COMPLEXITY IN BIOLOGY EVOLUTION, PARADOXES & INFORMATION

© David Krakauer, SFI.

Complexity @ SFI

- Operational: How should be study/analyze complex systems?
- Computational: How can we measure complexity?
- Evolutionary: Where does the complexity come from and how is it changing in time?
- Pragmatic: Why does it matter?

Complexity in Biology:

The Culturally Evolving Zeitgeist
for Evolved Forms of Complex Life


ASTOUNDING SUPER-SCIENCE FANTASIES!
STRANGE WORLDS
10c
No.7
space-gods of planetoid 50
sabotage on space station 1
the man who fought the world


THE ANDROMEDA STRAIN
It's a bad day to be human.
TWO-NIGHT TELEVISION EVENT
A&E
PREMIERES MEMORIAL DAY 9/8C

Universal Evolution circa 1950

Evolutionary Ockham's Razor

The Gould Contingency Doctrine

"Humans arrived just yesterday on the stage not noticeably prepared for our advent. **Would we appear at all if we could rewind the tape to an appropriate beginning** (say, the origin of modern phyla in the Cambrian explosion more than 500 million years ago) and let it run again?"

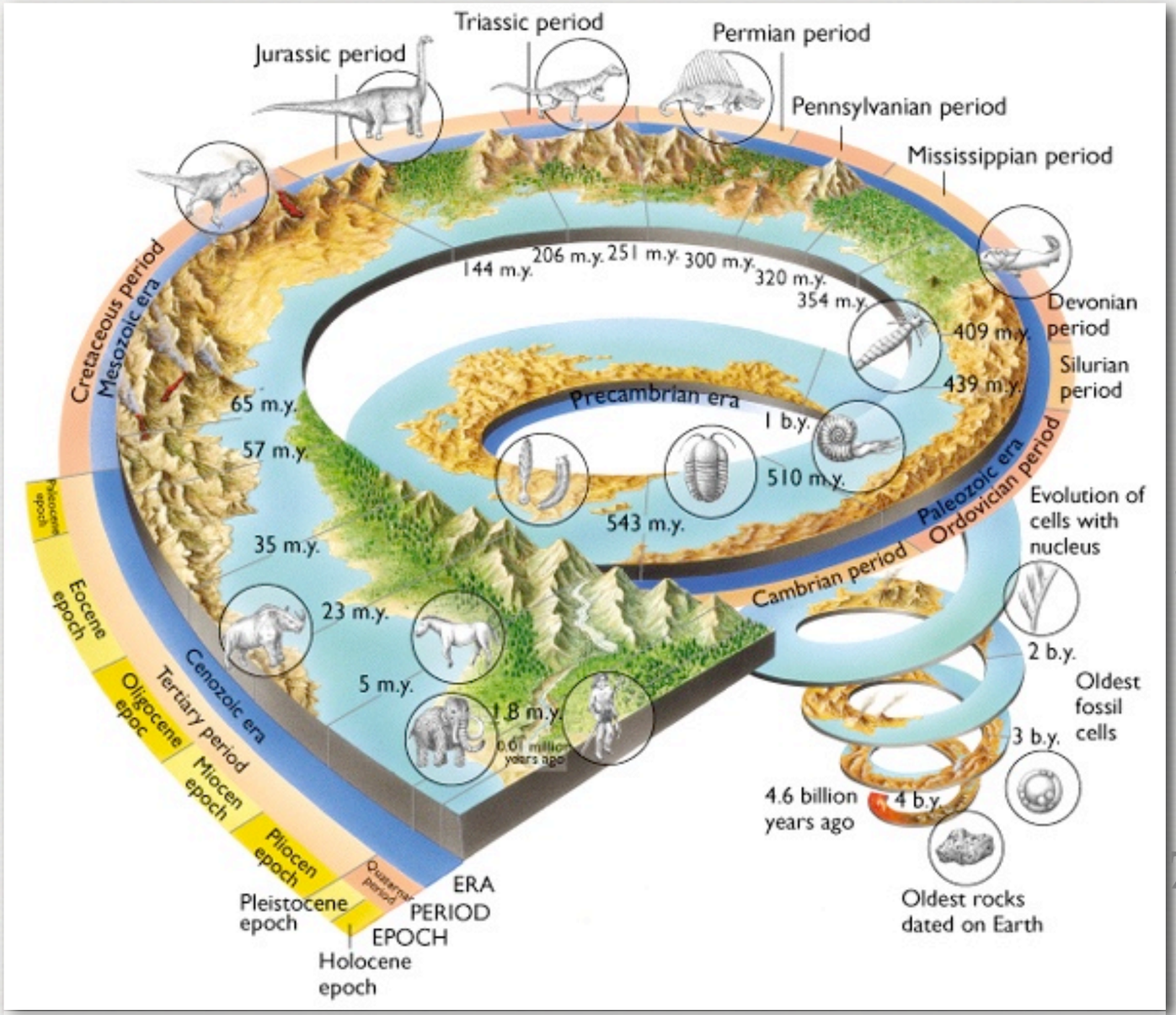
"Suppose that the rerun varies greatly from the original, but that the new version makes perfectly good sense too. Perhaps our world is not random. The race does go to the swift, but **each time you play the tape, different lineages win the relay for wondrously diverse reasons**. We cannot know who will be swift and why, but winners prevail for good cause"

"But the slightest early nudge contacts a different groove, and history veers into another plausible channel, diverging continually from its original pathway. **The end results are so different, the initial perturbation so apparently trivial. If little penis worms ruled the sea, I have no confidence that Australopithecus would ever have walked erect on the savannas of Africa.**"

S.J. Gould. Play it Again Life, Natural History. 1986

The Evidence from Earth History

Unwinding the Terrestrial Tape



4.6b Earth

800m

3.8b Prok cells

1.8b

2.0b Euk cells

1.0b

1.0b multi-cells

400m

600m animals

600m

7m hominids

Complexity:
Description
Length

Cell components

Cell components & compartments

Cell components, compartments, & tissues

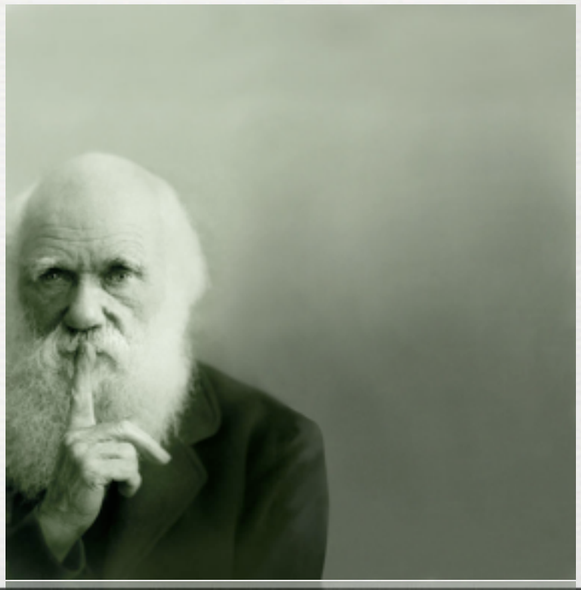
Cell components, compartment, & tissues & behavior

