Apocalypses, Firewalls, and Reasoning about Postselection,

Charles H. Bennett (IBM Research)

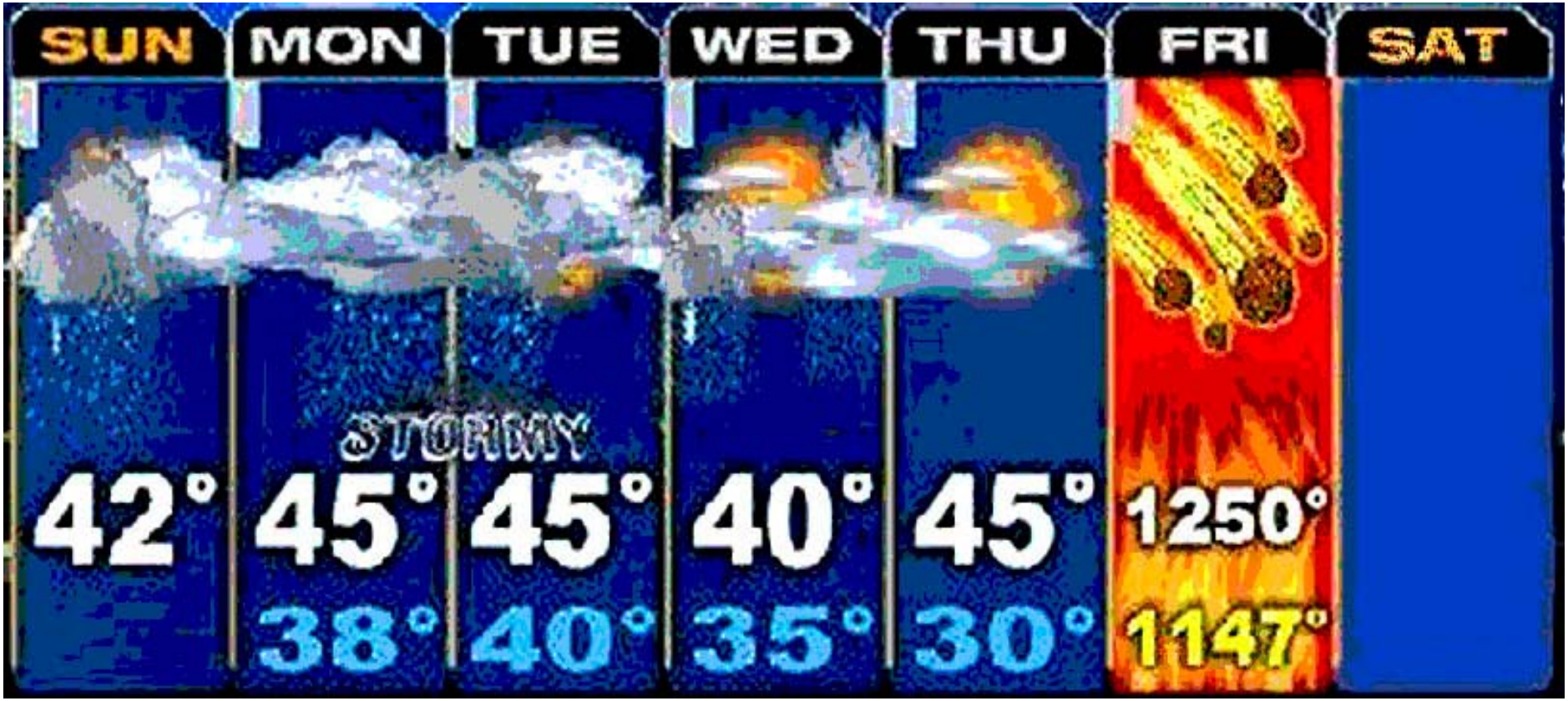
John Preskill's 60th Birthday Confererence
Pasadena, CA

Ides of March, 2013

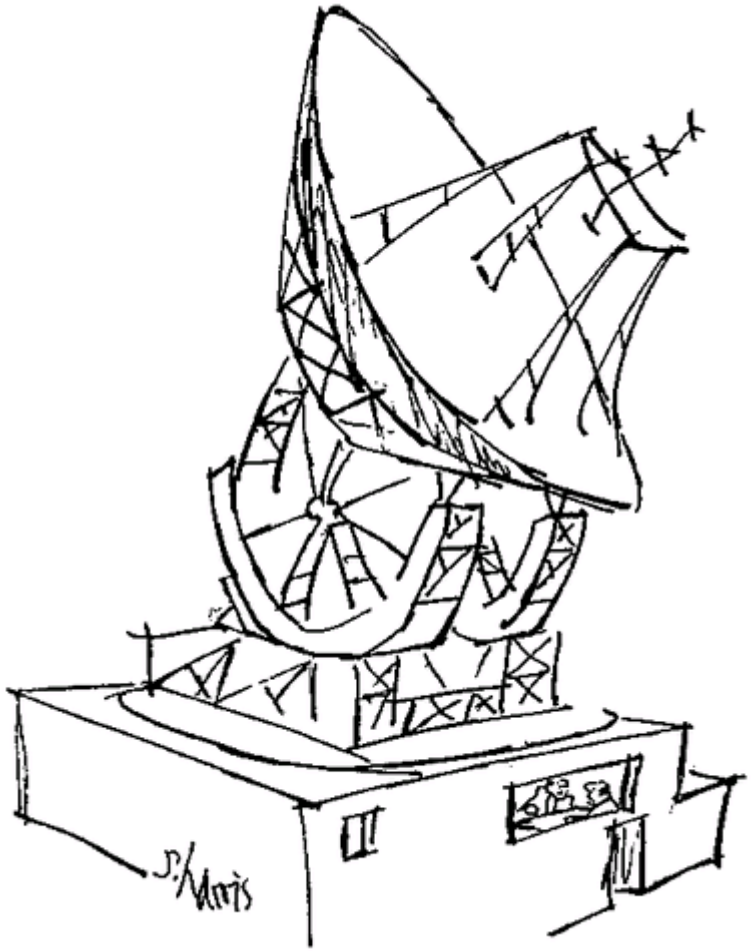
Yesterday Debbie argued that post-selecting CTCs probably cannot be made fault-tolerant.

If we try to use such a device to distinguish near-identical quantum states, nature will tend to avoid paradox in the least unlikely way possible, for example a funding cut or meteor strike that prevents the device from being used, instead of post-selecting against an otherwise near-certain outcome.

How do we reason in the presence of post-selection, for example about the likelihood of the universe ending soon? How well can we justify our gut feeling that the universe is unlikely to end before tomorrow, say by a vacuum phase transition?



Last year there were rampant warnings on the Internet that the world would end on Friday December 21, due to the wraparound of the Mayan Calendar. Hearing this, my 4 year old granddaughter said, “That’s silly. The world isn’t going to end.”



Principle of Mediocrity, or Copernican Principle, favors theories according to which phenomena actually observed are typical of those predicted by the theory.

