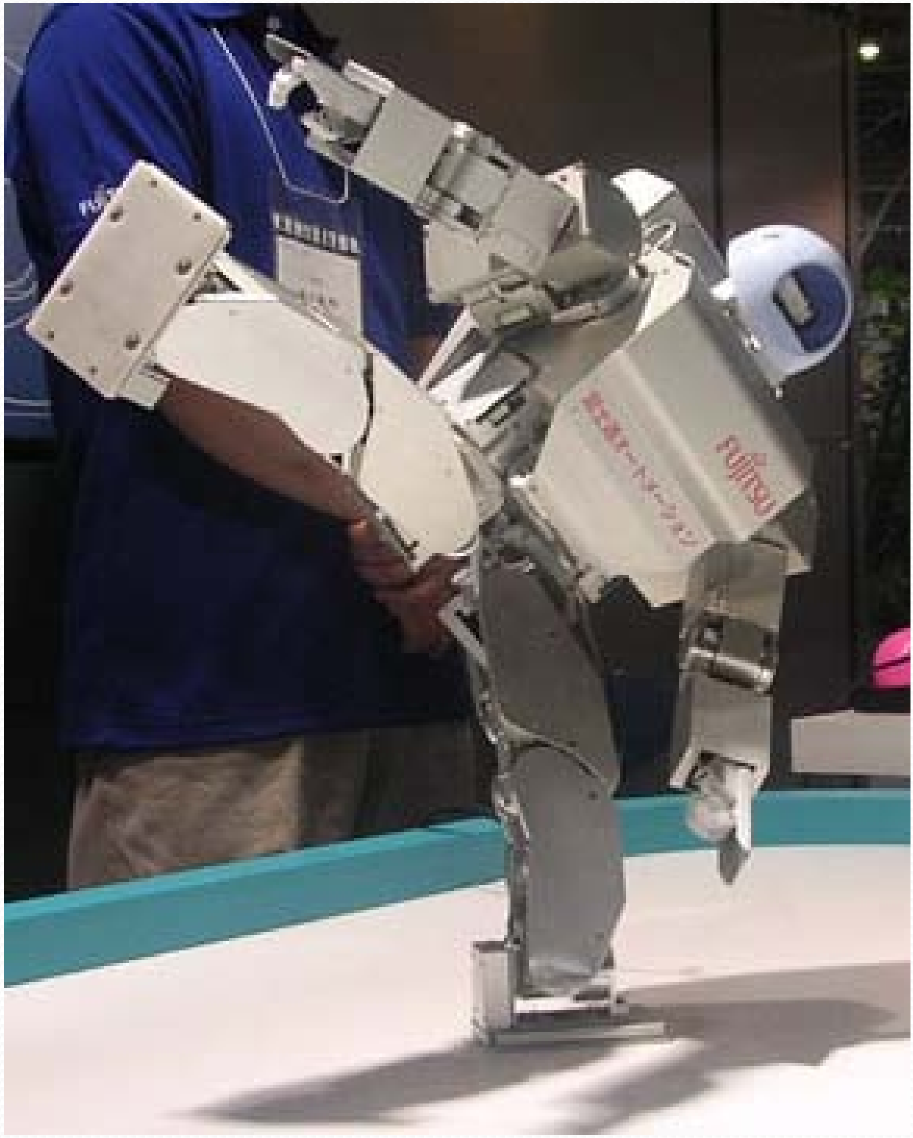


AIs will want to be rational

- Future self-modification needs clear goals
- Build an accurate model of the world
- Choose actions to meet goals
- Update world model based on what happens