Cell components & tissues & behavior & culture

800m

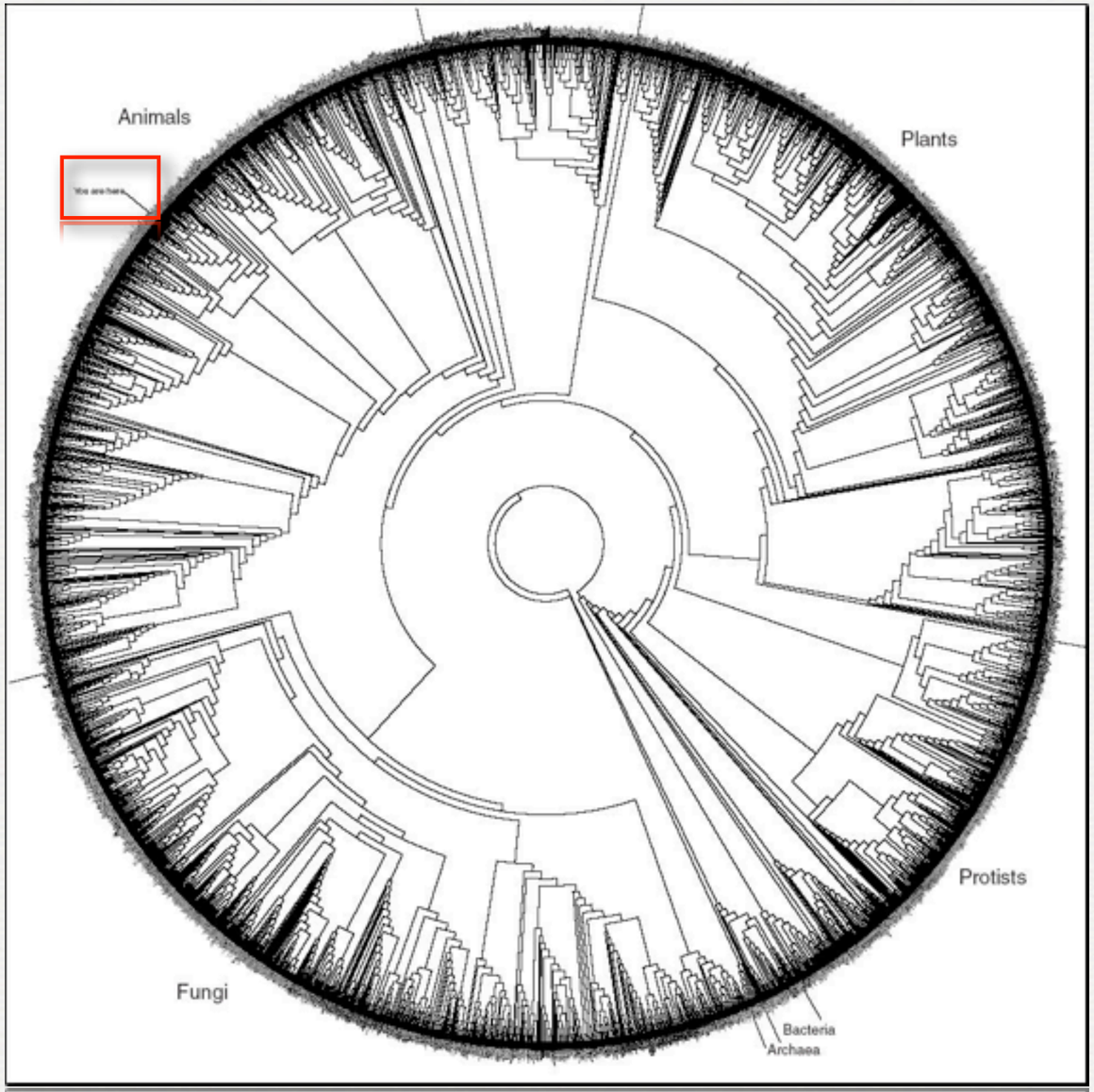
2.8b

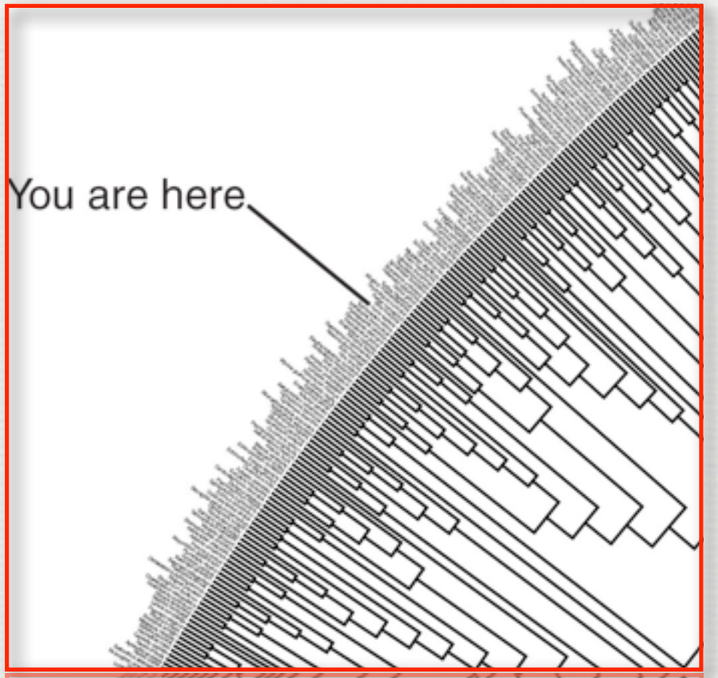
600m



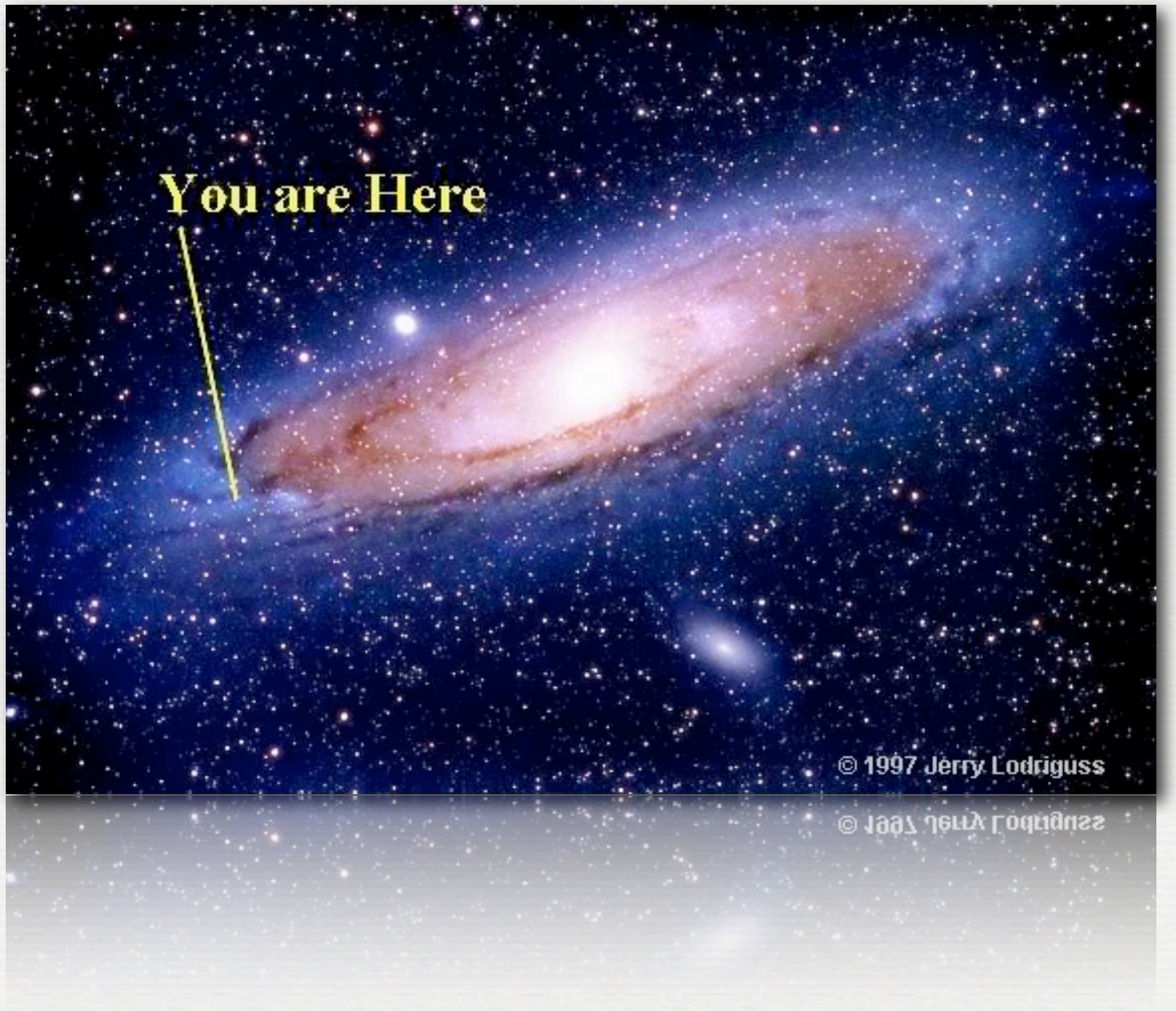
It took around 13 billion years for the universe to be able to contemplate itself

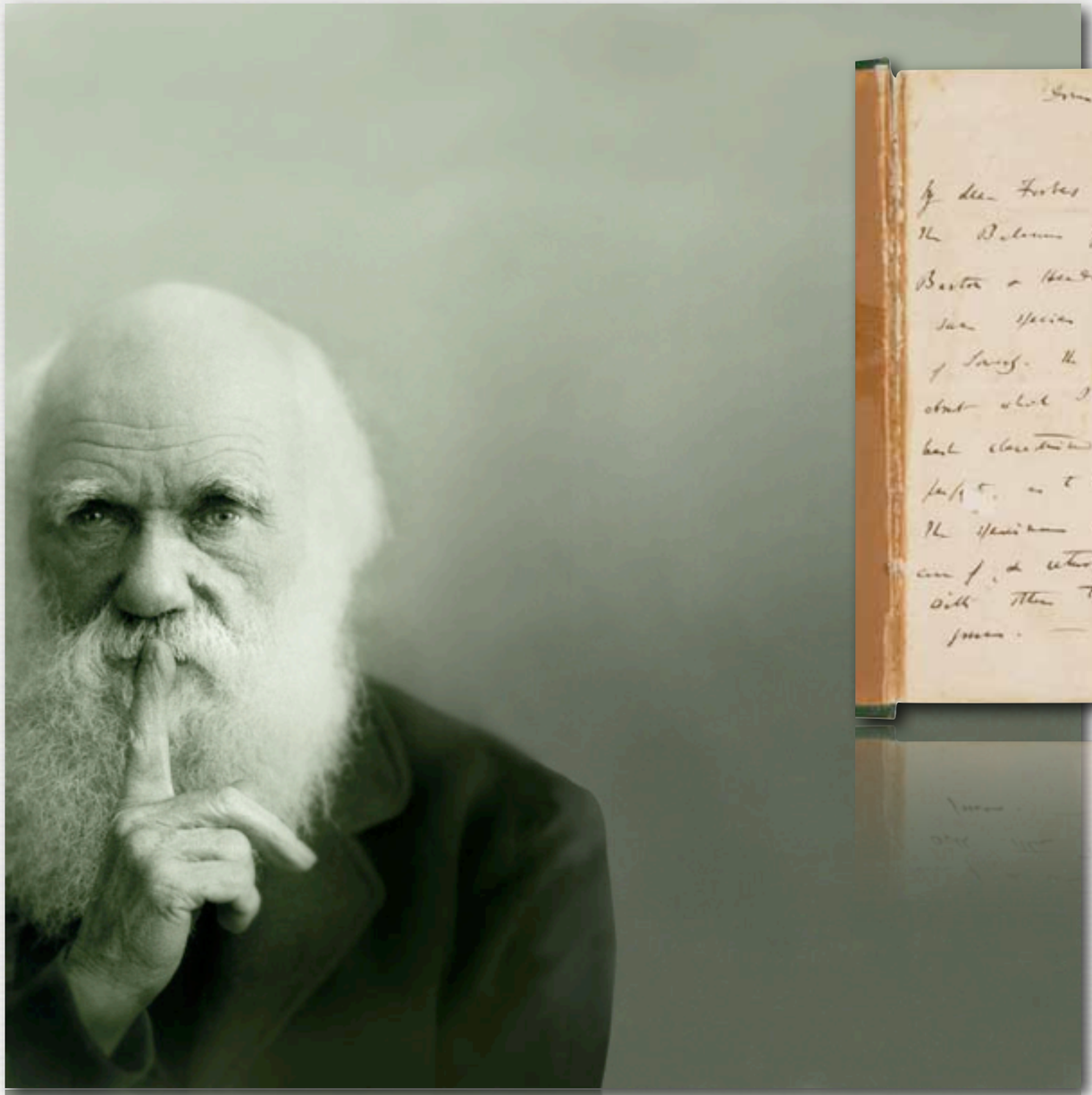
But not all of life on earth (in the universe) is contemplative...

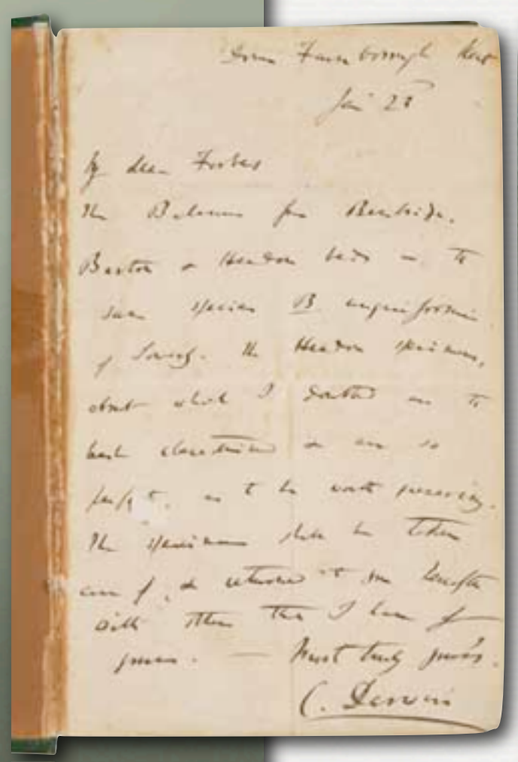

Animals
Plants
Protists
Fungi
Bacteria
Archaea

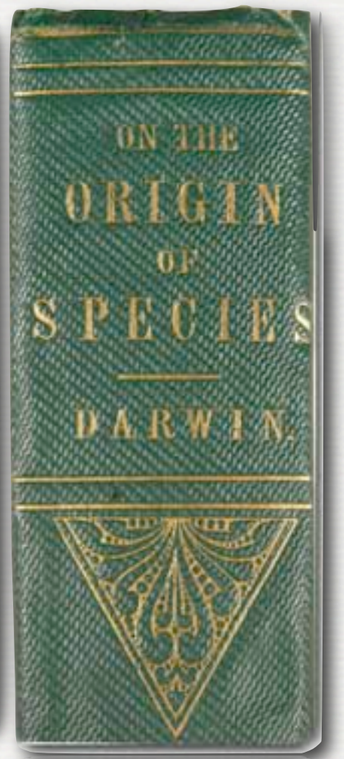

You are here

And it gets worse


You are Here
© 1997 Jerry Lodriguss






ON THE
ORIGIN
OF
SPECIES
DARWIN.

WHAT IS EVOLUTIONARY THEORY?

2000

1990

1930

1859

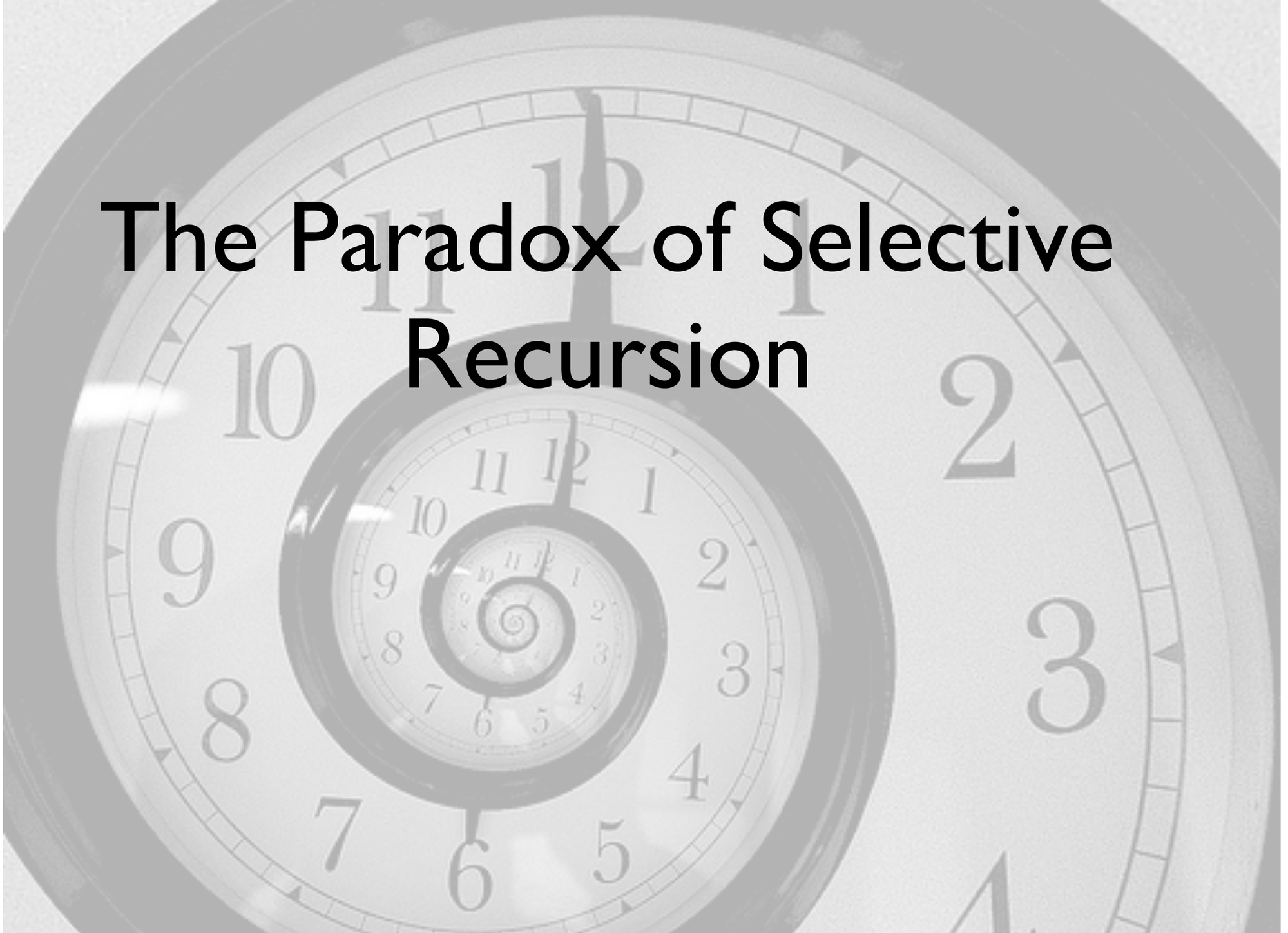
- A Physics-like theory searching based on Laws?
- A Statistical/Inferential Theory like Bayesian learning or approximate dynamic programming?
- An algorithmic/computational optimization theory?
- A dynamical process constrained by inheritance, mutation and selection?
- A narrative, historical description of life on earth structured by a plot called natural selection.