"If we take our current model of the universe and run it backward 13 billion years, we get something resembling Donald Duck. There must be something wrong here."

Doomsday arguments illustrate undisciplined reasoning about the future, neglecting selection bias:

"I am typical; therefore between 5 and 95 per cent of all people who will ever live already have."

or

"I am typical, therefore between 5 and 95 per cent of duration of human history has already passed."

Carlton Caves' birthday party rebuttal the doomsday argument, arXiv:0806.3538,

Imagine wandering into a birthday party and learning that the celebrant is 60 years old. Then there is a 1/2 chance he will live to be 120 years old and a 1/3 chance to 180.

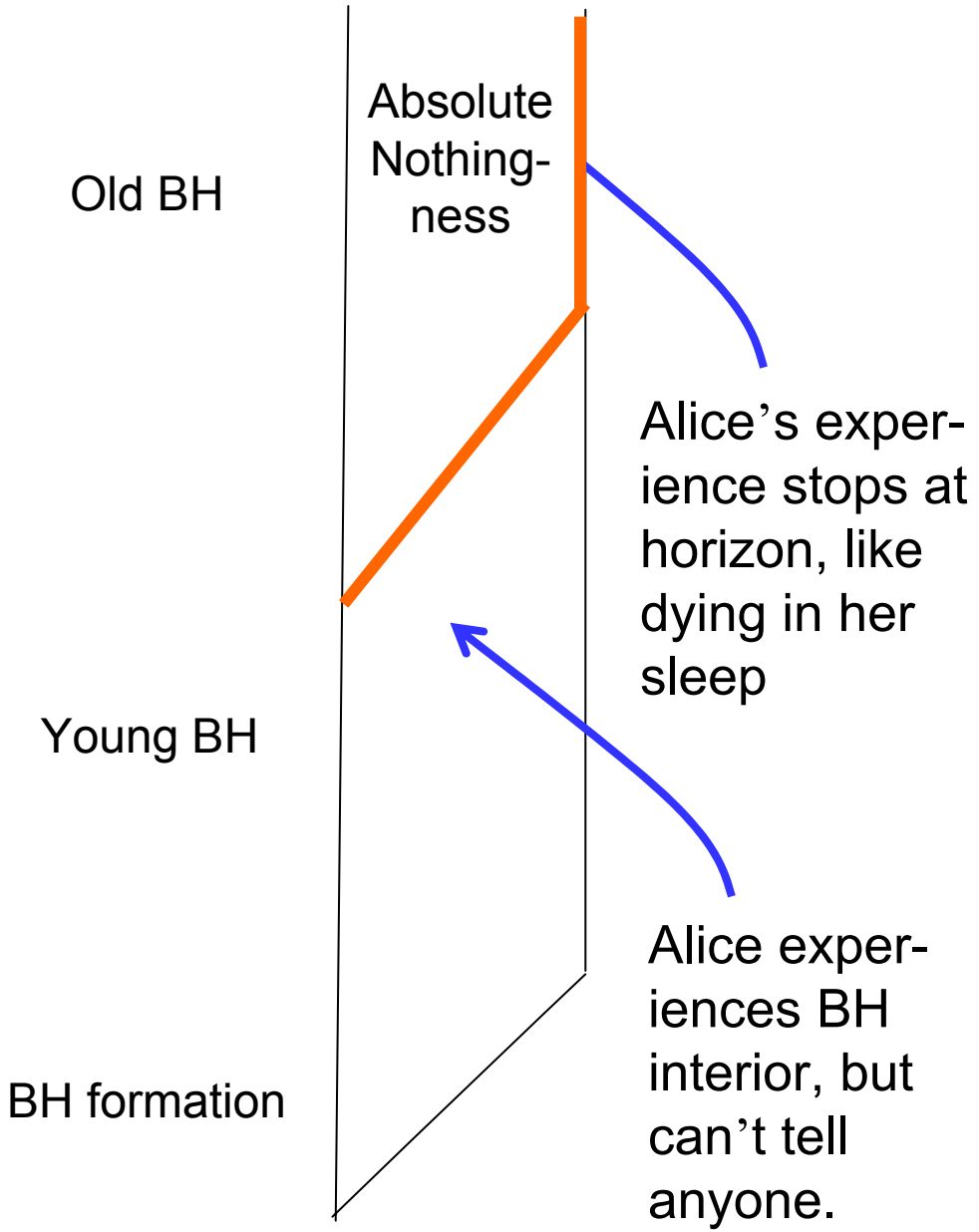
Conversely, on encountering a one day old baby, is it reasonable to warn the parents that it probably won't live to be more than 20 days old?

The *German Tank Problem* exemplifies more disciplined reasoning.

During WWII, the, Allies fairly accurately estimated the number of German tanks based on the serial numbers of a few captured tanks. Such estimates become rapidly more reliable the more tanks are captured.

# captured	estimate	95% confidence interval	(m is highest serial number observed)
1	2m +1	[m, 20m]	
2	1.5m +1	[m, 4.5m]	
5	1.2m +1	[m, 1.8m]	
20	1.05m +1	[m, 1.16m]	

Susskind's firewall picture



If firewalls don't exist, Alice can learn of it privately, but can't publish this "fact." Worse than 1 tank.

If firewalls do exist, Alice could gather statistical evidence for their existence by repeating the experiment many times and noting she always found herself outside, never inside. But she can only do the experiment once, nor can several Alices pool their results.

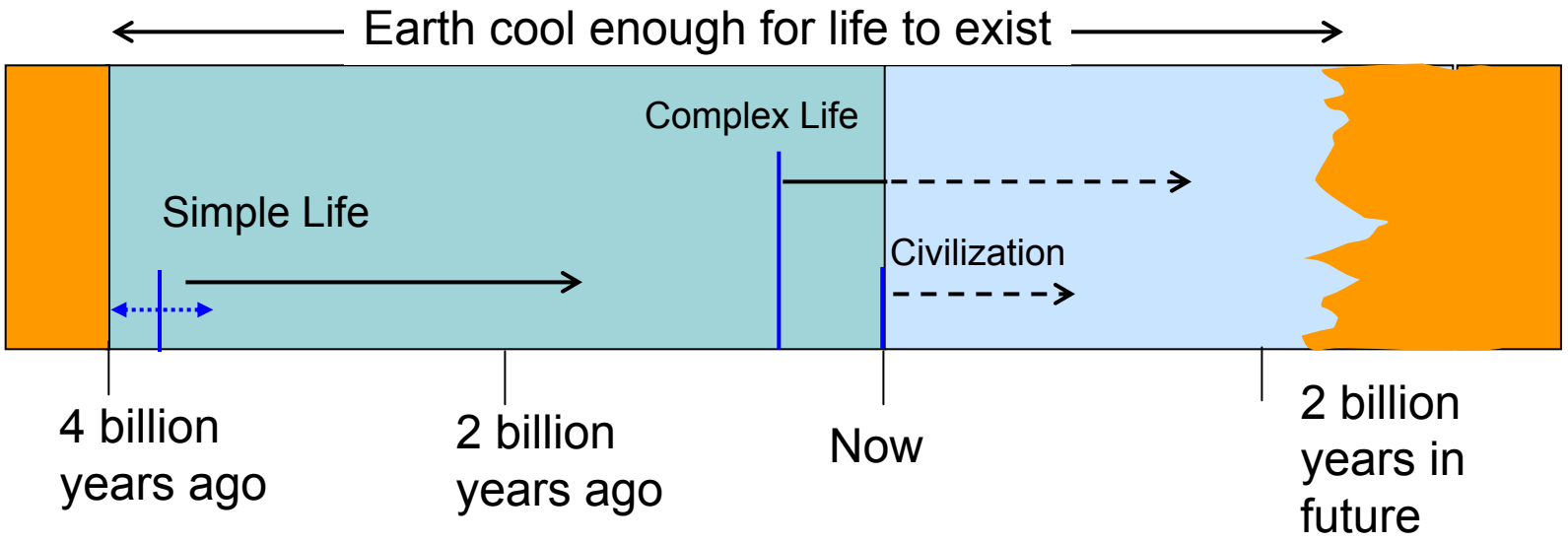
Does this mean that the presence or absence of firewalls is an ontologically moot or nearly moot question, by the same general principles as used to justify BH complementarity?