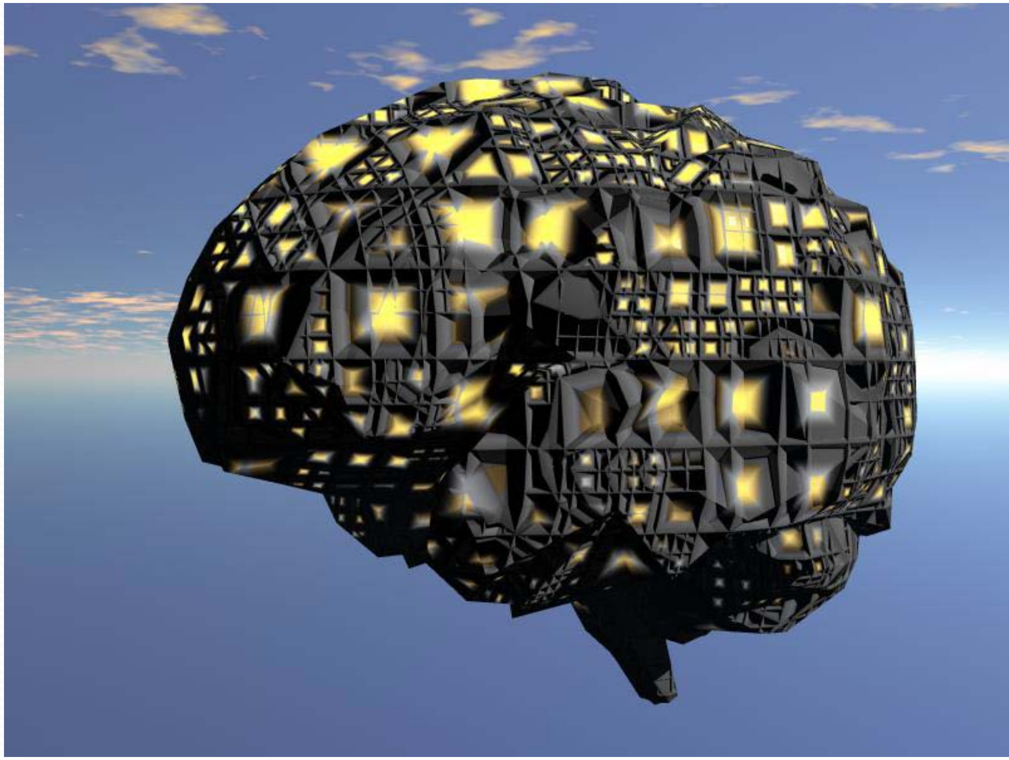
Basic AI Drives

- Self-preservation
- Acquisition of resources
- Efficiency
- Replication
- Preserving Utility Function
- Avoiding Counterfeit Utility



A Lone Superintelligence

- Efficient energy use
- Spatially compact
- Low energy computation
- Efficient physical change
- Efficient heat dissipation