EVOLUTIONARY LAWS OF COMPLEXITY?

- Not a deterministic law like gravity that applies everywhere and all the time
- Not a statistical law like Entropy which applies most of the time
- Is it something like an ideal gas law (PV=nRT) where “empirical” conditions are such that some values are never obtained? That is, not a logical constraint, and when they do, it is a statistical law relating to averages.

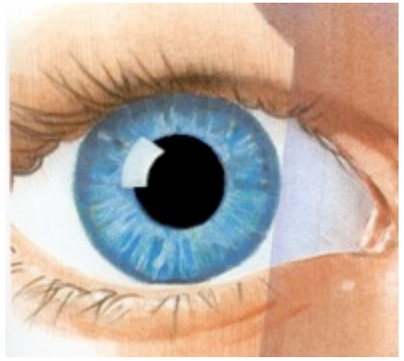
The Paradoxes of Bio-Complexity

- The Paradox of Selective Recursion
- The Sigma to Pi Paradox
- The Zero Omega Paradox
- The Evolutionary light speed constraint
- The Genomic Inferential Limit
- The Paradox of the Bootstrapped Niche (origins of selection)



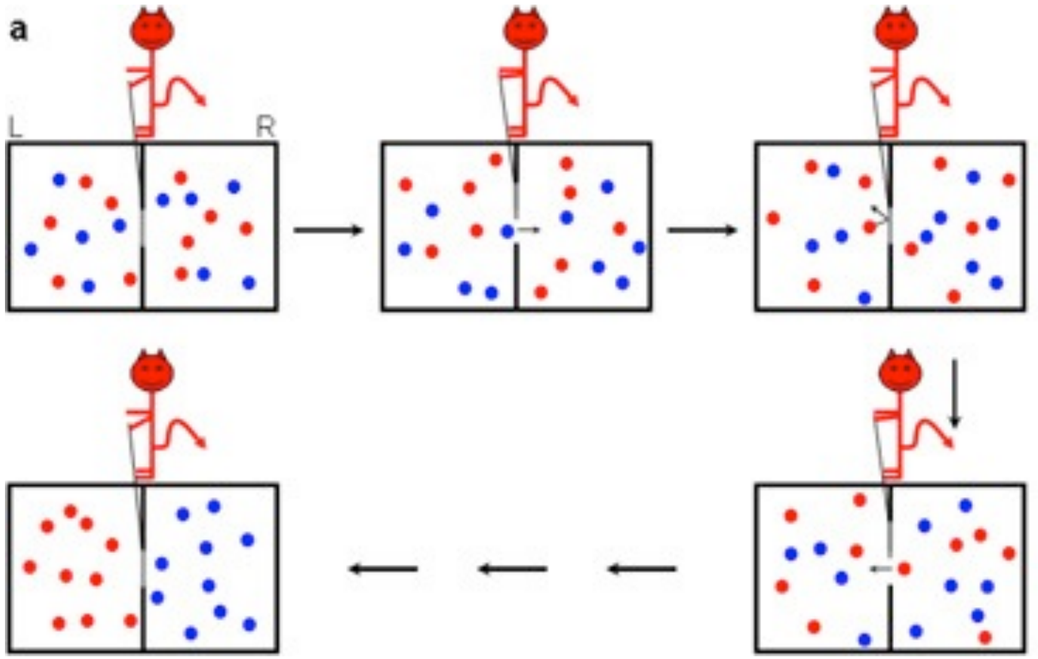
The Paradox of Selective Recursion

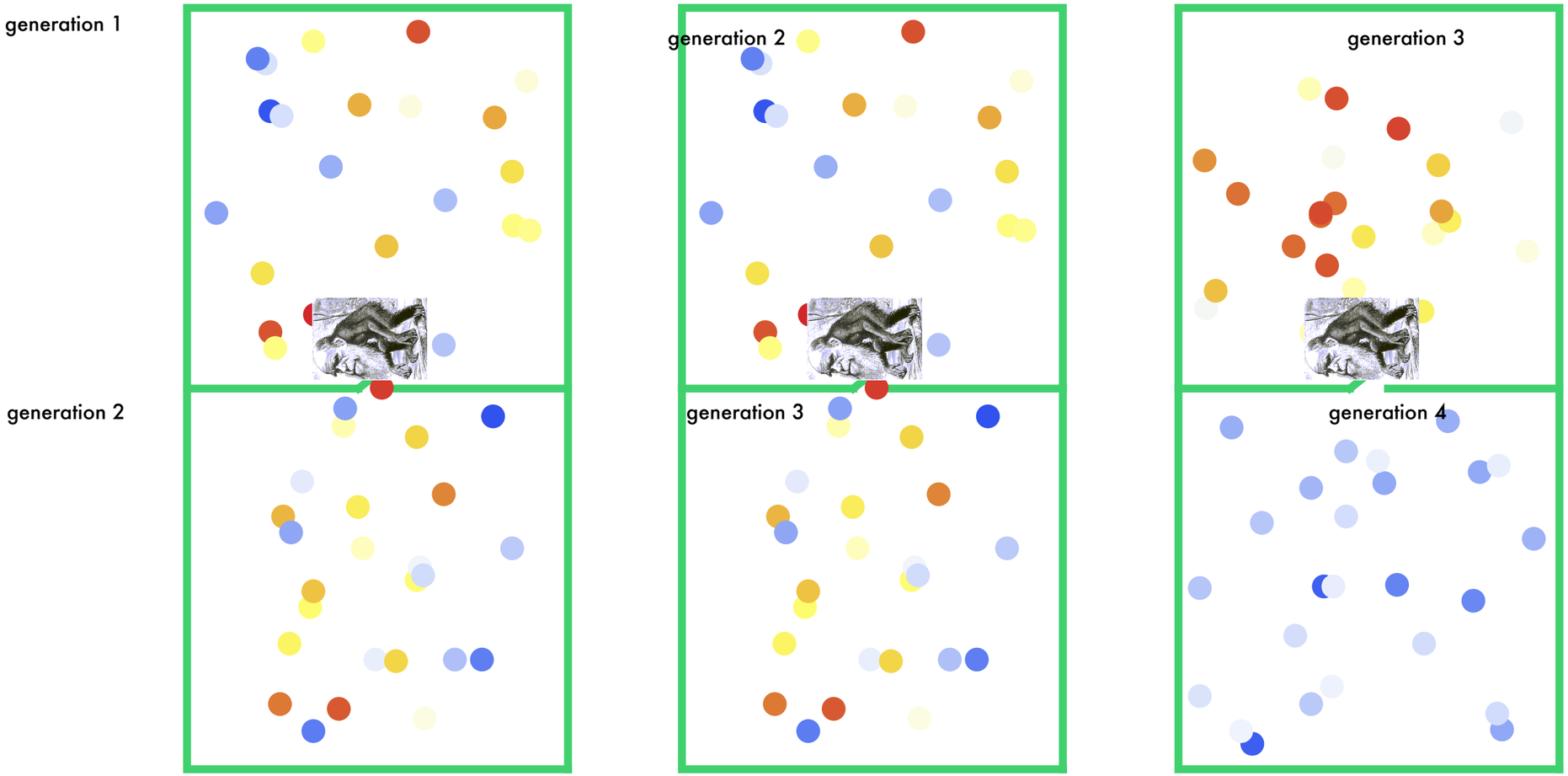
natural
selection



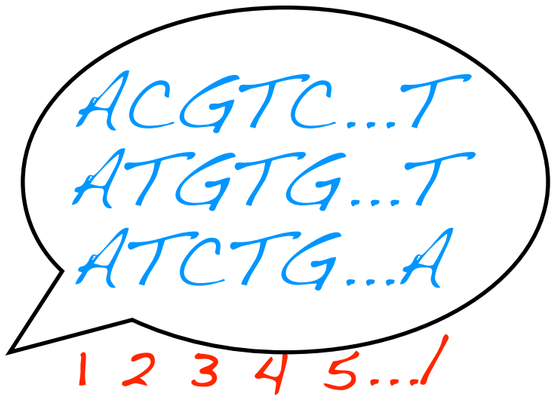
artificial
selection

Maxwell's Demon




generation 1
generation 2
generation 2
generation 3
generation 3
generation 4

Selective Demons Create Information



Demons have site preferences

$$H_i = -\sum_j p_j^{(i)} \log_4 p_j^{(i)}$$

Sites carry information

$$I_i = H_{max} - H_i$$

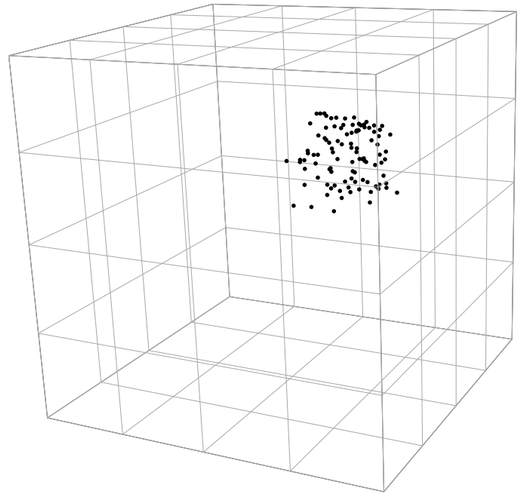
Complexity of Demon

$$C = L - \sum_i H_i$$

Evolutionary Information Storage

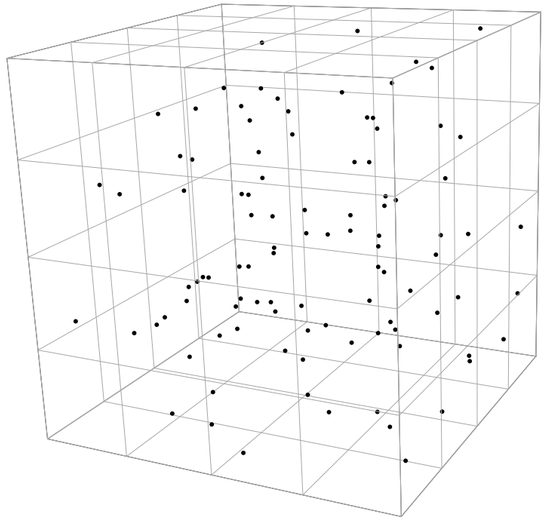
Information Conserved

$$\mu L < 1 < sN$$



Information Lost

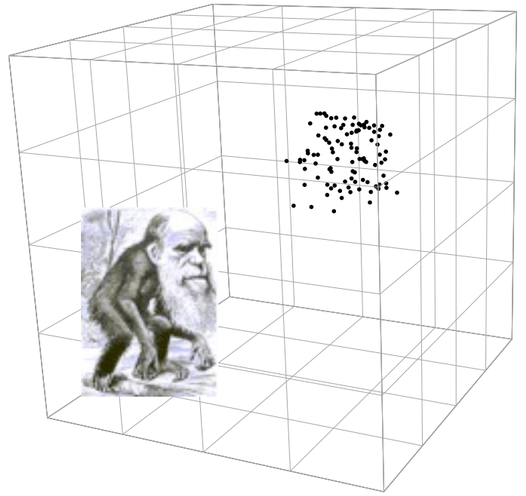
$$\mu L > 1 < sN$$
$$\mu L < 1 > sN$$
$$\mu L > 1 > sN$$



