Quantum Information & Cosmology

Charles H. Bennett
IBM Research, USA

YouQu, HRI, Allahabad, India
28 Feb 2017

- The relation between *Dynamics*—the spontaneous motion or change of a system obeying physical laws—and *Computation*—a programmed sequence of mathematical operations

- Self-organization, exemplified by cellular automata and *logical depth* as a measure of complexity.

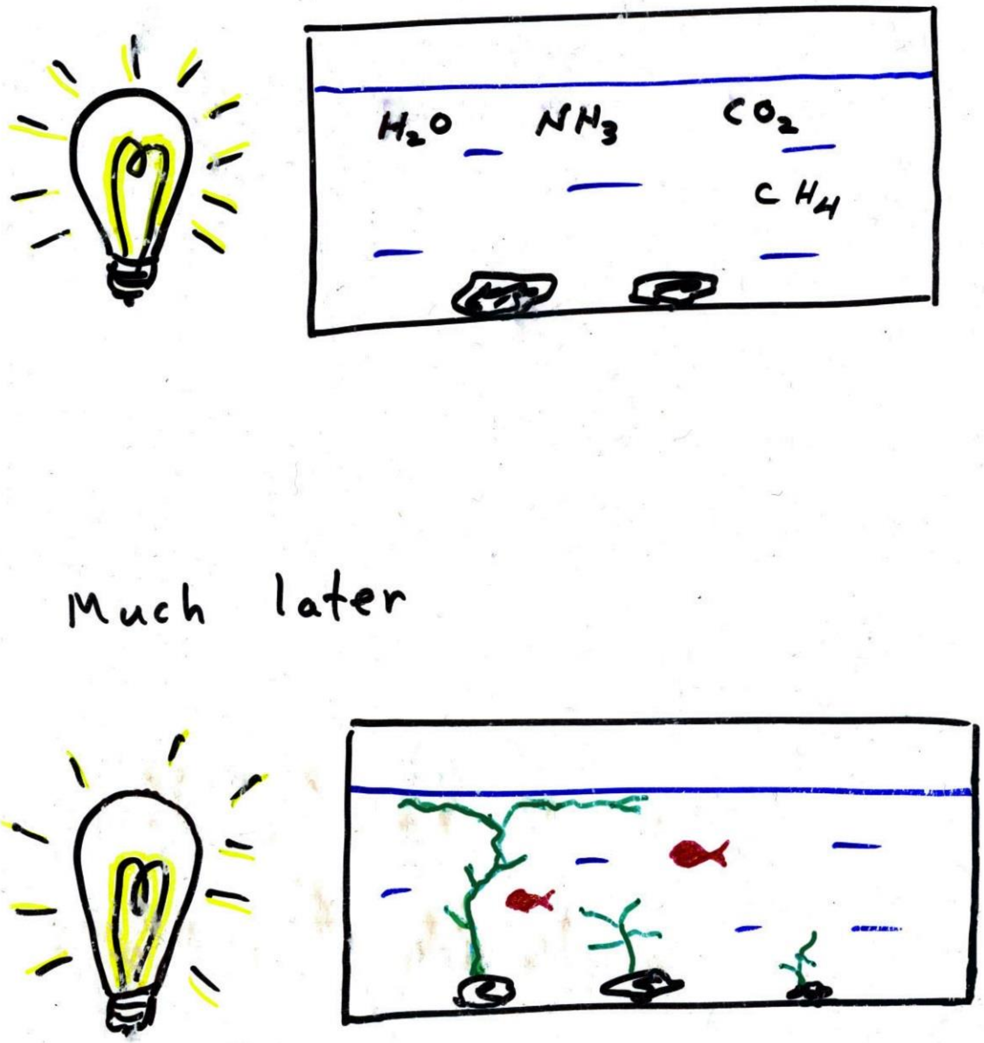
- True and False evidence—the Boltzmann Brain problem at equilibrium and in modern cosmology

- Wigner's Friend—what it feels like to be inside an unmeasured quantum superposition

How does the familiar complicated world we inhabit emerge cosmologically from the austere high-level laws of quantum mechanics and general relativity, or terrestrially from lower-level laws of physics and chemistry?

To attack this question in a disciplined fashion, one must first define complexity, the property that increases when a self-organizing system organizes itself.