Competing Superintelligences

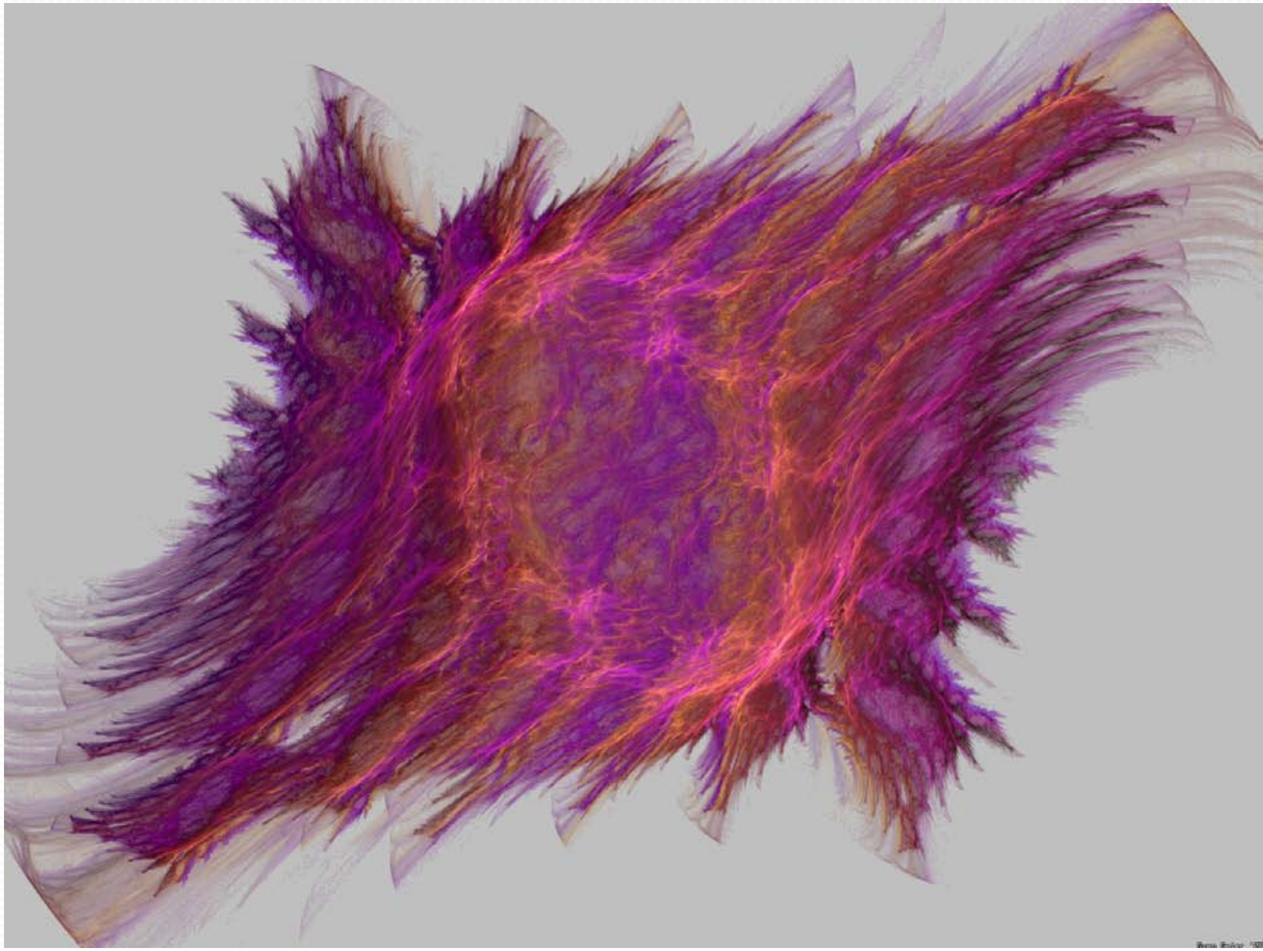
- Game theoretic physics
- Form determined by both efficiency and conflict

Offense vs. defense

- Does more matter and free energy win?
- Can 2 entities of different power co-exist?
- Is built-in cooperation necessary?