Evolutionary Information Storage

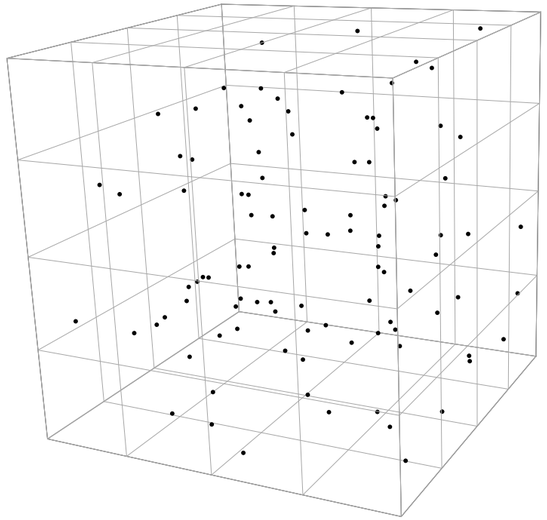
variables localized

$$\mu L < 1 < sN$$



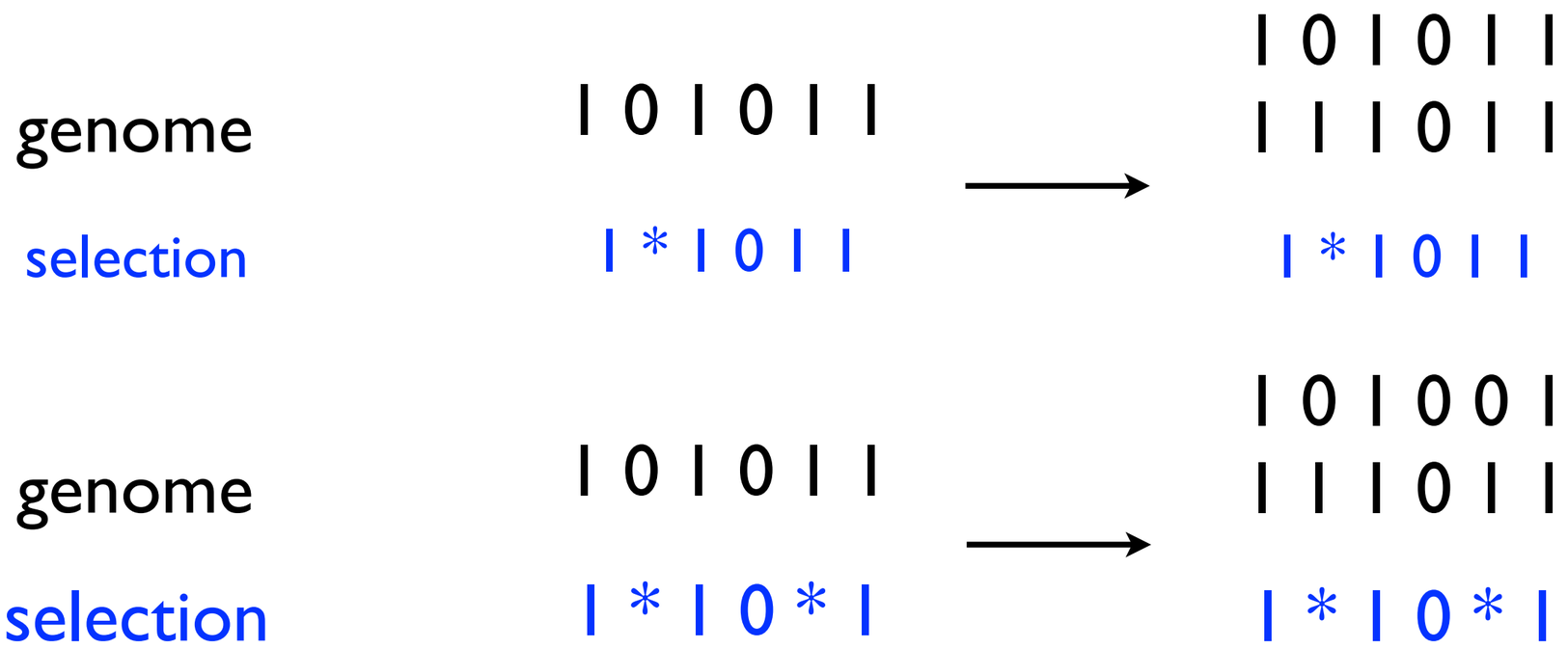
variables Distributed

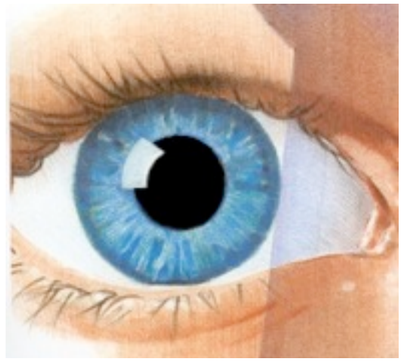
$$\mu L > 1 < sN$$
$$\mu L < 1 > sN$$
$$\mu L > 1 > sN$$

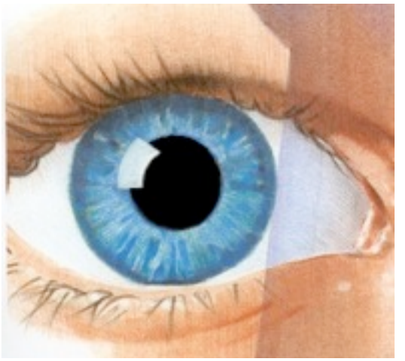


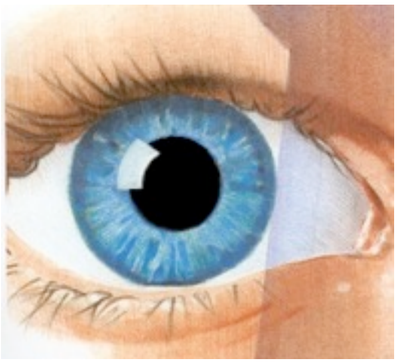
Sequence Selection = Template Complentarity