In fact many people, especially dictators, fancy themselves as *atypical*, occupying a privileged temporal position at the very beginning of a long future era.

Not only dictators, but also ordinary people, like to think they're special. I learned this inspirational slogan as a hippie, during the Summer of Love (1967) in Berkeley and San Francisco.

Later I found that it was probably coined by Charles Dederich, founder of the ill-fated drug rehabilitation organization Synanon.

Mediocrity and Earth History



Origin of simple (prokaryotic) life appears to have been likely, given earth-like conditions, because it happened so early; but origin of complex life and civilization seem rather unlikely (per billion years), because these did not occur until after a significant fraction of the available time had elapsed.

If simple life were *unlikely*, and had only occurred by virtue of being anthropically post-selected as a prerequisite civilization, one would expect it to have originated nearer the middle of the available time.

Civilization would seem to be severely non-mediocre, requiring some explanation of why we find ourselves in the first few millionths of the presumable remaining time available.

Possible explanations of the youngness of civilization

• **Anthropogenic destruction** of civilization (though less likely of life or entire human species). This is distasteful and has various problems.
    • Why can't we protect ourselves from it?
        • by becoming cooperative and peaceful
        • by colonizing space
    • Why don't we see the remains of previous civilizations?
        • Maybe because civilization may be rare even given life. That explanation comports with Fermi paradox.

• **Survivor selection against vacuum phase transitions** (VPT) happening at an expected rate around once every few thousand years?
    • How world look different if VPTs were expected once per second?
        • Like a Boltzmann brain, but not of the classic sort of Boltzmann brain that can't tell it's alone. One able to be conscious and think, but not having fake memories of remote places and distant past.

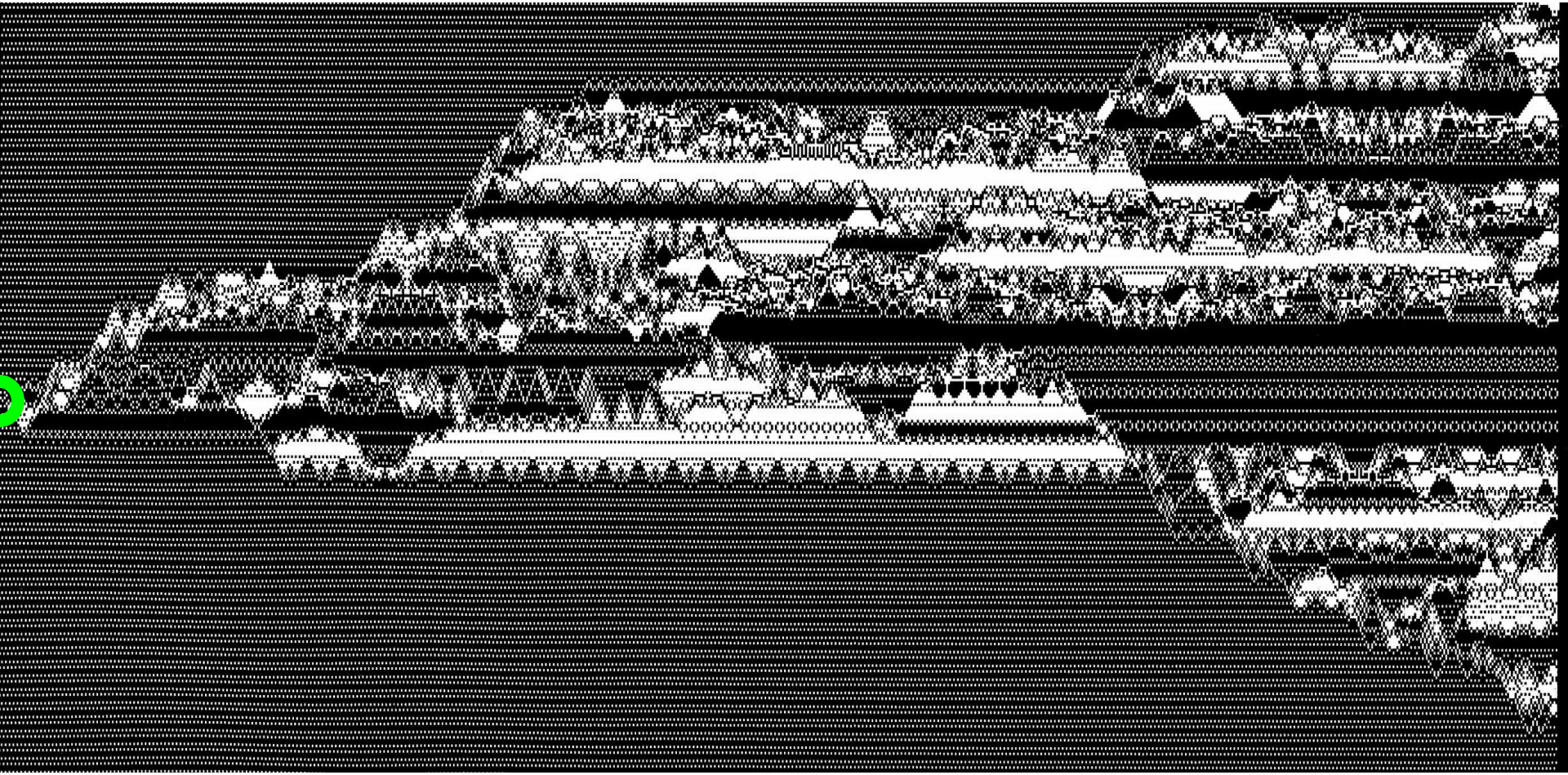
• **Perpetual newness.** Maybe 1 billion years from now there will still be people, or our cultural descendants, but they will be preoccupied by some other qualitatively new feature of their existence and ask why ***it*** didn't happen earlier. They will still worry, that by the mediocrity principle, life ***as they know it*** may be about to disappear.