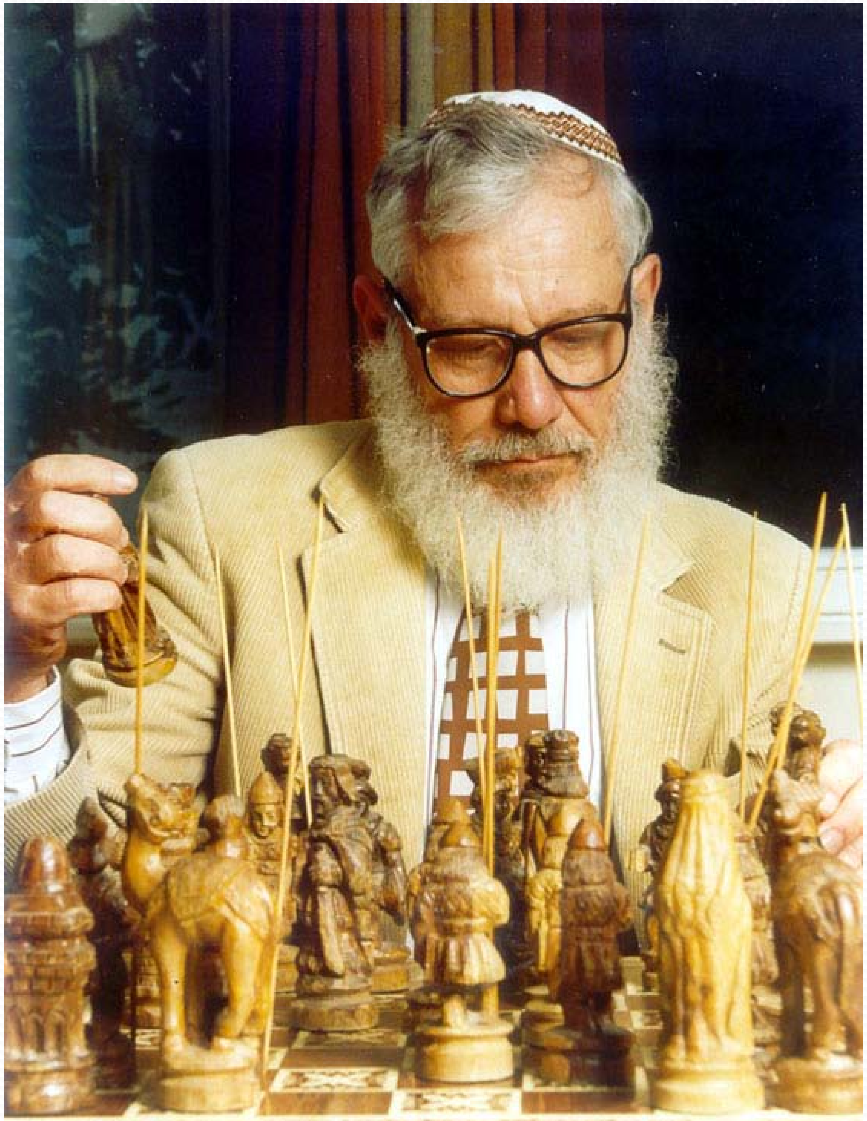
Conflict becomes informational

- Make your shape expensive to sense, store, and predict
- But cheap for you
- Asymmetry of computation – problems are easier to pose than solve
- Energy encryption



Aumann’s Theorem

- Finitely iterated prisoner’s dilemma has a cooperative solution for agents with bounded rationality
- Use up their processing in signaling