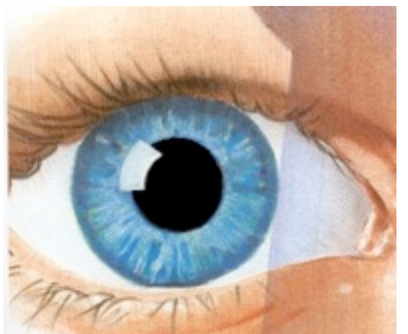
$$C(\mathbf{g}) = kLD(s)$$



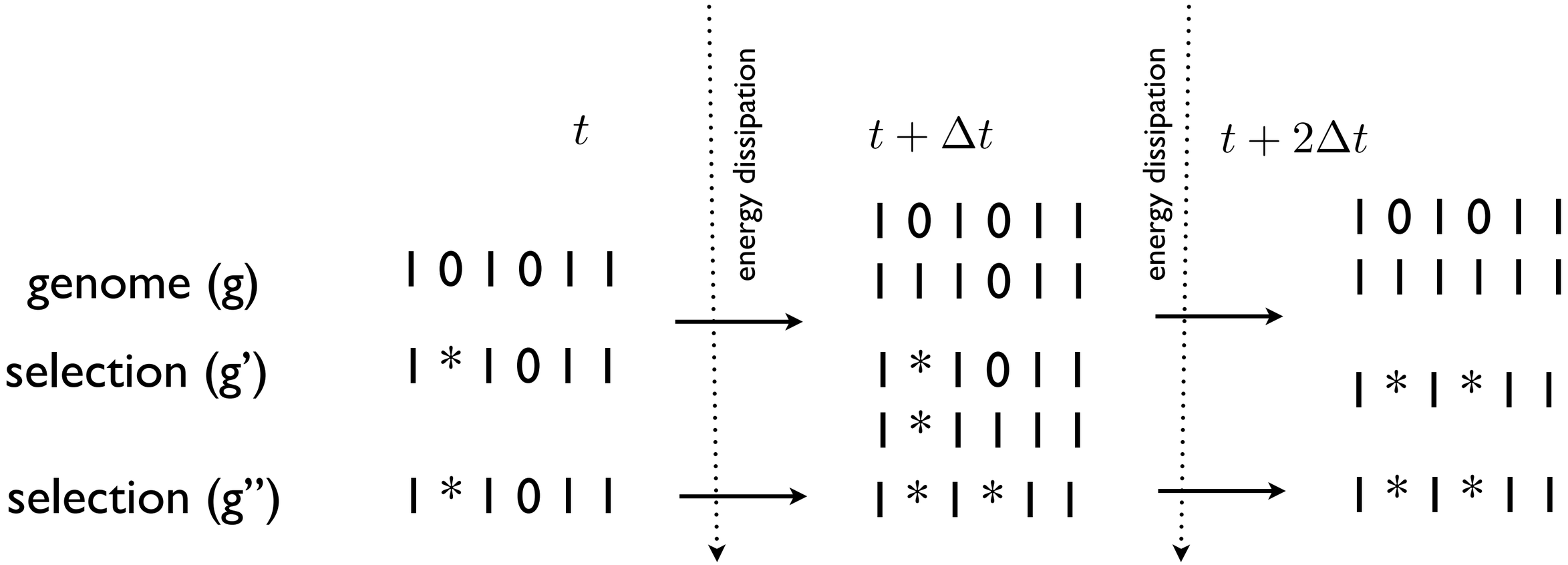








Selective Recursion & Template Dissipation

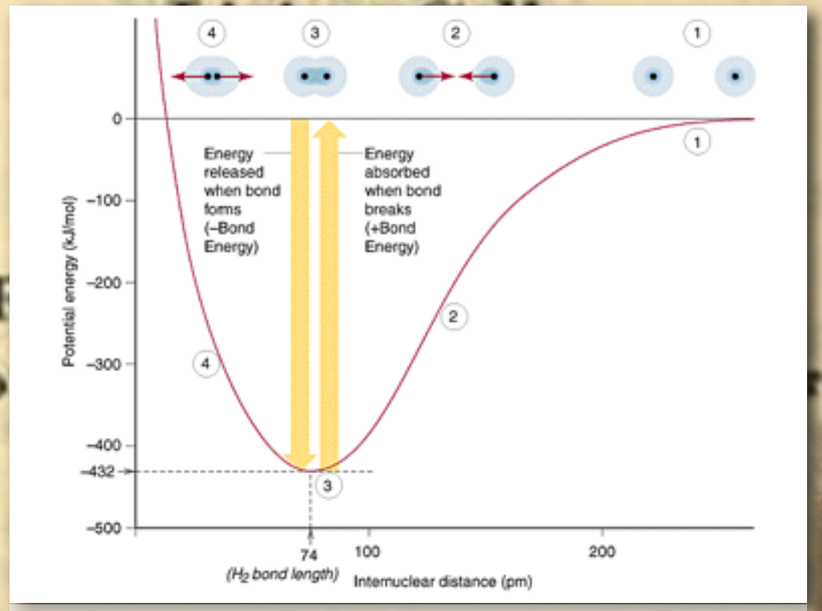


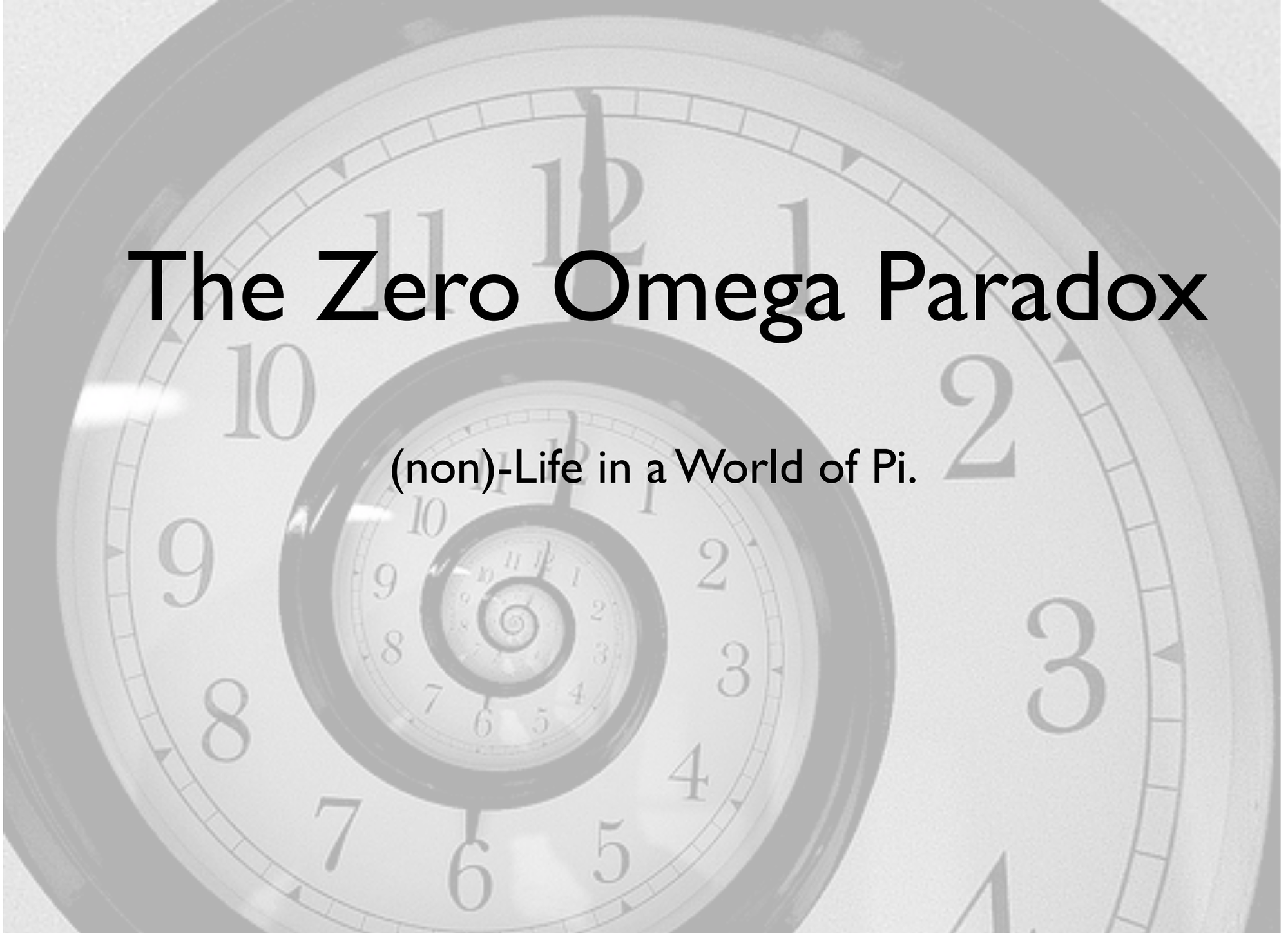
Limiting Distribution: ******

Fund' selection (= Physics?)

WHAT IS LIFE?

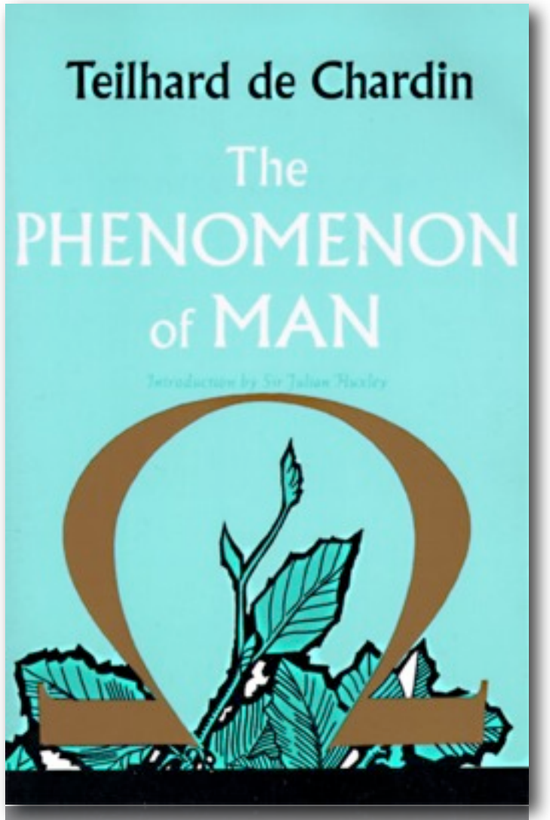
The Physical Aspect of the



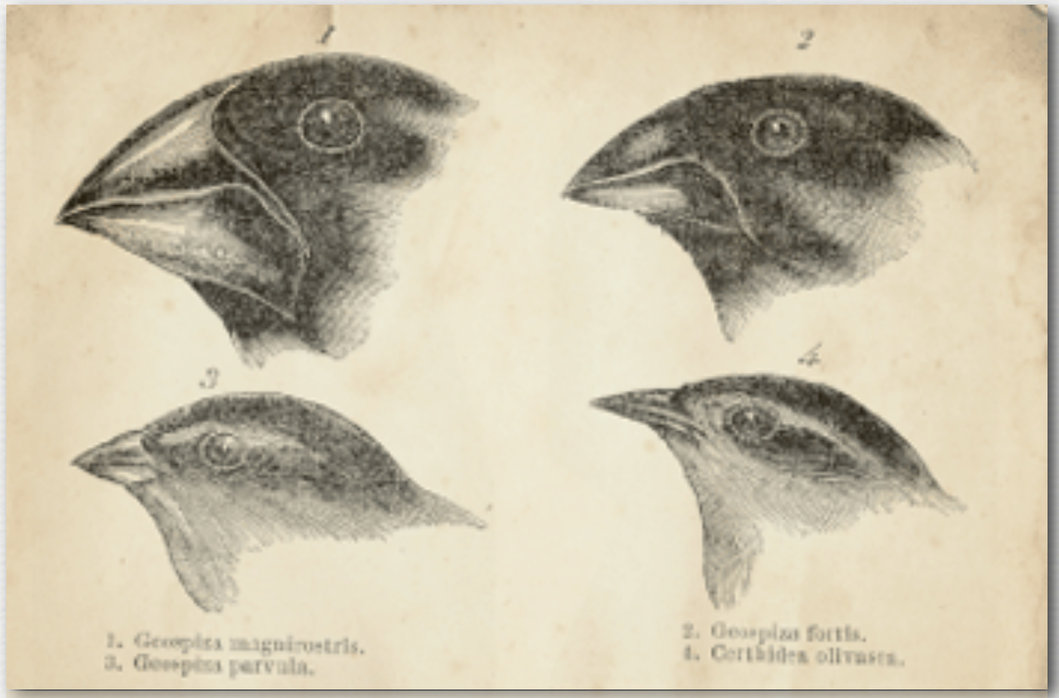


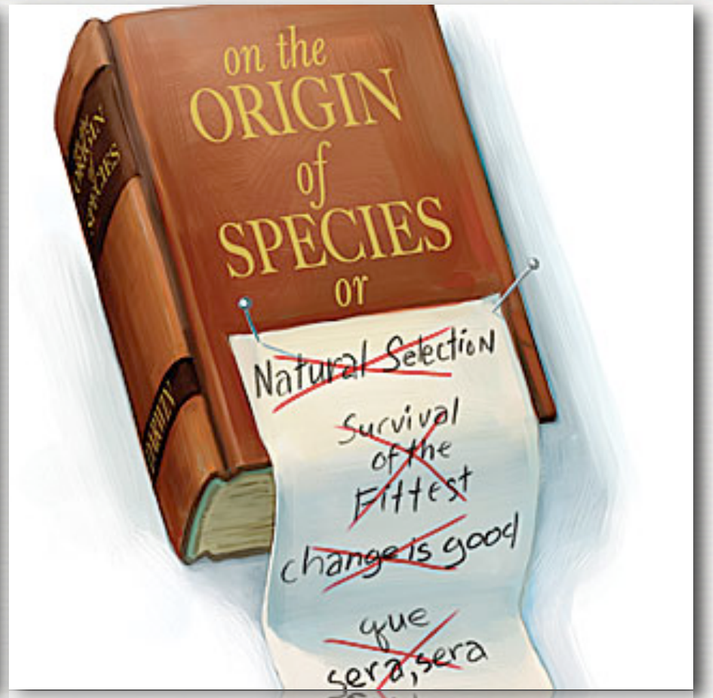
The Zero Omega Paradox

(non)-Life in a World of Pi.