Returning now to the origin of life, one may ask: How likely is it under auspicious nonequilibrium boundary conditions? (This godless creation—a bright flip side to the godless hell of heat death—nowadays seems to worry creationists even more than Darwin's initially more inflammatory thesis that humans are descended from apes.)

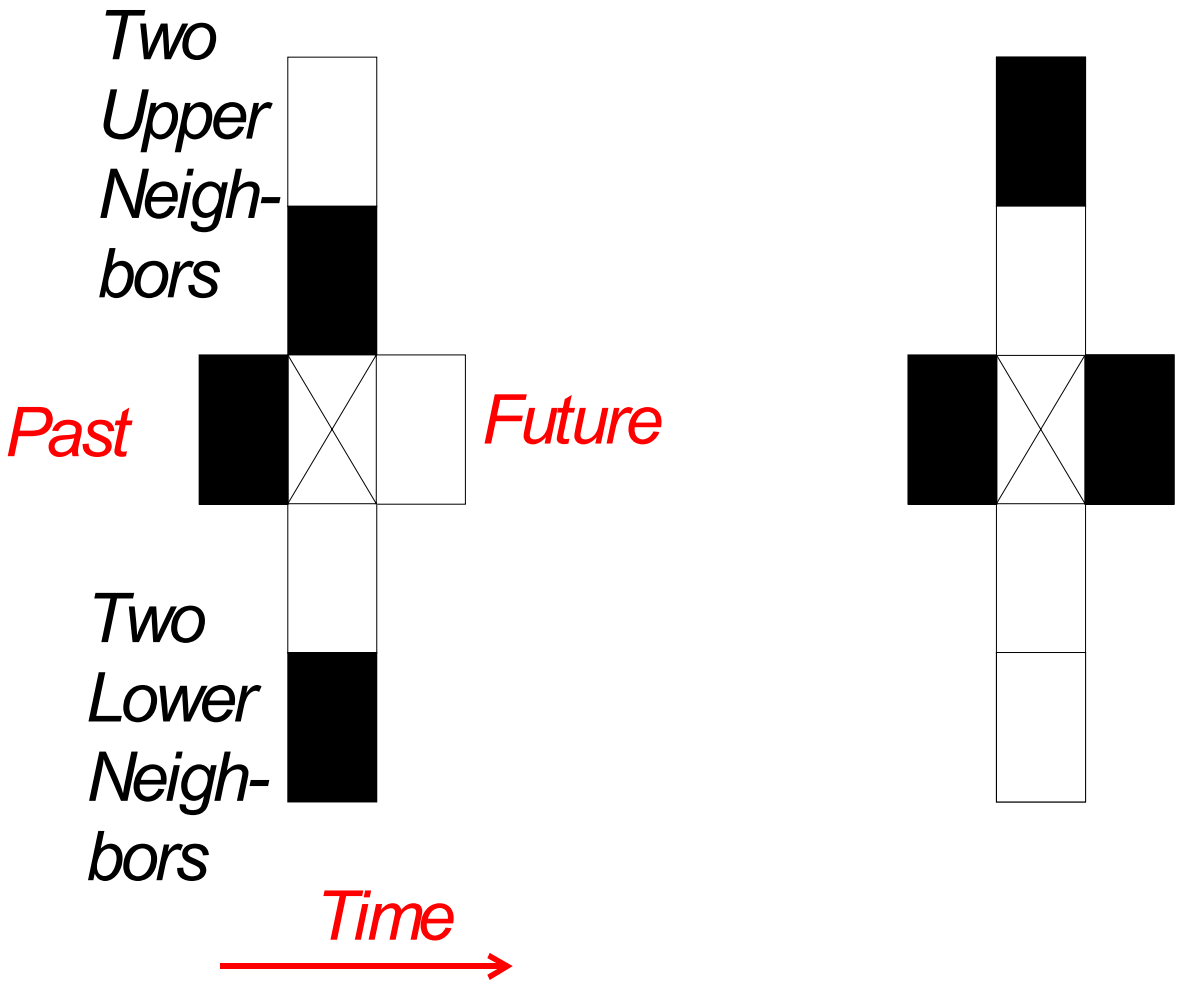
primitive earth

much later

Simple dynamical processes (such as this 1 dimensional reversible cellular automaton) are easier to analyze and can produce structures of growing "complexity" from simple initial conditions. time →



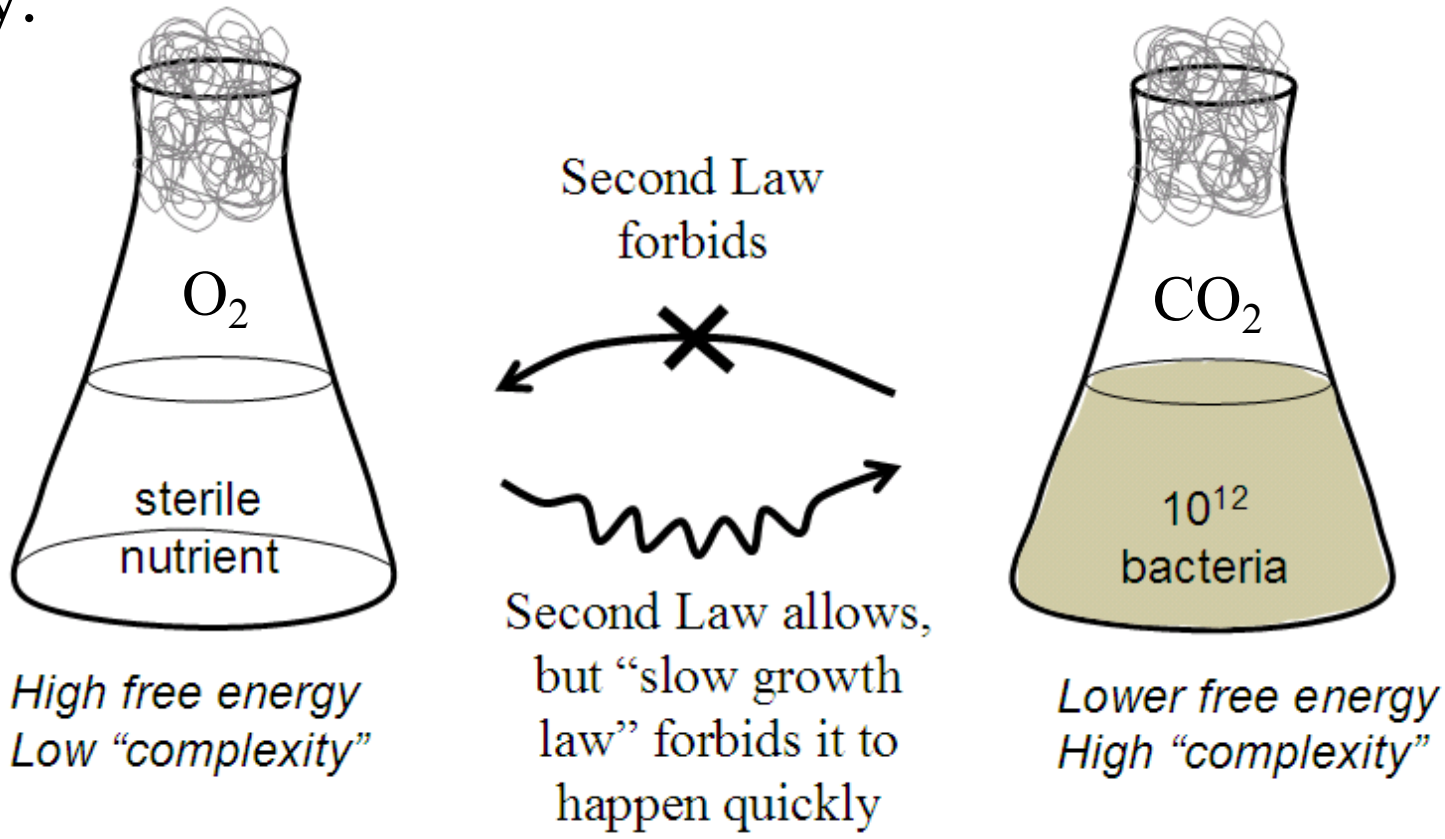
Small irregularity (green) in otherwise periodic initial condition produces a complex deterministic wake.