A simple cause can have a complicated effect, but not right away.



A good scientific theory should give predictions relative to which the phenomena it seeks to explain are typical.

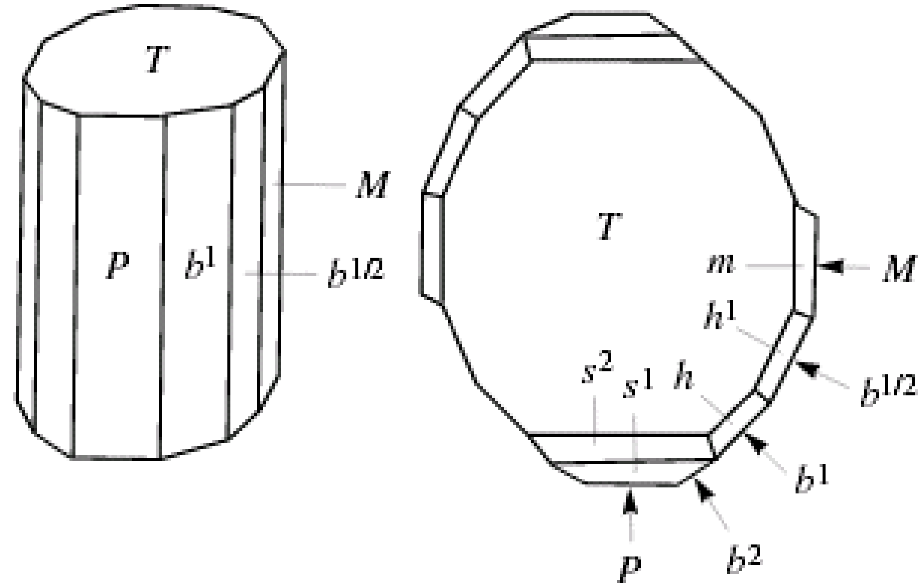
A cartoon by Sidney Harris shows a group of cosmologists pondering an apparent typicality violation

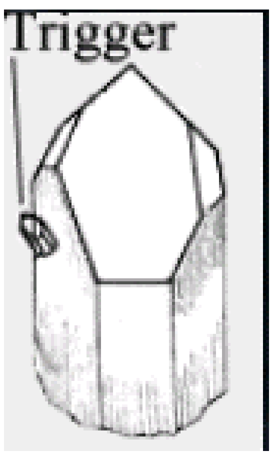
"Now if we run our picture of the universe backwards several billion years, we get an object resembling Donald Duck. There is obviously a fallacy here."

(This cartoon is not too far from problems that actually come up in current cosmology)

Scientific vs. Magical or Anthropocentric Thinking

Pasteur's sketch of sodium ammonium tartrate crystal. Chiral location of hemihedral faces e.g. ***h*** is determined by chirality of molecules within.





A typical crystal site on the Internet
http://www.neatstuff.net/avalon/texts/Quartz-Configurations.html