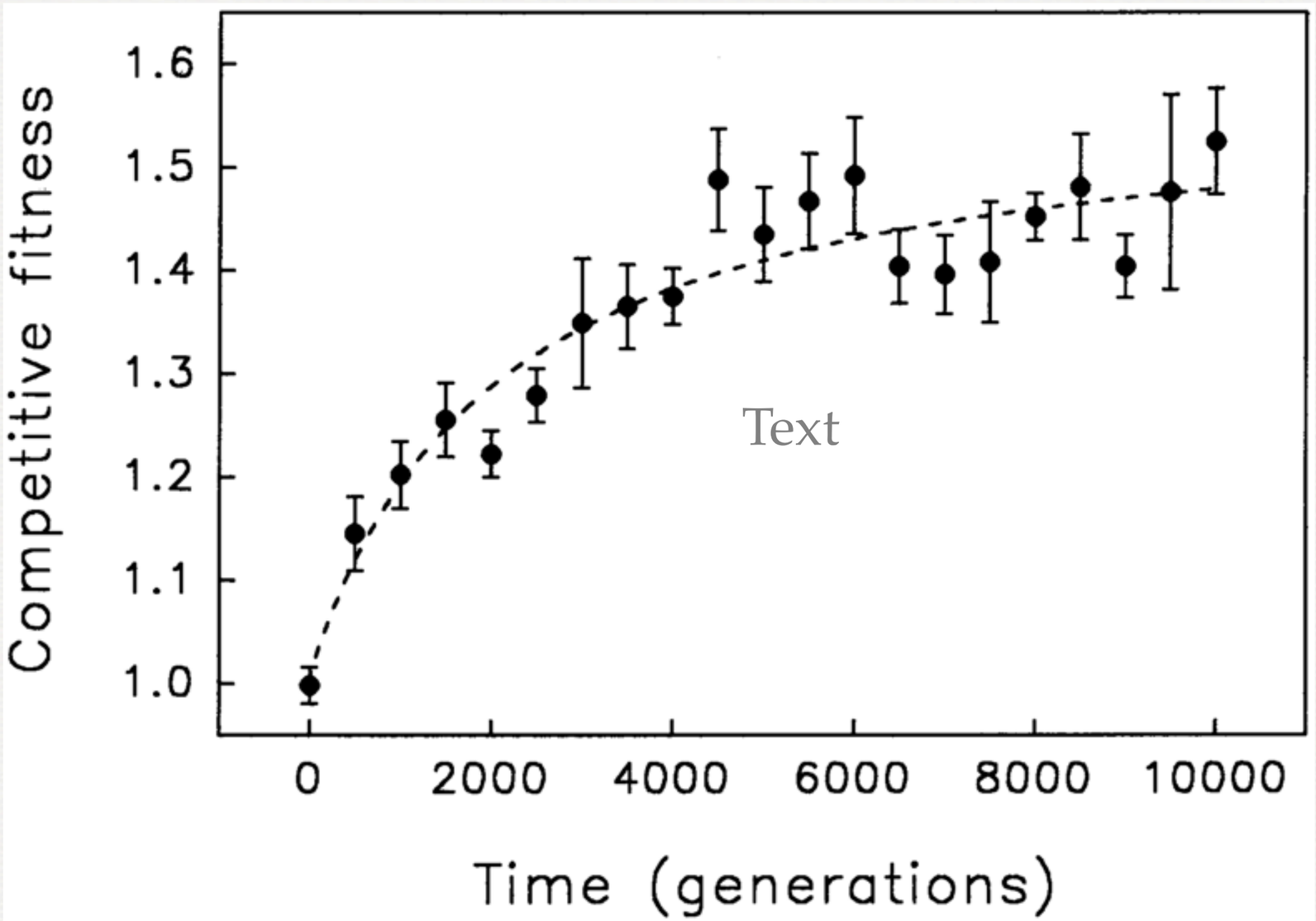

Teilhard de Chardin
The
PHENOMENON
of MAN

Experimental determination of selective regularties



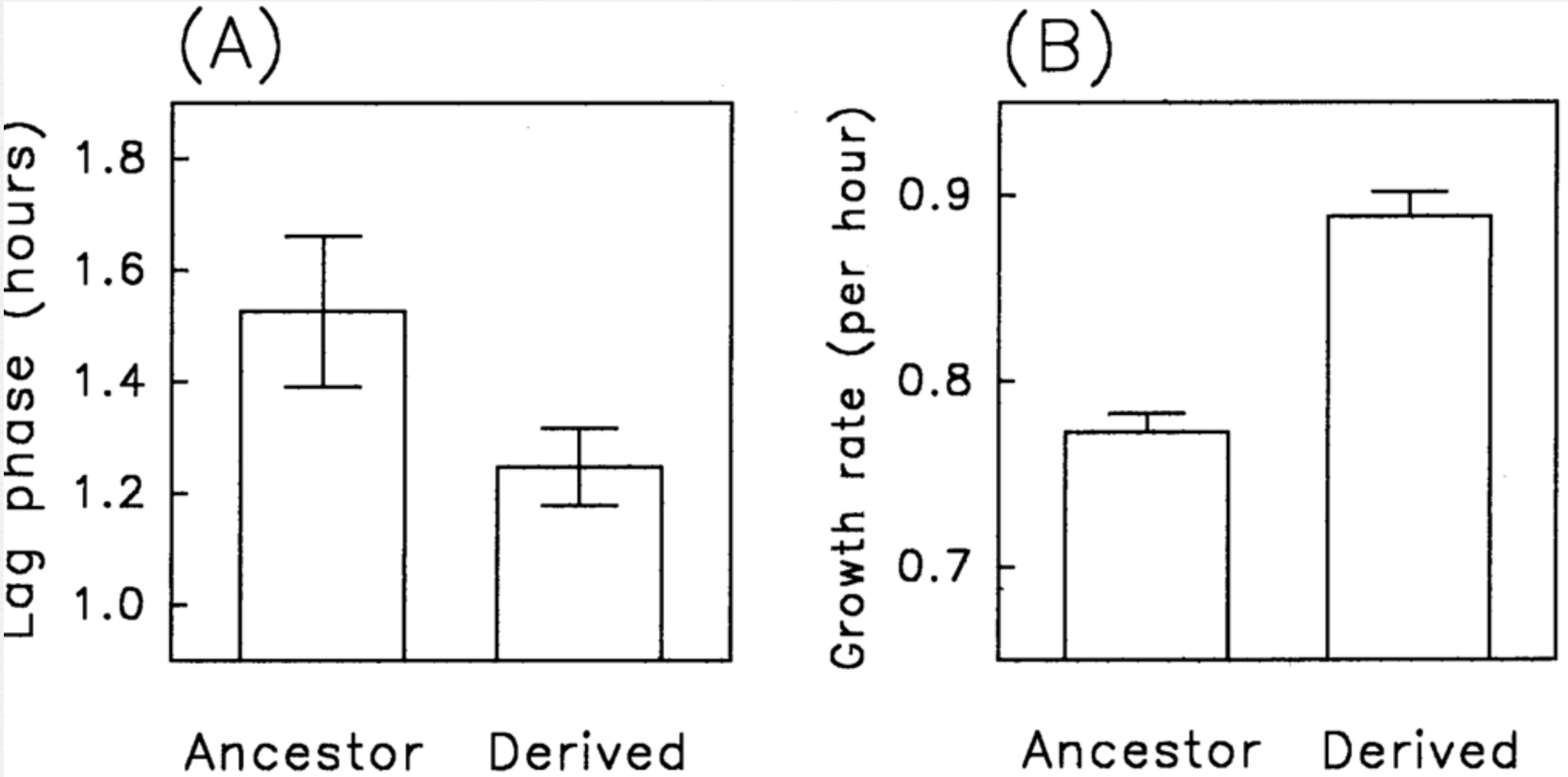


Increasingly selected E. coli in competition against its ancestor



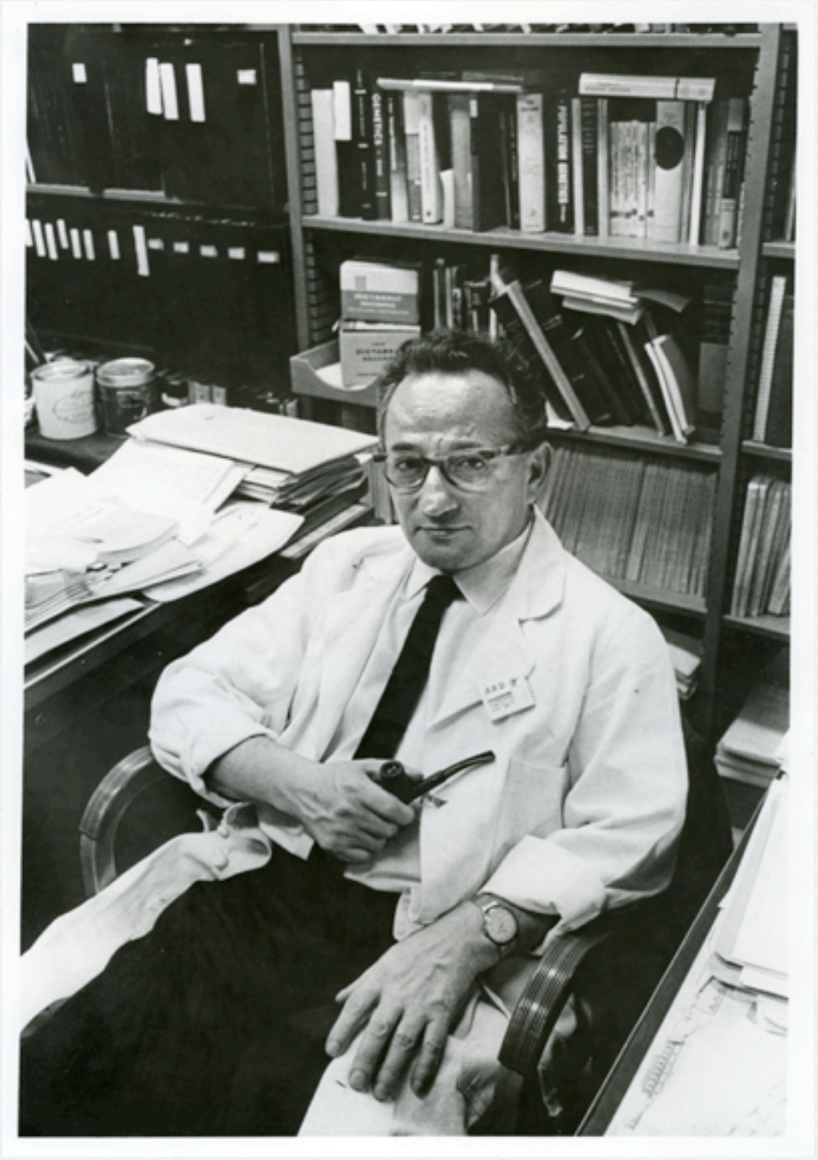
Lenski et al 1998

Absolute increases in performance

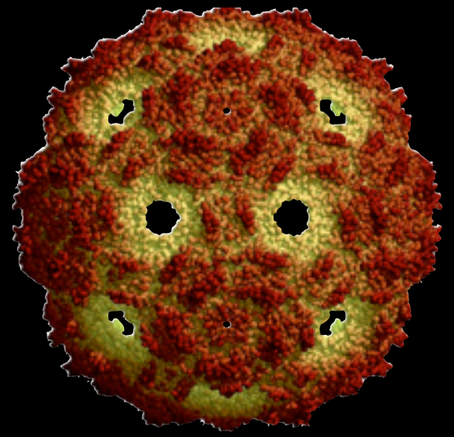


Lenski et al 1998

Sol Speigelman & his Monster



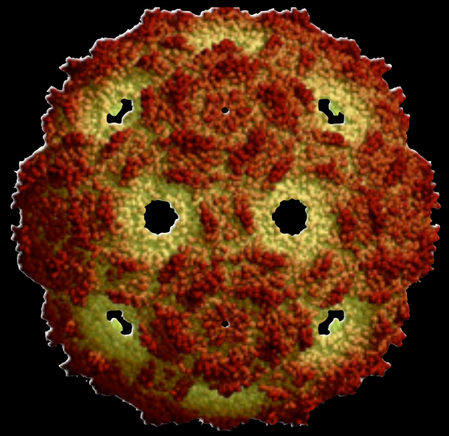
Q Beta Phage



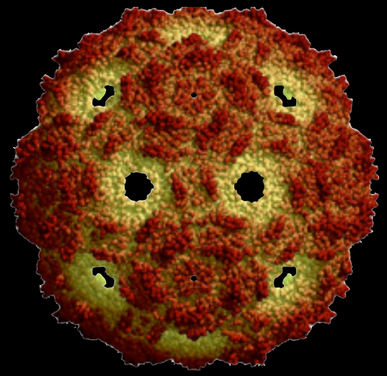
QBeta RNA replicase

1st gen

4.5kbp RNA



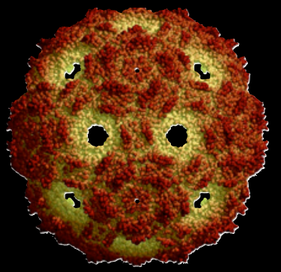
Q Beta Phage



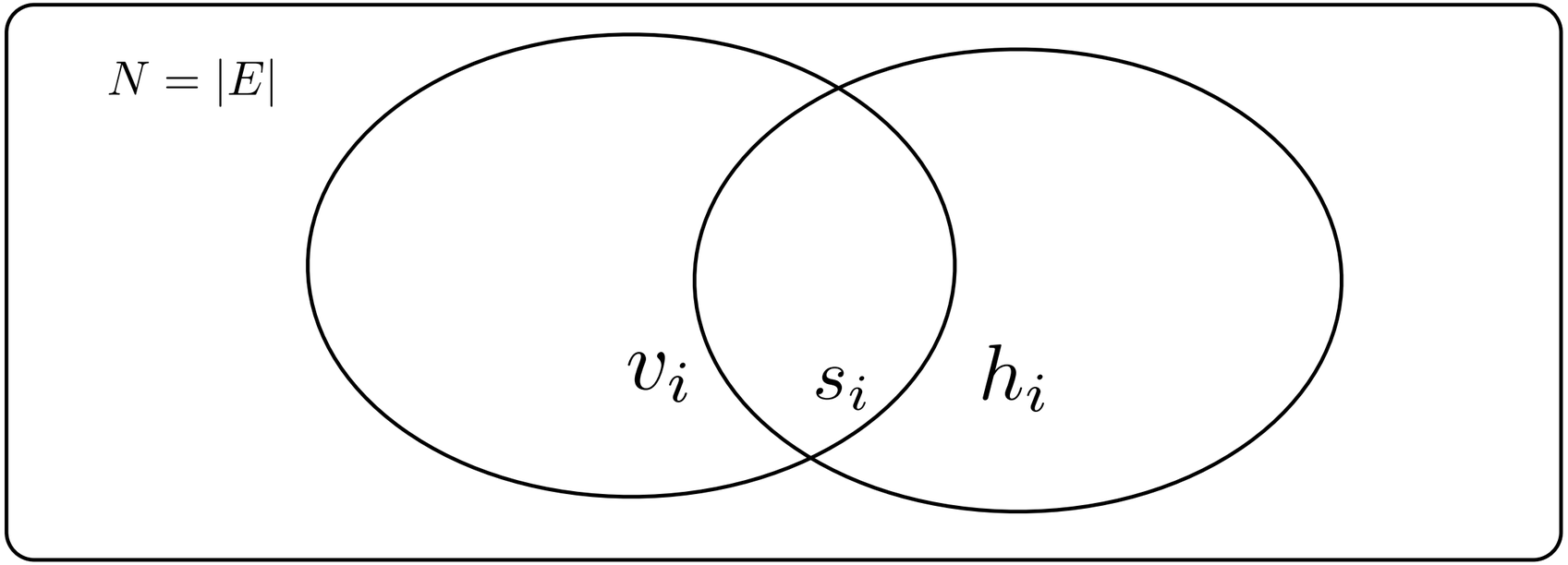
QBeta RNA replicase

74th gen

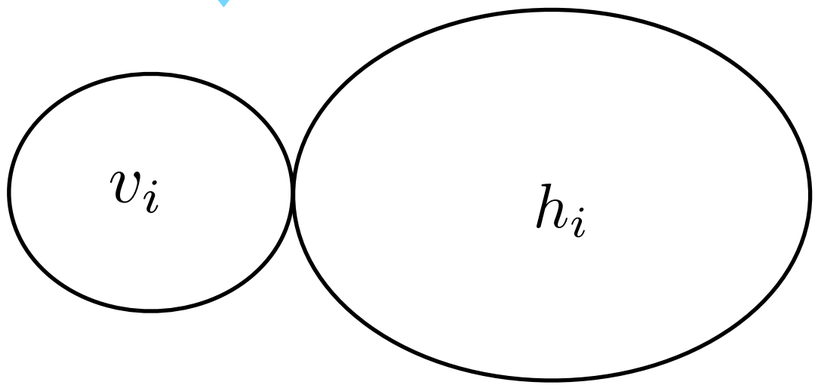
218bp RNA



The Speigelman Monster

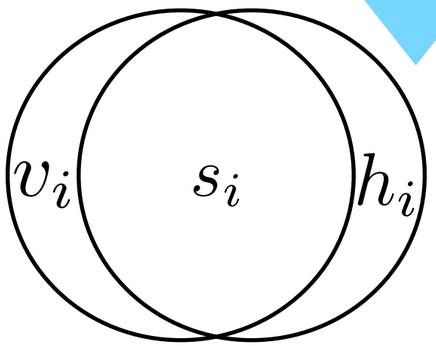


Minimality

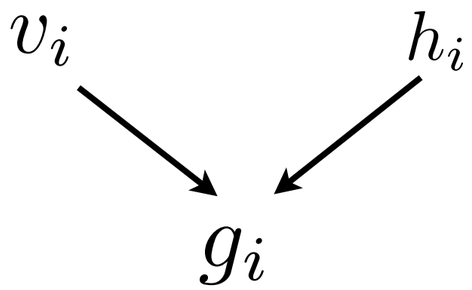


$$|v \cup h| = N$$
$$|v \cap h| = 0$$

Autonomy



$$|v \cup h| = N$$
$$|v \cap h| \approx N$$



$$g_i = h_i \vee v_i$$

$$Prob(h_i = 1) = q \quad Prob(v_i = 1) = p$$

$$L_v = \sum_i v_i, \qquad < L_v >= pN$$

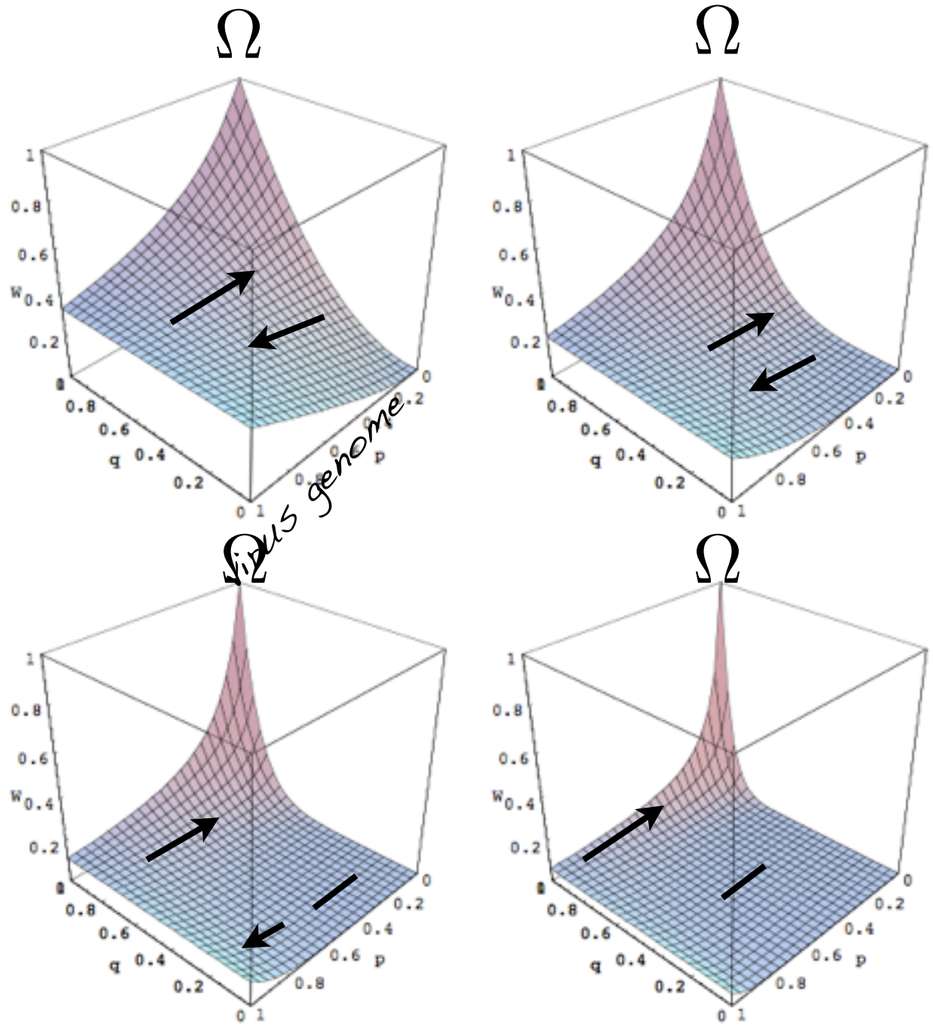
$$w = \frac{\prod_{i \in N} g_i}{(1 + L_v)} \qquad < w >= \frac{(q + p - qp)^N}{1 + pN}$$

$N=4$

$N=8$

host genome

virus genome



$N=16$

$N=32$

The Story Thus Far

- Selective demons need to be as complex as the organism
- Demons also dissipate, and so need their own demons
- Increased predictability reduces the need to encode adaptive information
- Increased fitness does not imply increased coding capacity (regulatory complexity)