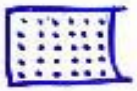
Range-2, deterministic, 1-dimensional Ising rule. Future differs from past if exactly two of the four nearest upper and lower neighbors are black and two are white at the present time.

"Complexity" cannot be identified with situations of maximum order, or maximum disorder, or indeed any intermediate amount of order.

It is not a thermodynamic potential, like entropy or free energy.



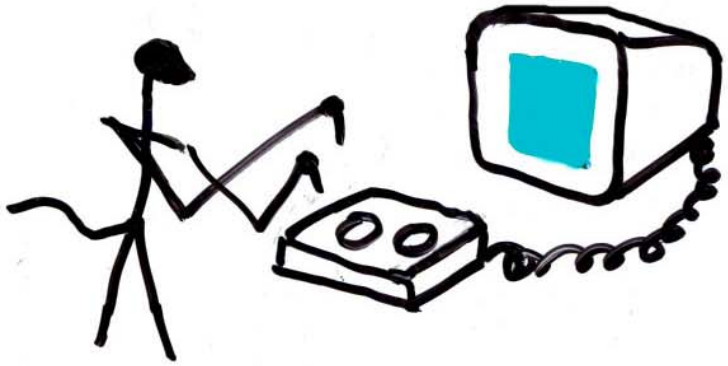
Candidates for "Complexity"

- Ability to grow, reproduce, metabolize  ; Fitness 
- Randomness  
- Order  
- Free Energy  
- Information , Correlation  
- Logical Depth – internal evidence of a long and eventful history.

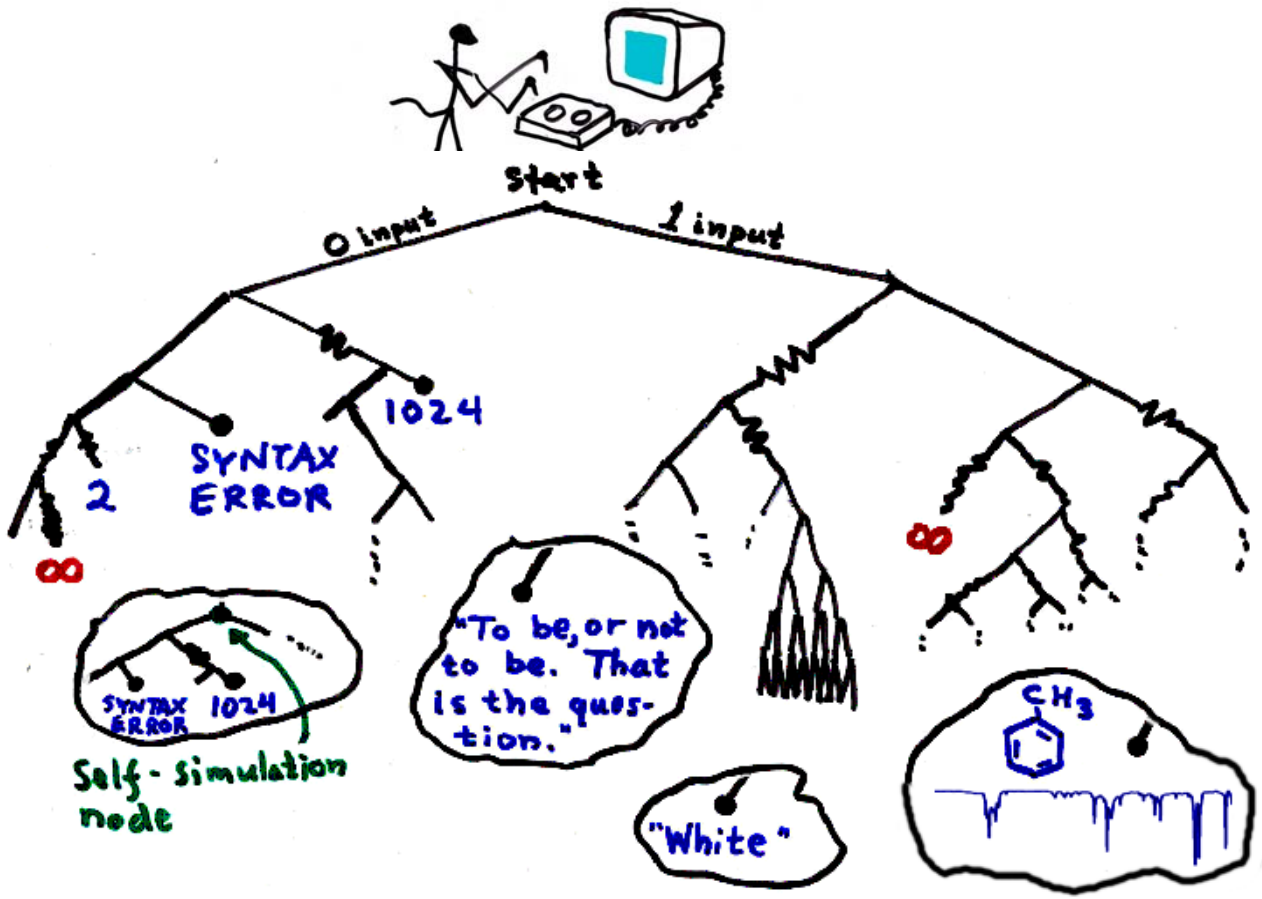
- Hierarchical Structure 
- Computational Universality 

∴

Defining complexity: use a computerized version of the old idea of a monkey at a typewriter eventually typing the works of Shakespeare. Of course a modern monkey uses a computer instead of a typewriter.