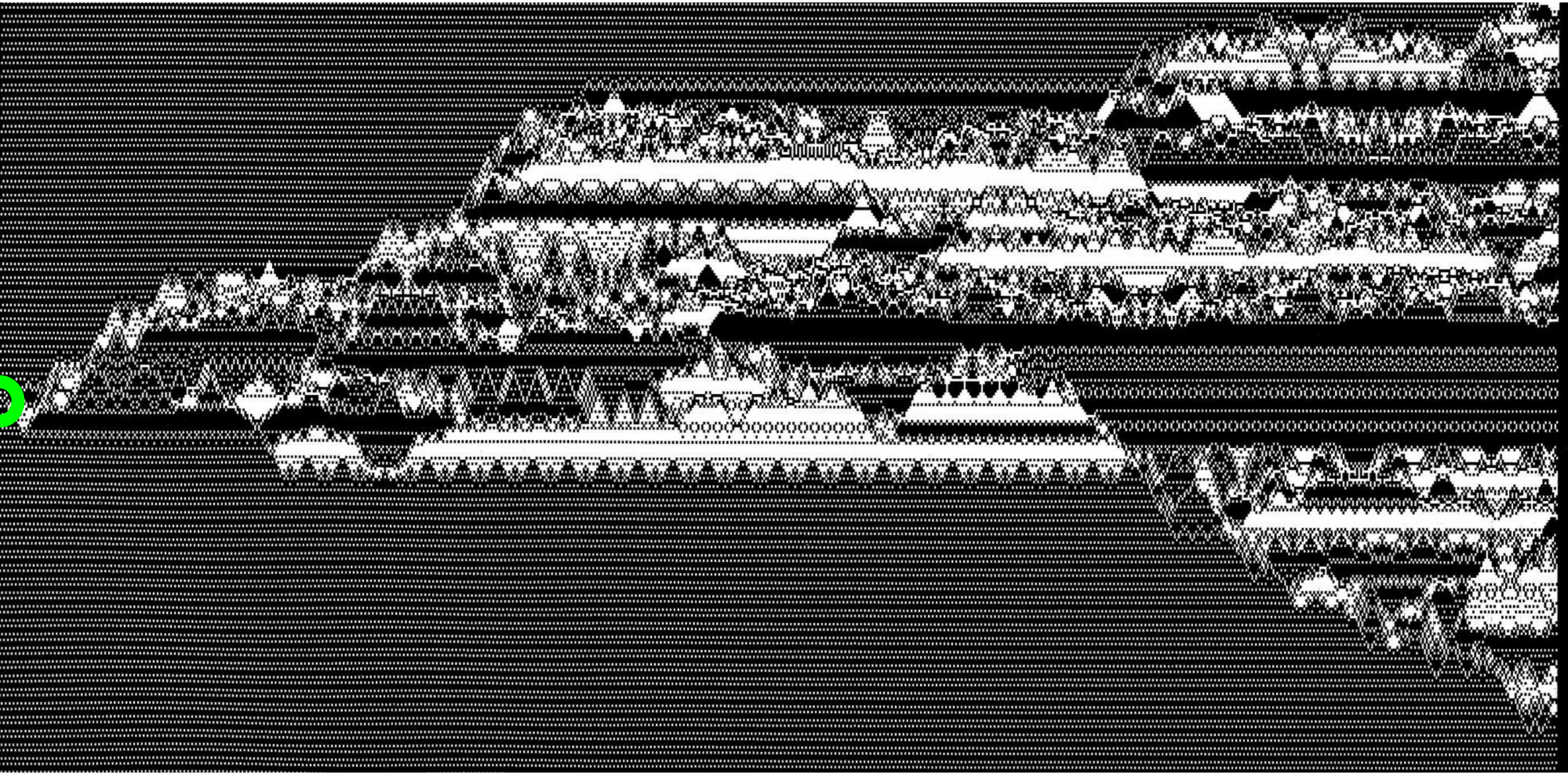
TRIGGER CRYSTALS:

have a **smaller crystal growing out from them**. This `trigger' can be gently squeezed to activate the power of the crystal and strengthen its attributes. These are just used for a surge of a particular kind of energy.