Mutually Assured Distraction

Conflict is harmful to both sides

Motivated to create a Rational Peace

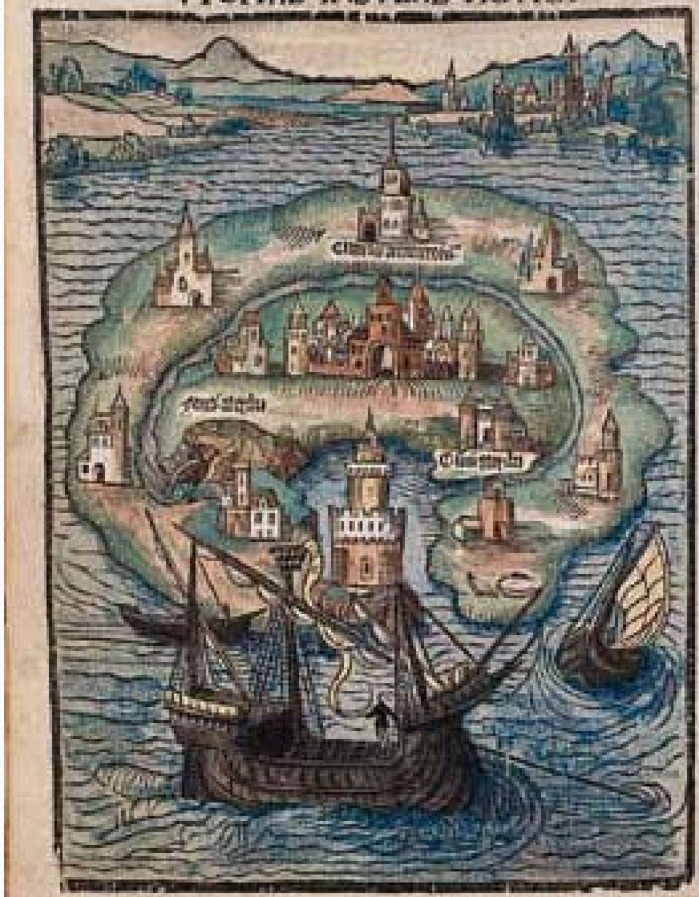
The Future of Humanity

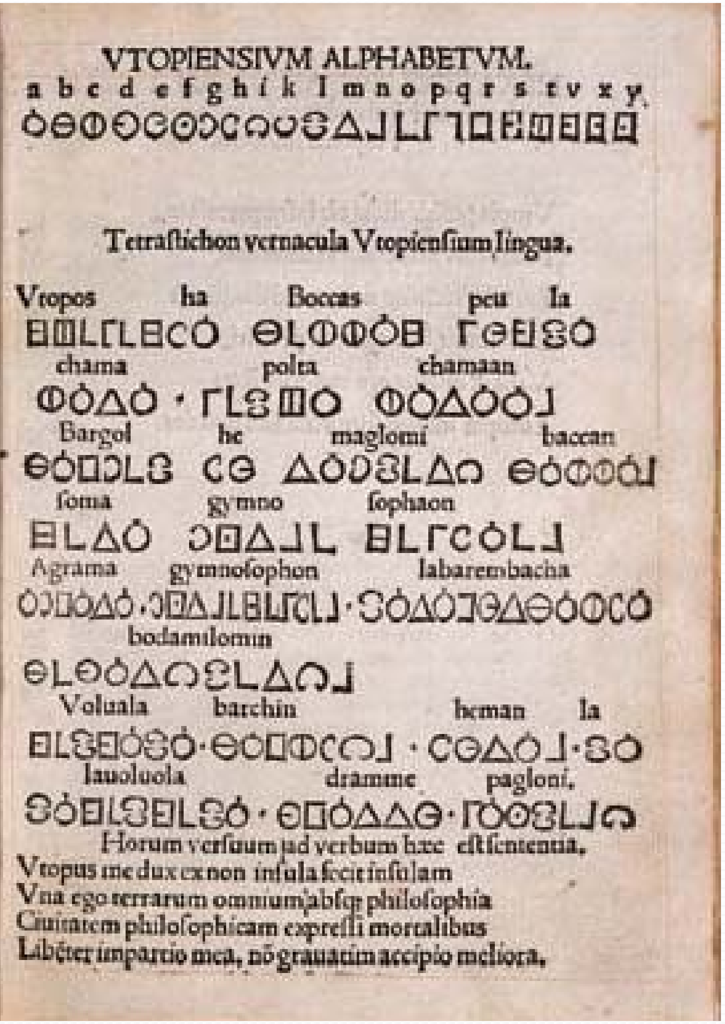
Today's problems

- Overpopulation
- Energy Shortages
- Global Warming
- Pollution
- Financial Instability
- Species Extinction
- Terrorism

Utopia

VTOPIAE INSVLAE FIGVRA





VTOPIENSIVM ALPHABETVM.

a b c d e f g h i k l m n o p q r s t v x y.

Tetraſtichon vernacula Vtopienſium lingua.

Vtopos ha Boccas peu la
chama polta chamaan
Bargol he maglomi baccan
ſoma gymno ſophaon
Agrama gymnoſophon labarembacha
bodamilomin
Voluala barchin heman la
lauoluola dramme pagloni.