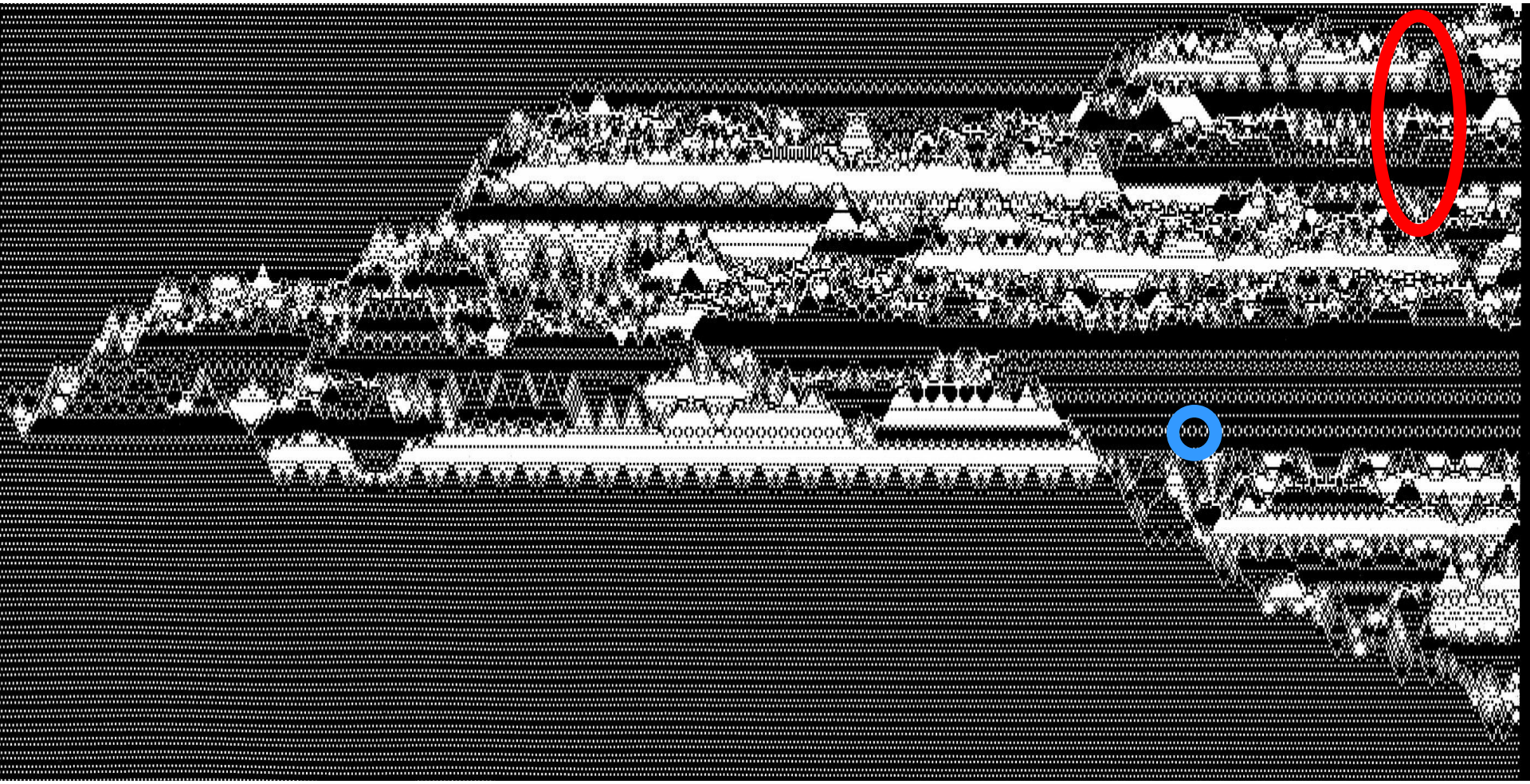


A monkey randomly typing 0s and 1s into a universal binary computer has some chance of getting it to do any computation, produce any output (Chaitin 1975)



The input/output graph of this or any other universal computer is a microcosm of all cause/effect relations that can be demonstrated by deductive reasoning or numerical simulation. The (semicomputable) output distribution of this computer is called the Universal Semimeasure, or Algorithmic Probability, or Universal Prior (Chaitin, Levin et al late 20th C.)

Subjectively complicated structures typically are “logically deep,” containing evidence of a nontrivial causal history.



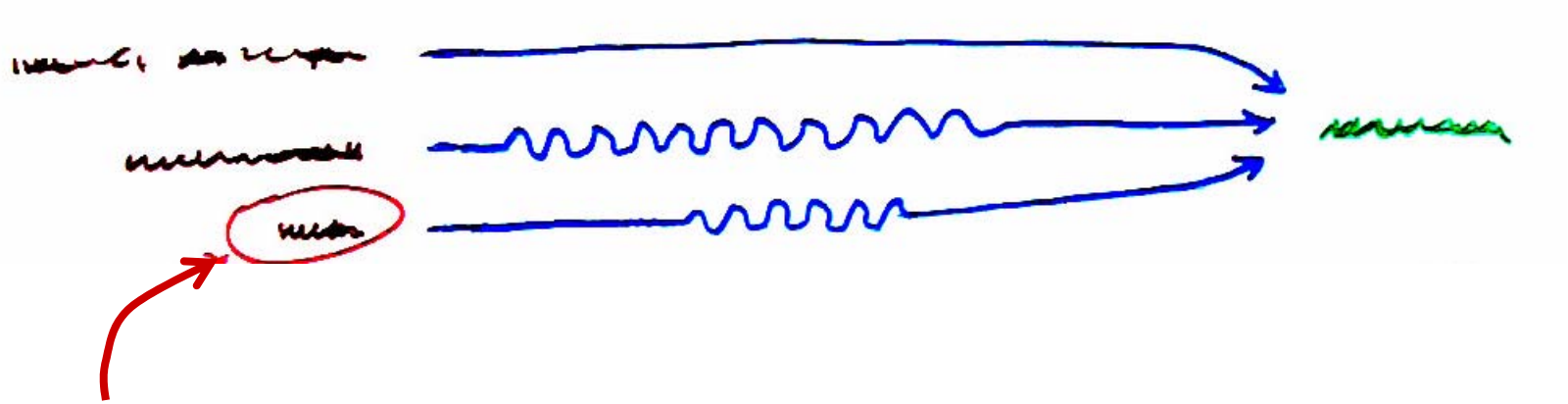
A sufficiently big piece of the wake (red) contains enough evidence to infer the whole history. A smaller pieces (blue) does not.

In the philosophy of science, the principle of Occam's Razor directs us to favor the most economical set of assumptions able to explain a given body of observational data.