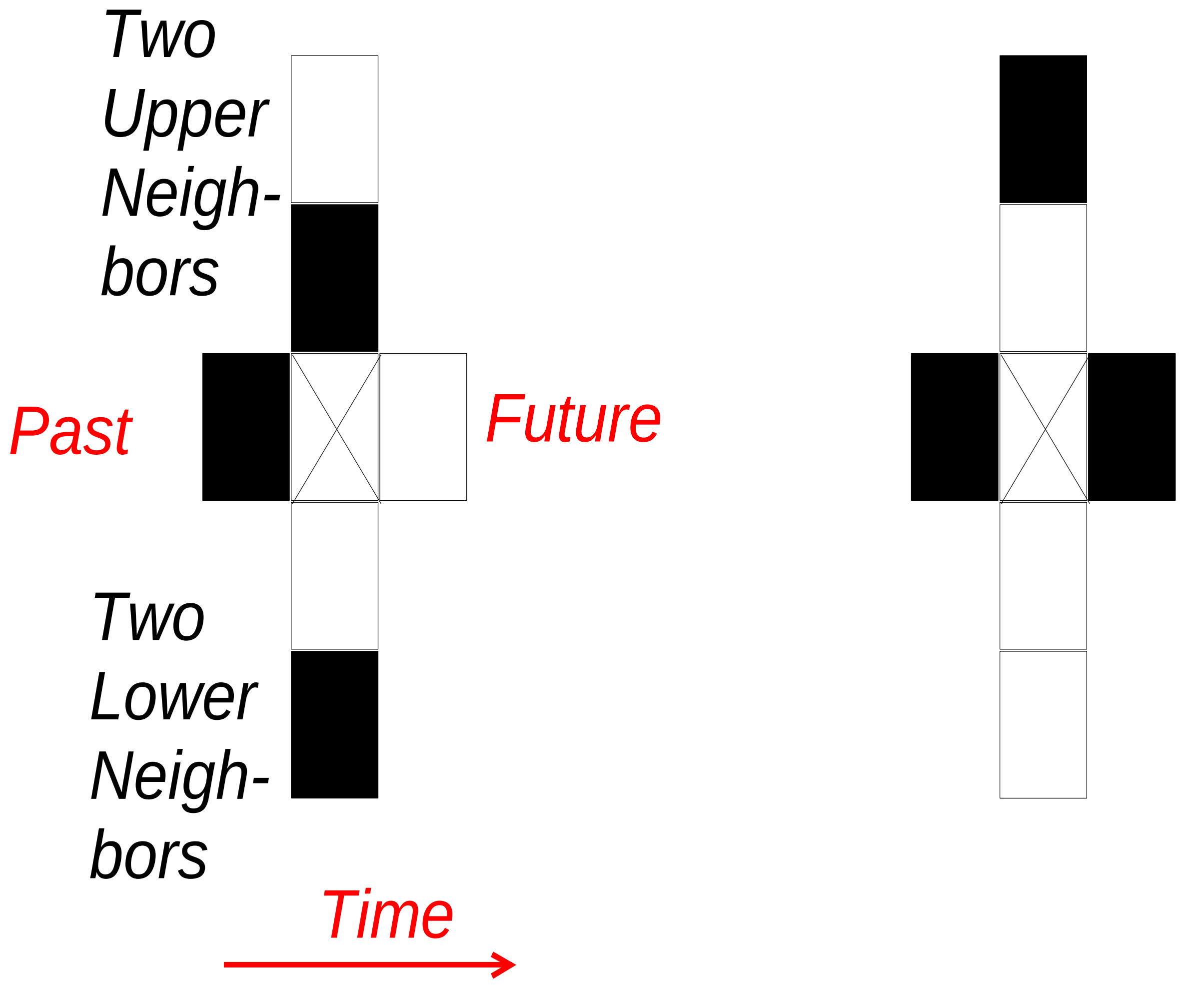
Conclusion: To understand molecules, learn to think like one.

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.



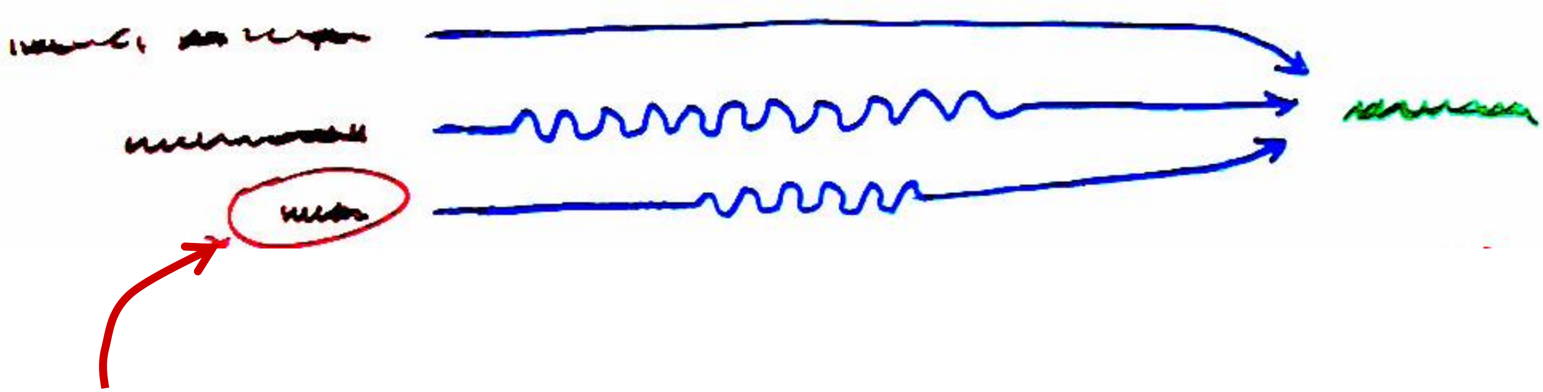
Range-2, deterministic, 1-dimensional Ising rule. Future differs from past if exactly two of the four nearest upper and lower neighbors are black and two are white at the present time.