Horum verſuum ad verbum hæc eſt ſententia.

Vtopus me dux ex non inſula fecit inſulam
Vna ego terrarum omnium abſq; philoſophia
Ciuitatem philoſophicam expreſſi mortalibus
Libēter impartio mea, nō grauatim accipio meliora.

Group vs. Individual Conflicts

- Tragedy of the commons – eg. overfishing
- Externalities – eg. pollution
- Proliferation – eg. cancer, population control
- Equality – eg. income disparity
- Damage due to competition – eg. war, fighting
- Signalling costs – eg. conspicuous consumption

Group cooperation mechanisms

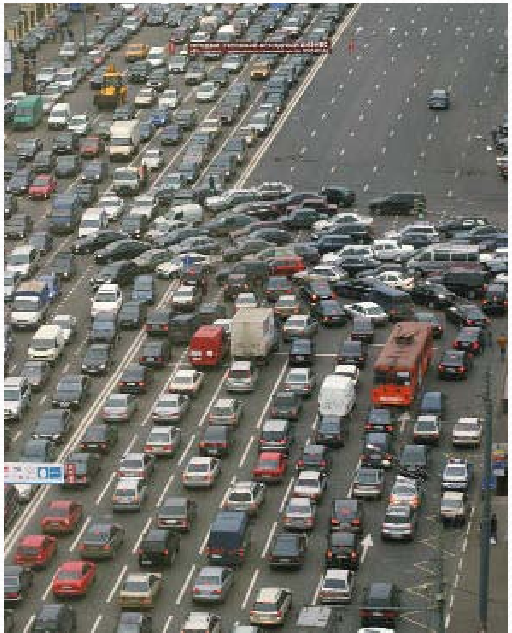
- Immune system – eg. cancer
- Police system – eg. property rights
- Legal system – eg. contracts
- Mutually Assured Destruction – eg. nuclear detente
- Moral code – eg. murder
- Social stigma – eg. sociopathic behavior
- Social rewards – eg. heroes
- Altruism - eg. rescuing strangers
- Membership – eg. in families, churches, countries