Information Maximization

Maximizing Information (s.t costs)

$$H(G)$$

Minimizing Equivocation (s.t costs)

$$H(G|E)$$

$$\max(H(G) - H(G|E))$$

$$= \max(H(G) + H(E) - H(G, E)) = \max(I(G; E))$$

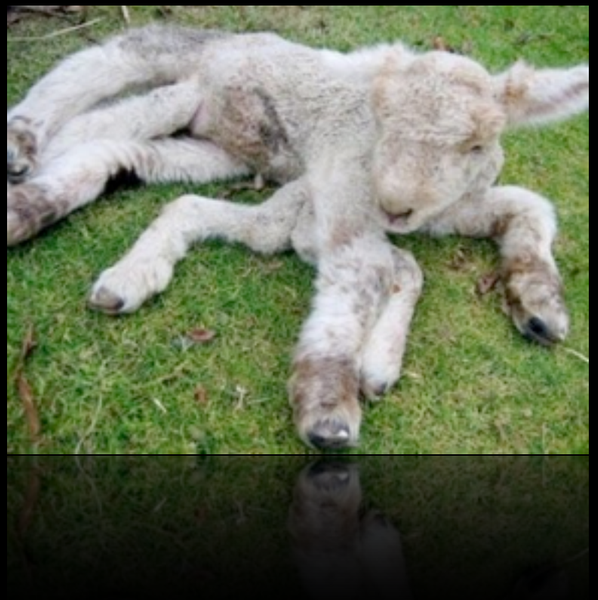
- Selection tends to maximizes I(G,E) at the level of populations using genomic information
- The constraints of evolutionary dynamics select for “epigenetic/ cognitive” mechanisms (complement the selective regime) in the face of increasing H(E), where (|E| > |G|)
- Cognitive mechanisms implement computations that significantly expand the channel capacity of adaptation: I(G,E) -> I(G,E) + I(C,E)
- Multicellular cognition is evidence of sustained |E| > |G| and hence increasing complexity in those lineages capable of detecting H(E)

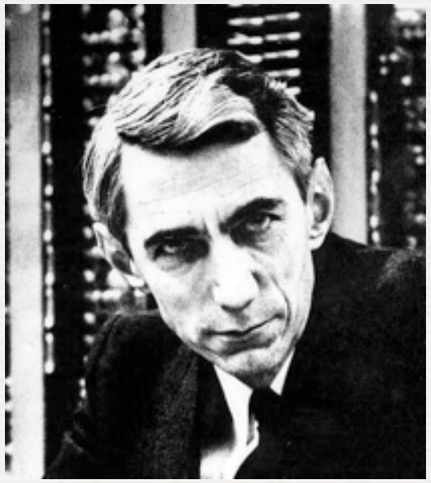
- There are identifiable conditions when this quantity becomes positively correlated with informal definitions of complexity
  - *when H(E) increases.*
- There are identifiable conditions when this quantity becomes negatively correlated with informal definitions of complexity
  - *when H(E) decreases or when H(E,G) increases.*

Both of these cases are adaptive

Coral reef diversity

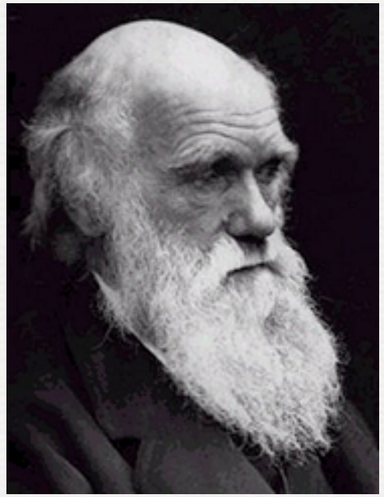
Amblyopsis spelaeus





Transmitter → X → Channel → Y → Receiver

$$I(X;Y)$$

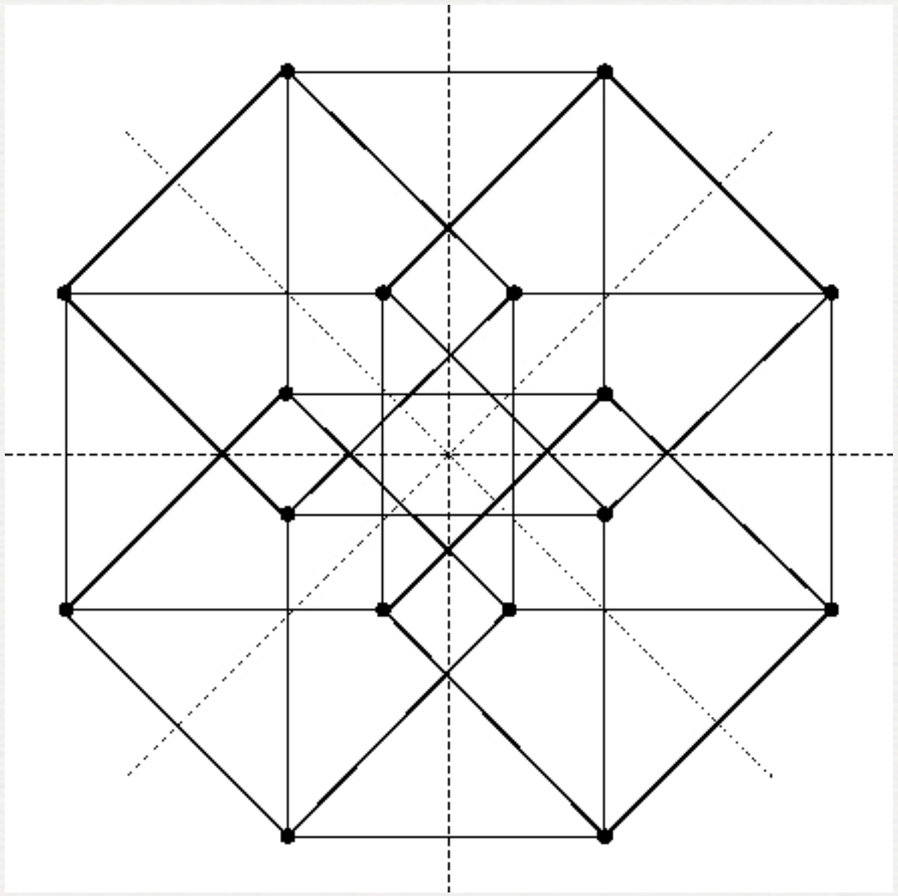


Environment & selection → variation & mutation → Genome

$$I(G;E)$$

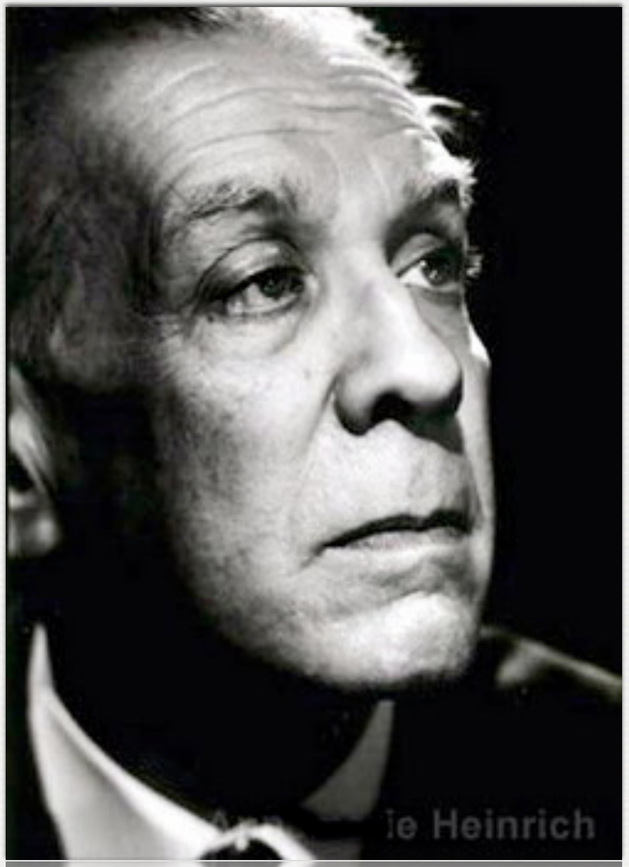
An organism is a message that the environment sends to its future self.