Occam's Razor

Alternative hypotheses

Deductive path

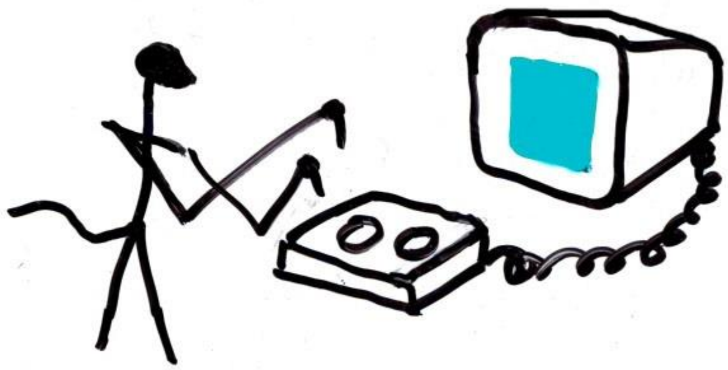
Observed Phenomena



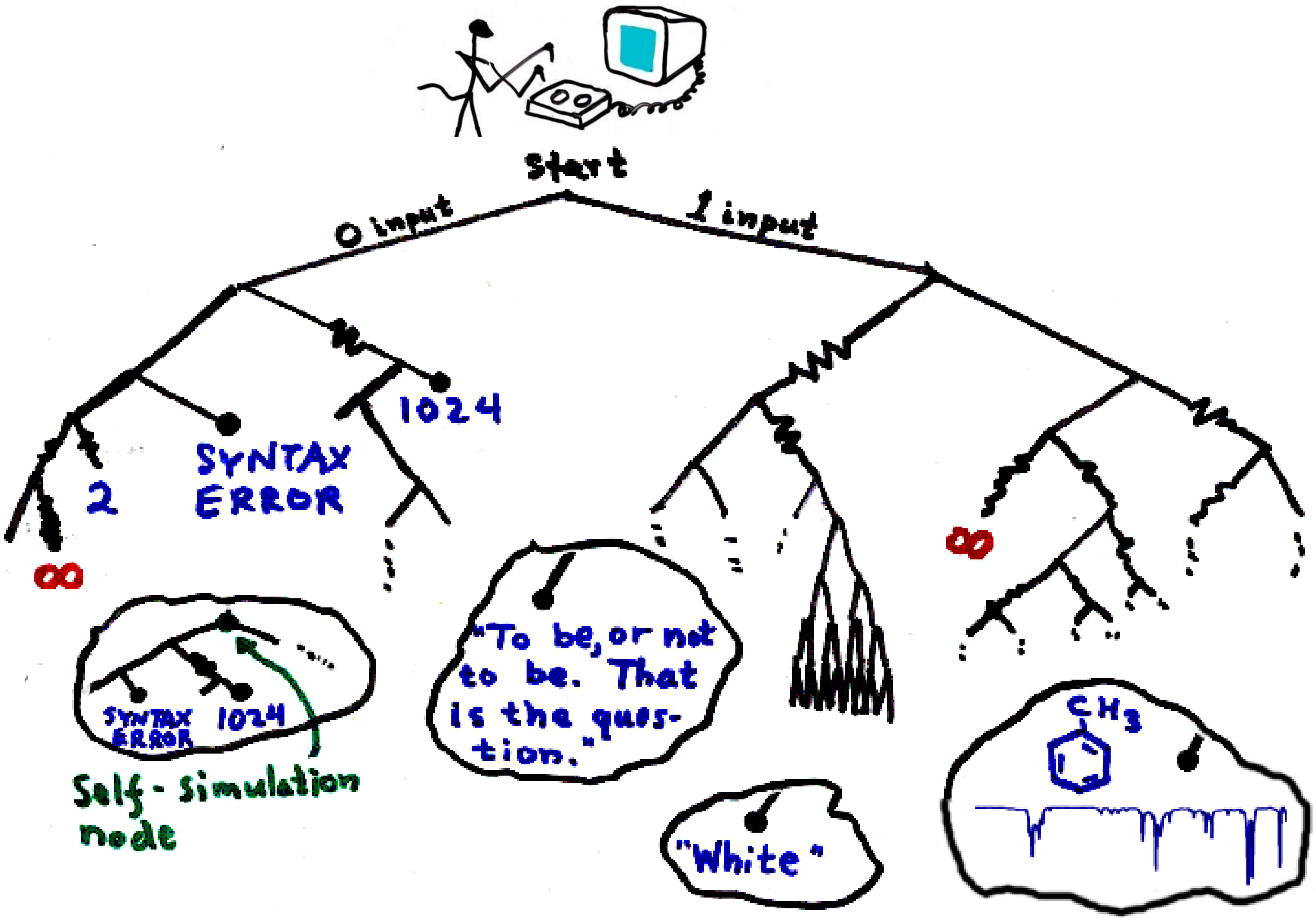
The most economical hypothesis is to be preferred, even if the deductive path connecting it to the phenomena it explains is long and complicated.