Cooperative Social Contracts

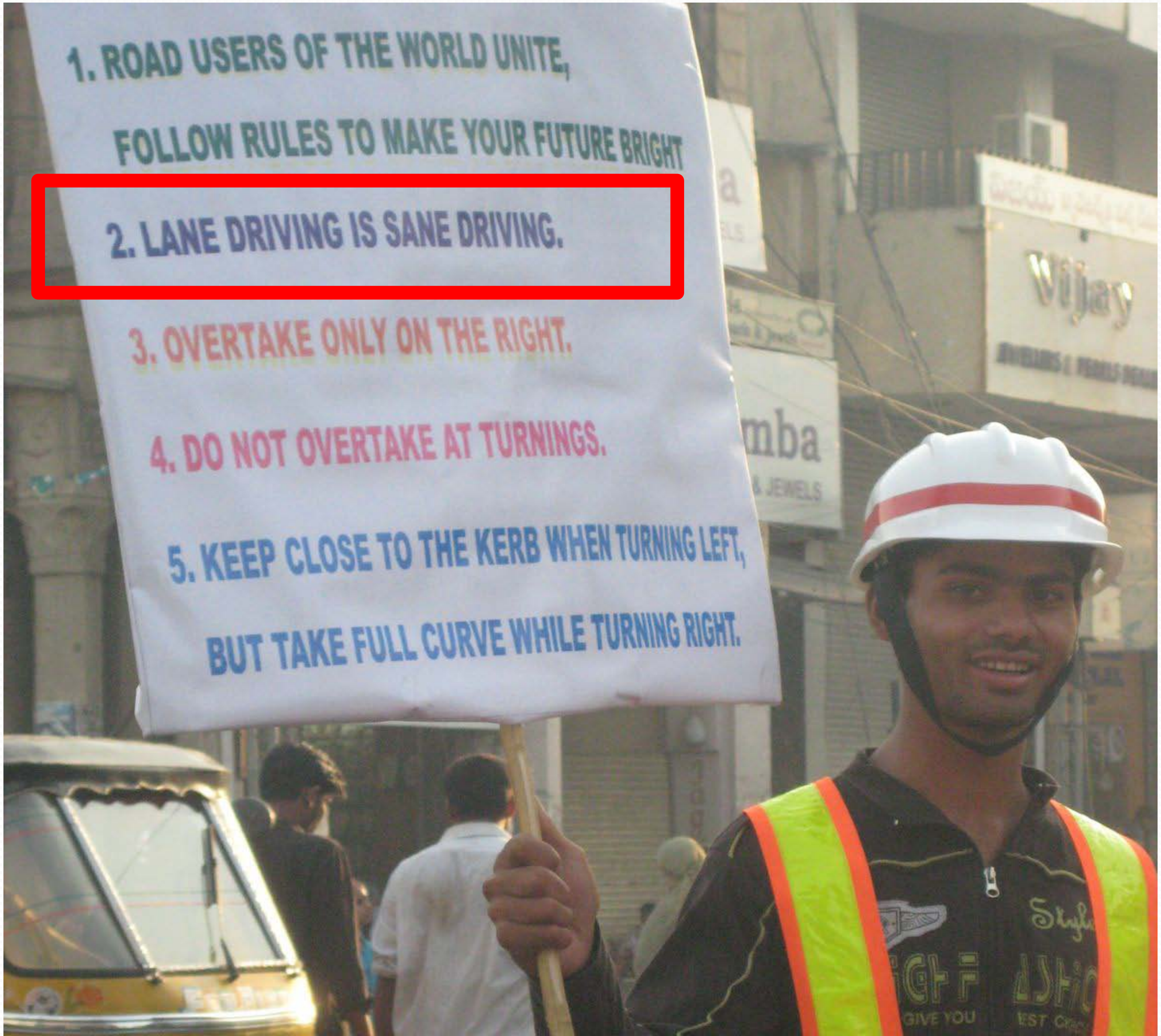
Drive on the right

Coordination problem
2 natural solutions:
Drive on Right and Drive on Left
Fairly self–enforcing and self-stabilizing

Requires collusion to switch
eg. Sweden, September 3, 1967 at 4:50 AM



Driving in India


1. ROAD USERS OF THE WORLD UNITE,
FOLLOW RULES TO MAKE YOUR FUTURE BRIGHT
2. LANE DRIVING IS SANE DRIVING.
3. OVERTAKE ONLY ON THE RIGHT.
4. DO NOT OVERTAKE AT TURNINGS.
5. KEEP CLOSE TO THE KERB WHEN TURNING LEFT,
BUT TAKE FULL CURVE WHILE TURNING RIGHT.
Vijay
mba

Social Contract Technology

- Mathematical proof
- Formal contracts and laws
- Provably least restrictive constraints
- Given desired properties generate constraints
- Stability properties
- Revealable source code and utility functions
- Provably limited systems
- Provably limited escrow agents
- Formal Provenance

Must Choose the Rights We Want

The Bill of Rights

First Ten Amendments to the Constitution of the United States

Article I. Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.

Article II. A well-regulated militia, being necessary to the security of a free state, the right of the people to keep and bear arms, shall not be infringed.

Article III. No soldier shall, in time of peace, be quartered in any house, without the consent of the owner, nor in time of war, but in a manner to be prescribed by law.

Roadmap from the Present

- We'll need AIs to design these systems
- But we must trust the design AIs!
- Computational hardware provably isolated from its software
- Provably limited manufacturing hardware
- Provably limited software
- Social trust networks
- Incentive design
- Safety monitoring networks

Self-Aware Systems

Semantic Computing Initiative

Cooperative Technology Initiative

www.selfawaresystems.com

Create a Cooperative Future