Alternative hypotheses

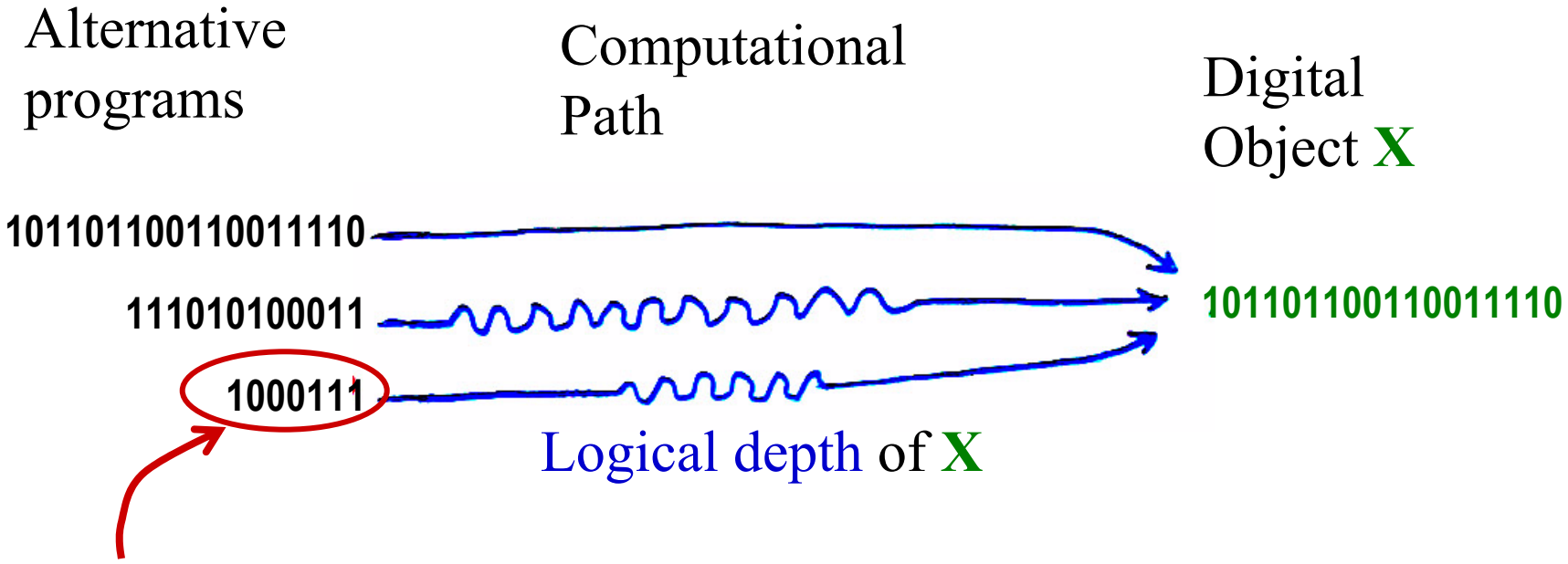
Deductive path

Observed Phenomena



The most economical hypothesis is preferred, even if the deductive path connecting it to the phenomena it explains is long and complicated.

In a computerized version of Occam's Razor, the hypotheses are replaced by alternative programs for a universal computer to compute a particular digital or digitized object **X**.



The shortest program is most plausible, so its *run time* measures the object's logical depth, or plausible amount of computational work required to create the object.

To make the quantitative definition of logical depth more stable with respect small variations of the string x, and the universal machine U, a two-parameter definition is used:

A string x has depth d, at significance level s if for all programs p such that $U(p)=x$, the program p is compressible by at least s bits (i.e. there is another program p^*, at least s bits shorter than p, such that $U(p^*)=p$).

This formalizes the notion that all hypotheses for producing x in less than d steps suffer from at least s bits worth of ad-hoc assumptions.

Thus defined, depth obeys the slow-growth law.

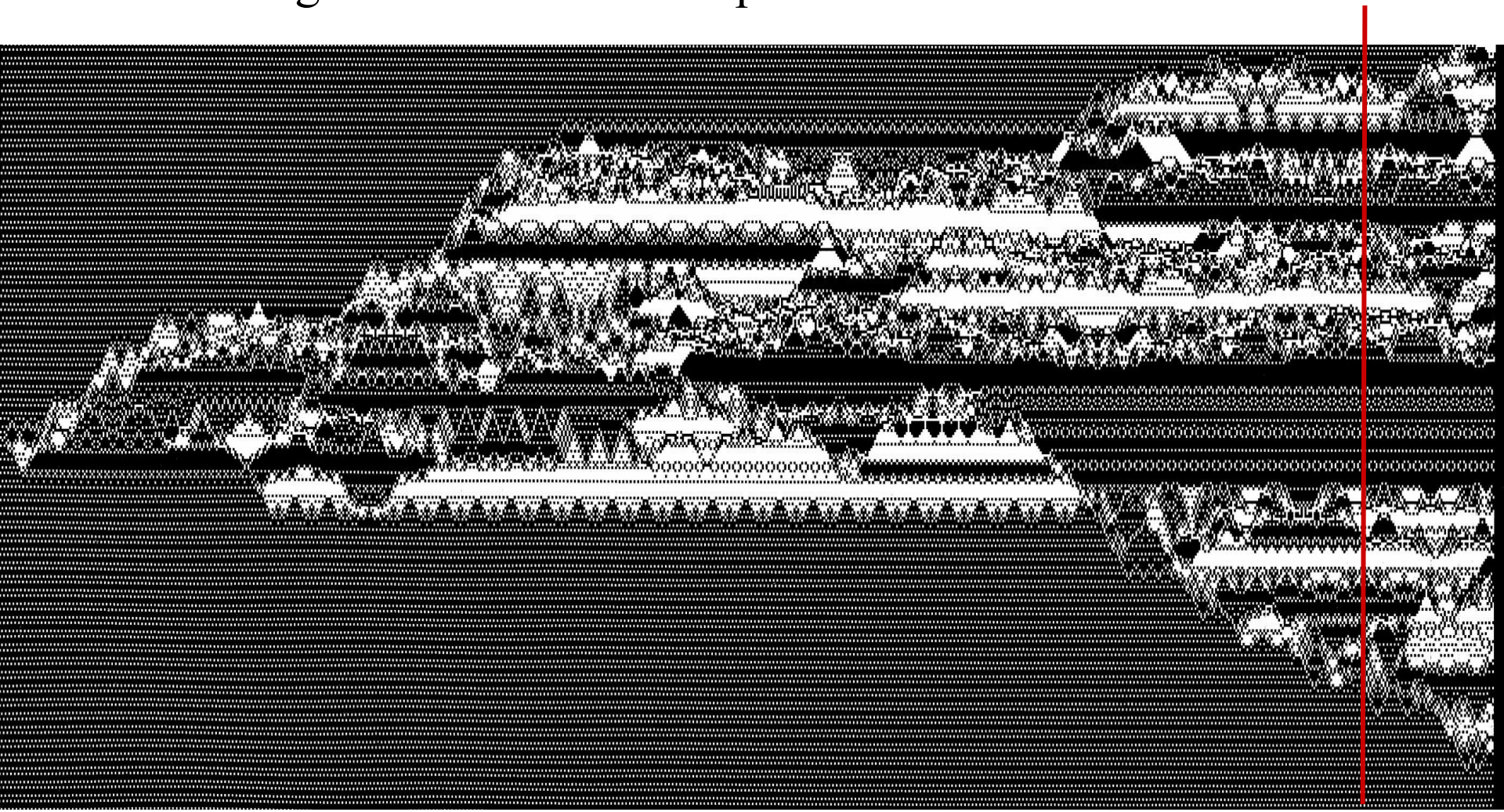
A trivially orderly sequence like 111111… is logically shallow because it can be computed rapidly from a short description.