But how does one compare economy of hypotheses in a disinterested way?

Algorithmic information, devised in the 1960's by Solomonoff, Kolmogorov, and Chaitin, uses a computerized version of the old idea of a monkey at a typewriter eventually typing the works of Shakespeare.

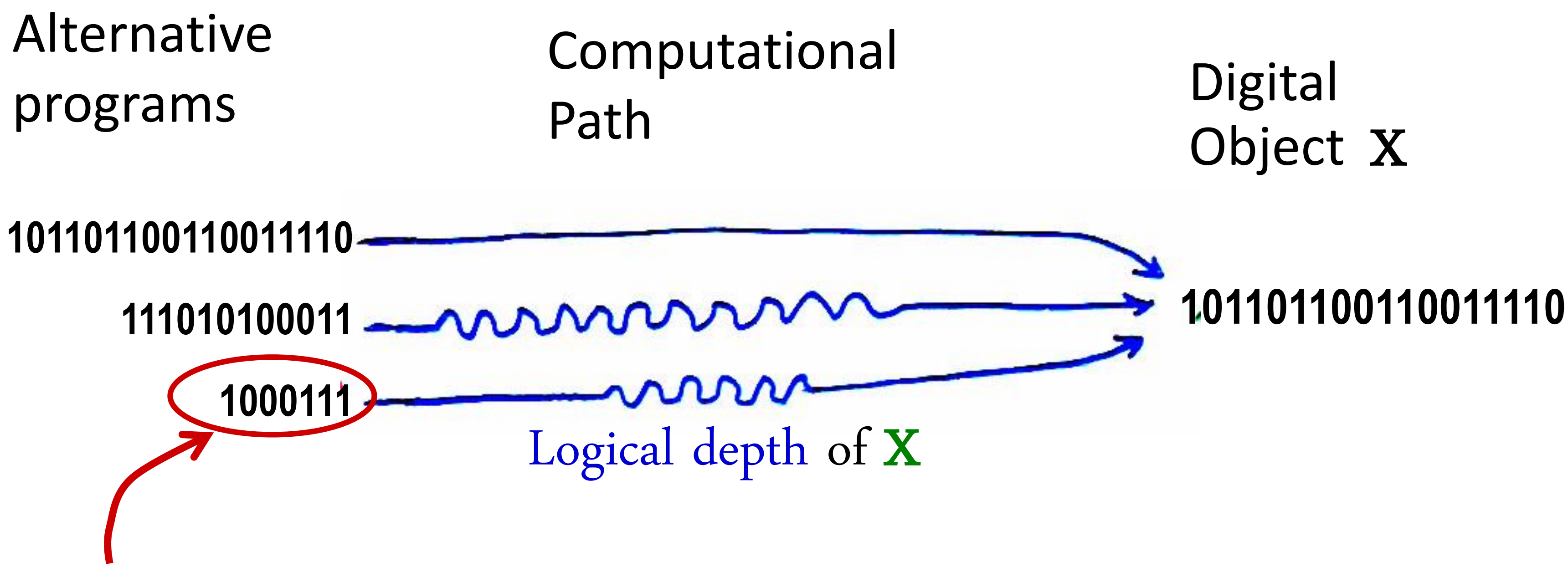


A monkey randomly typing 0s and 1s into a universal binary computer has some chance of getting it to do any computation, produce any output.