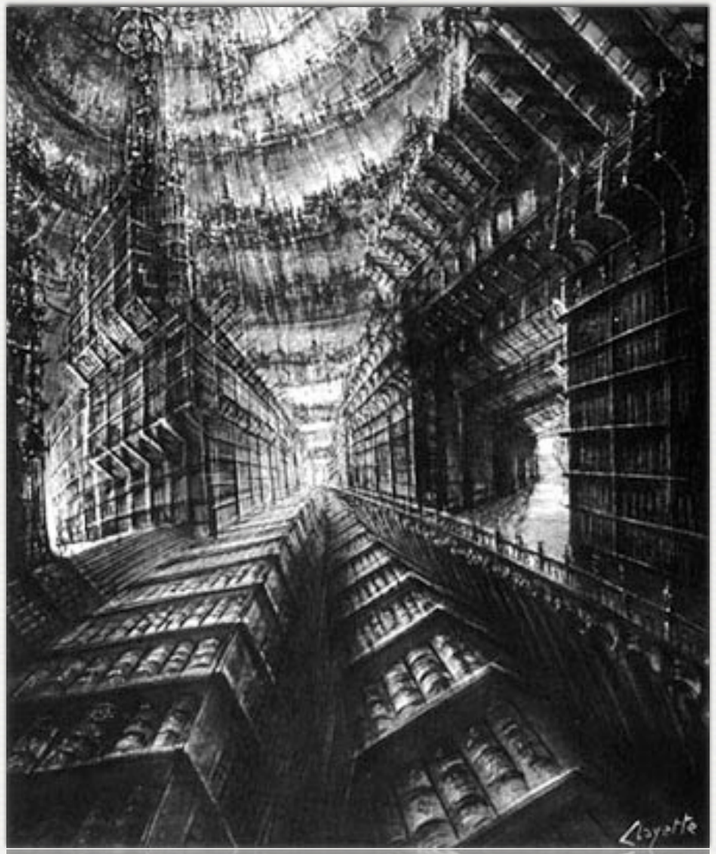
COMPLEXITY: INVENTED OR DISCOVERED?



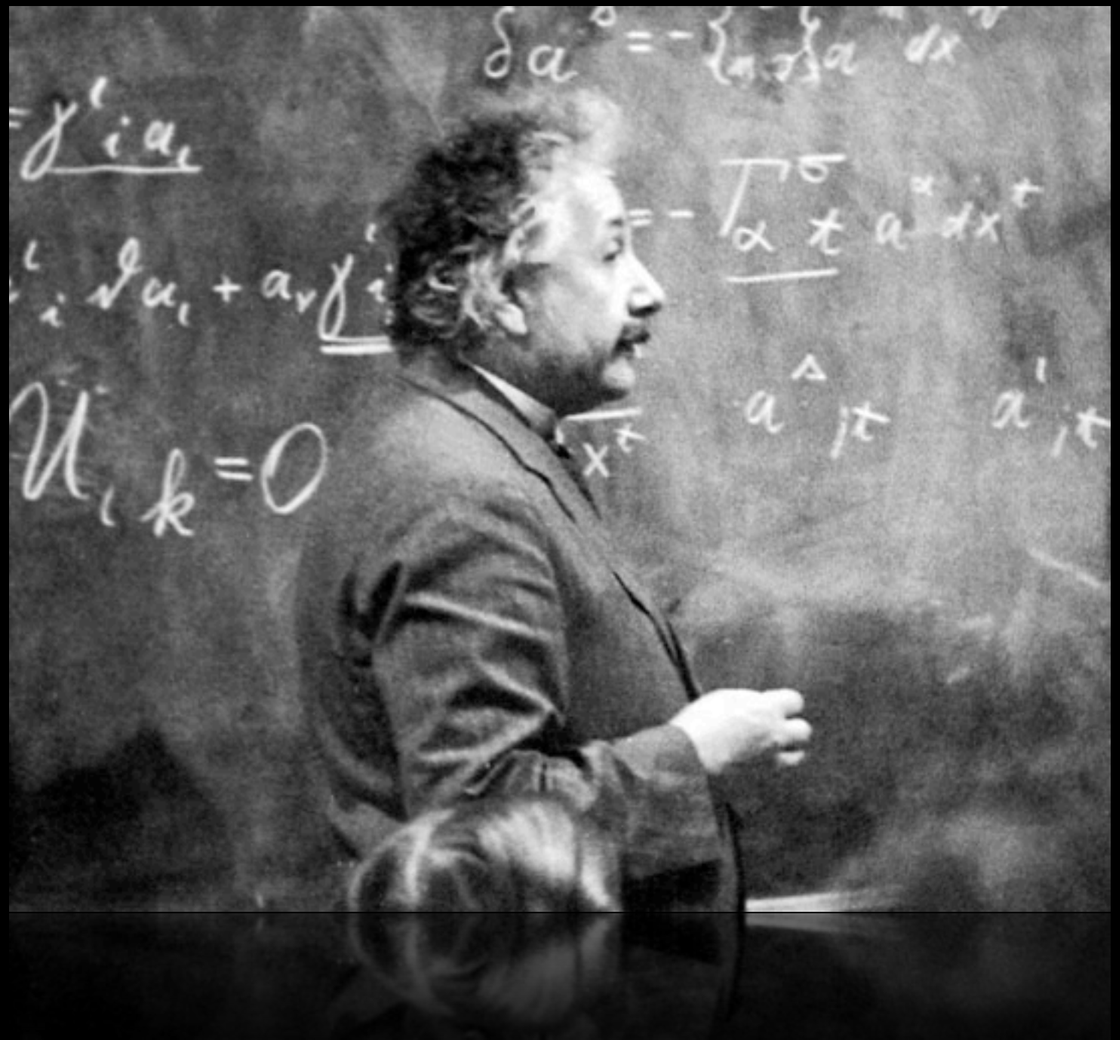
4^L dimensional hypercube

THE INFINITE LIBRARY





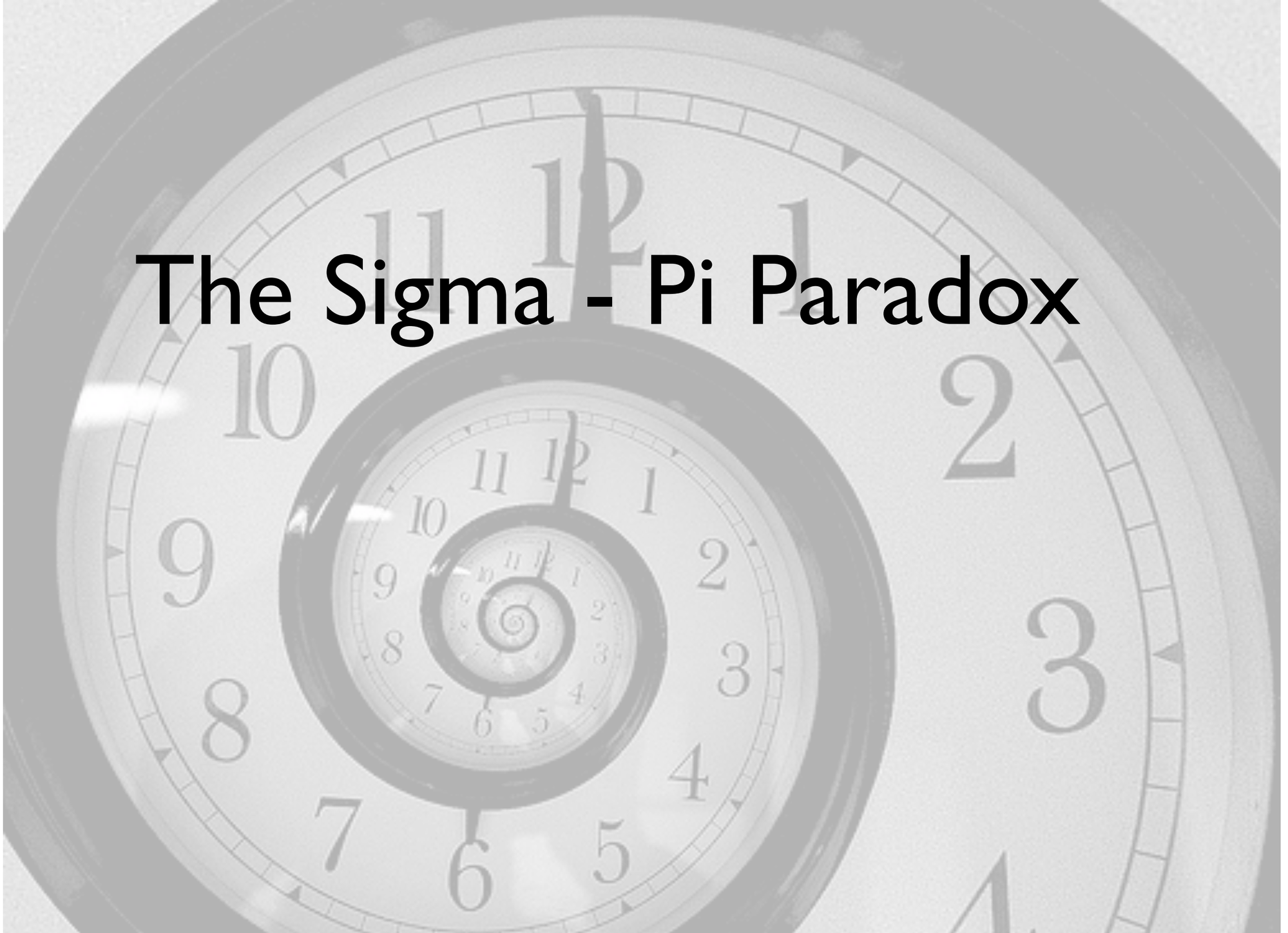
Biological Complexity Consists in Overcoming Evolutionary Light Speed
: transcending the constraints of natural selection



2% sequence translates into Extra-genomic Memory & Inference mechanisms

“When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness. All men felt themselves to be the masters of an intact and secret treasure. There was no personal or world problem whose eloquent solution did not exist in some hexagon. The universe was justified, the universe suddenly usurped the unlimited dimensions of hope. At that time a great deal was said about the Vindications: books of apology and prophecy which vindicated for all time the acts of every man in the universe and retained prodigious arcana for his future. Thousands of the greedy abandoned their sweet native hexagons and rushed up the stairways, urged on by the vain intention of finding their Vindication. These pilgrims disputed in the narrow corridors, proferred dark curses, strangled each other on the divine stairways, flung the deceptive books into the air shafts, met their death cast down in a similar fashion by the inhabitants of remote regions. Others went mad ... The Vindications exist (I have seen two which refer to persons of the future, to persons who are perhaps not imaginary) but the searchers did not remember that the possibility of a man's finding his Vindication, or some treacherous variation thereof, can be computed as zero.”

J. L. Borges. The Library of Babel. 1941.



The Sigma - Pi Paradox

FEBS 15305

FEBS Letters 362 (1995) 257–260

Hypothesis

An estimation of minimal genome size required for life

Mitsuhiro Itaya*

Mitsubishi Kasei Institute of Life Sciences, 11-Minamiooya, Machida-shi, Tokyo 194, Japan

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10004–10006, September 1996

Commentary

The minimal cell genome: "On being the right size"

*Jack Maniloff**

Department of Microbiology and Immunology, University of Rochester, Medical Center Box 672, Rochester, NY 14642

COMPARATIVE GENOMICS, MINIMAL GENE-SETS AND THE LAST UNIVERSAL COMMON ANCESTOR

Eugene V. Koonin

$$F(L) = \prod_{i}^{L} w_i \Rightarrow F(L \backslash 1) = 0$$

Sigma - Pi $$F(L) = \sum_{i}^{L} w_i \rightarrow \prod_{i}^{L} w_i$$

No minimum $$F(L) = \sum_{i=1}^{j} w_i + \prod_{i=j+1}^{L} w_i$$

$$j = L \rightarrow j = L - 1 \rightarrow ... \rightarrow j = 1$$