A typical random sequence, produced by coin tossing, is also logically shallow, because it essentially **its own** shortest description, and is rapidly computable from that.

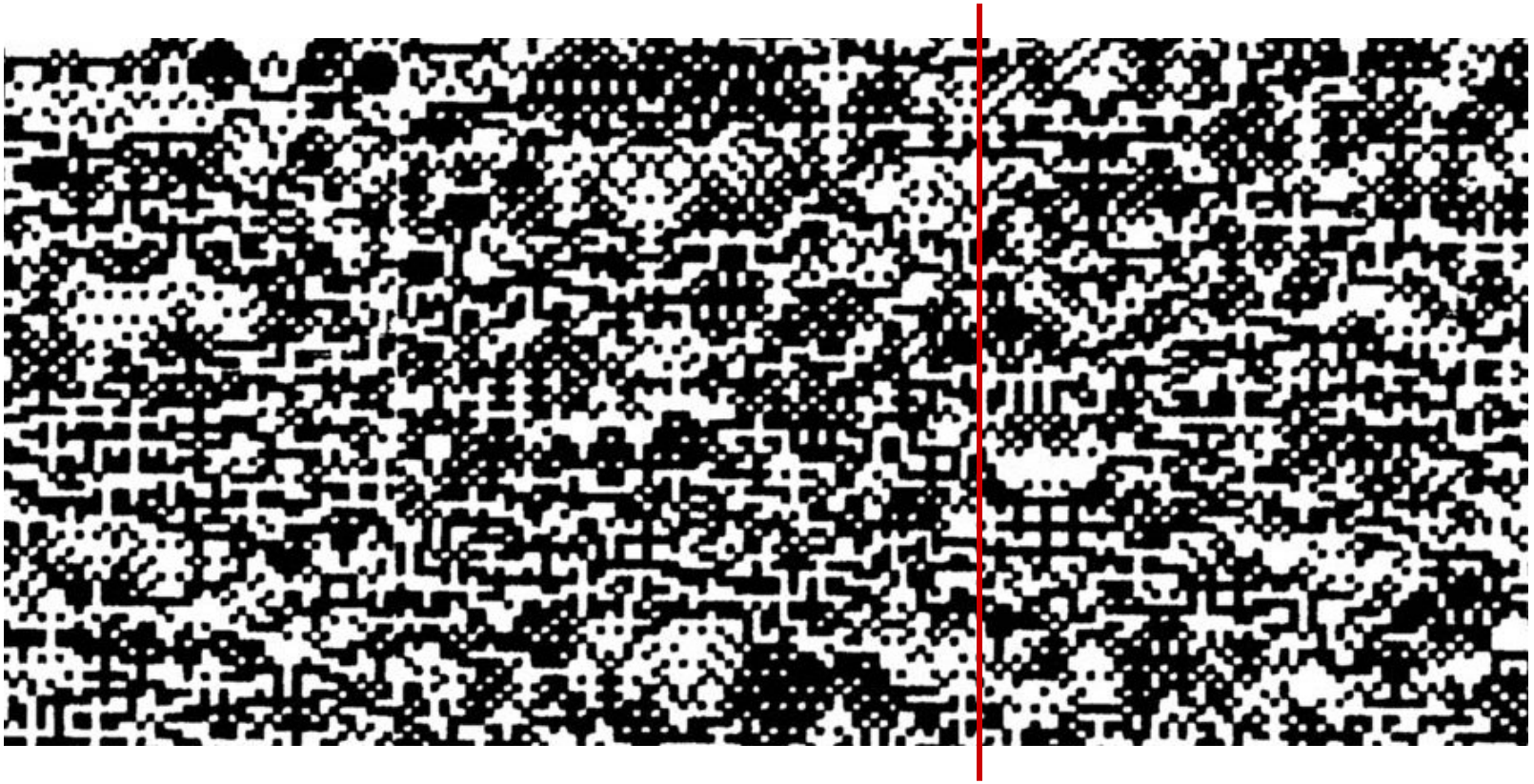
Trivial semi-orderly sequences, such as an alternating sequence of 0's and random bits, are also shallow, since they are rapidly computable from their random part.

(Depth is thus distinct from, and can vary independently from *Kolmogorov complexity* or *algorithmic information content*, defined as the **size** of the minimal description, which is high for random sequences. A sequence's Kolmogorov complexity measures its randomness, not its complexity in the sense intended here.)

Initially, and continuing for some time, the logical depth of a time slice increases with time, corresponding to the duration of the slice's actual history, in other words the computing time required to simulate its generation from a simple initial condition.



But if the dynamics is allowed to for a large random time after equilibration (comparable to the system's Poincaré recurrence time, exponential in its size), the typical time slice becomes shallow and random, with only short-range correlations.



The minimal program generating this time slice not by retracing its actual long history, but a short computation short-circuiting it.

Cosmologists worry about typicality, especially in connection with eternal inflation, where it is hard to find a non-pathological prior distribution over "all possible universes"

e.g.

- D. Page, *Typicality Defended* hep-th arxiv:707.4169
- A. Garriga and J. Valenkin *Prediction and Explanation in the Multiverse* hep-th arxiv:0711.2559v3

Cosmological models like eternal inflation resemble the rest of science in being based on evidence acquired from observation and experiment.

But one could use instead look to embed the set of "all possible universes" in a purely mathematical construct like the Monkey Graph, which is untainted by physics.

Lloyd and Dreyer *the Universal Path Integral* arxiv:1302.2859 recently sketched how to embed path integral theories within the monkey graph.

Thinking about selection bias helps us think take anthropic effects into account while with as little anthropocentrism as possible.

- What prior distribution to use? Answer: Universal Prior.

Do we want to include any "universal" physical principles?

- Reversibility/thermodynamics?
- Superposition – quantum mechnaics
- Locality / field theories?

- What sort of "sentient observer" criterion we should post-select on?
  - Computational universality (then self-organization is likely).
  - *Science* itself: How to define?
    - Gell-Mann's IGUS or Information Gathering and Utilizing Systems-–a universe that studies and stores information about itself. That would seem to presume thermodynamic disequilibrium emergent classicality, and perhaps locality.