This tree of all possible computations is a microcosm of all cause/effect relations that can be demonstrated by deductive reasoning or numerical simulation.

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.

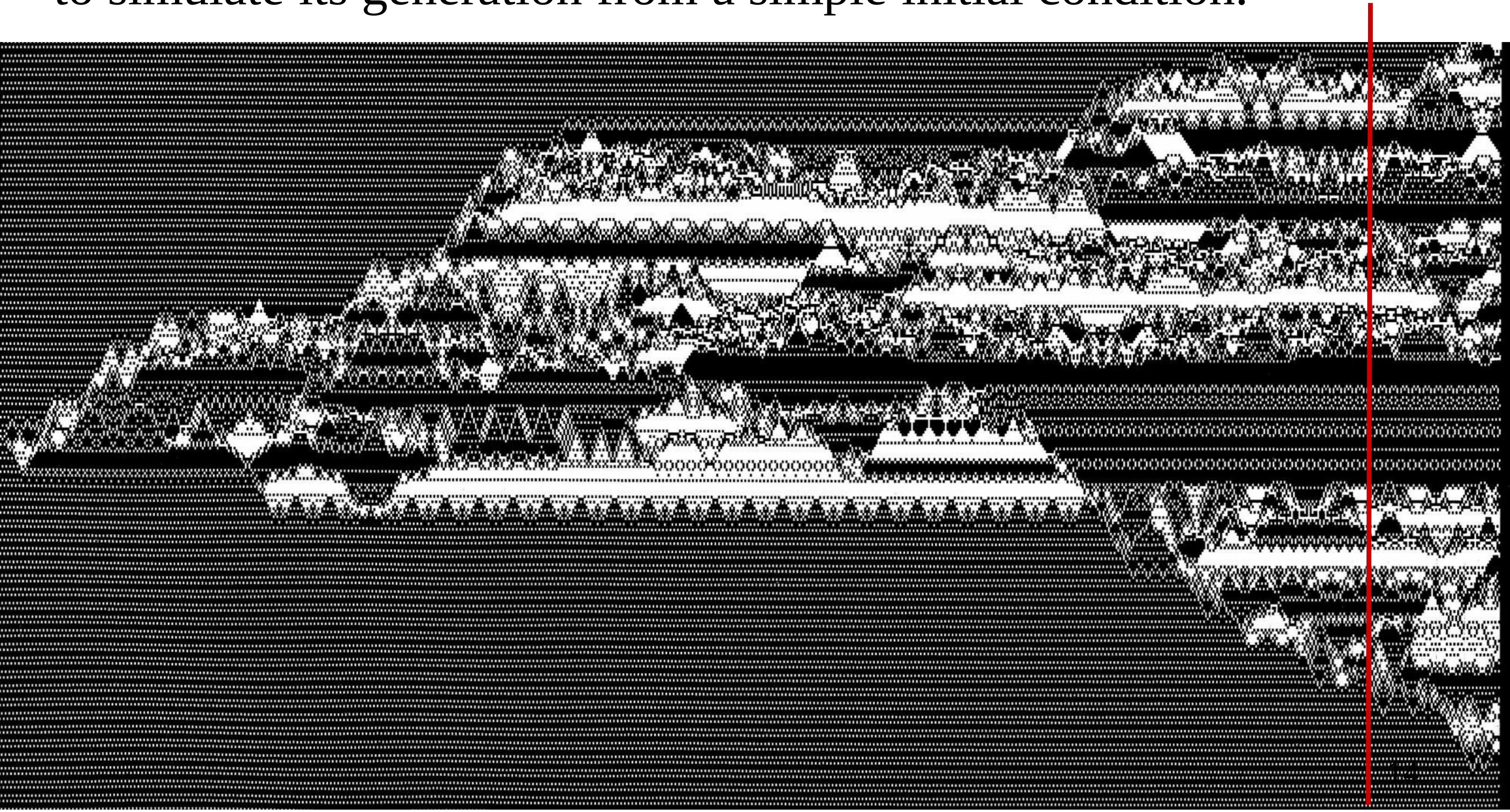
A trivially orderly sequence like 111111… is logically shallow because it can be computed rapidly from a short description.

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.

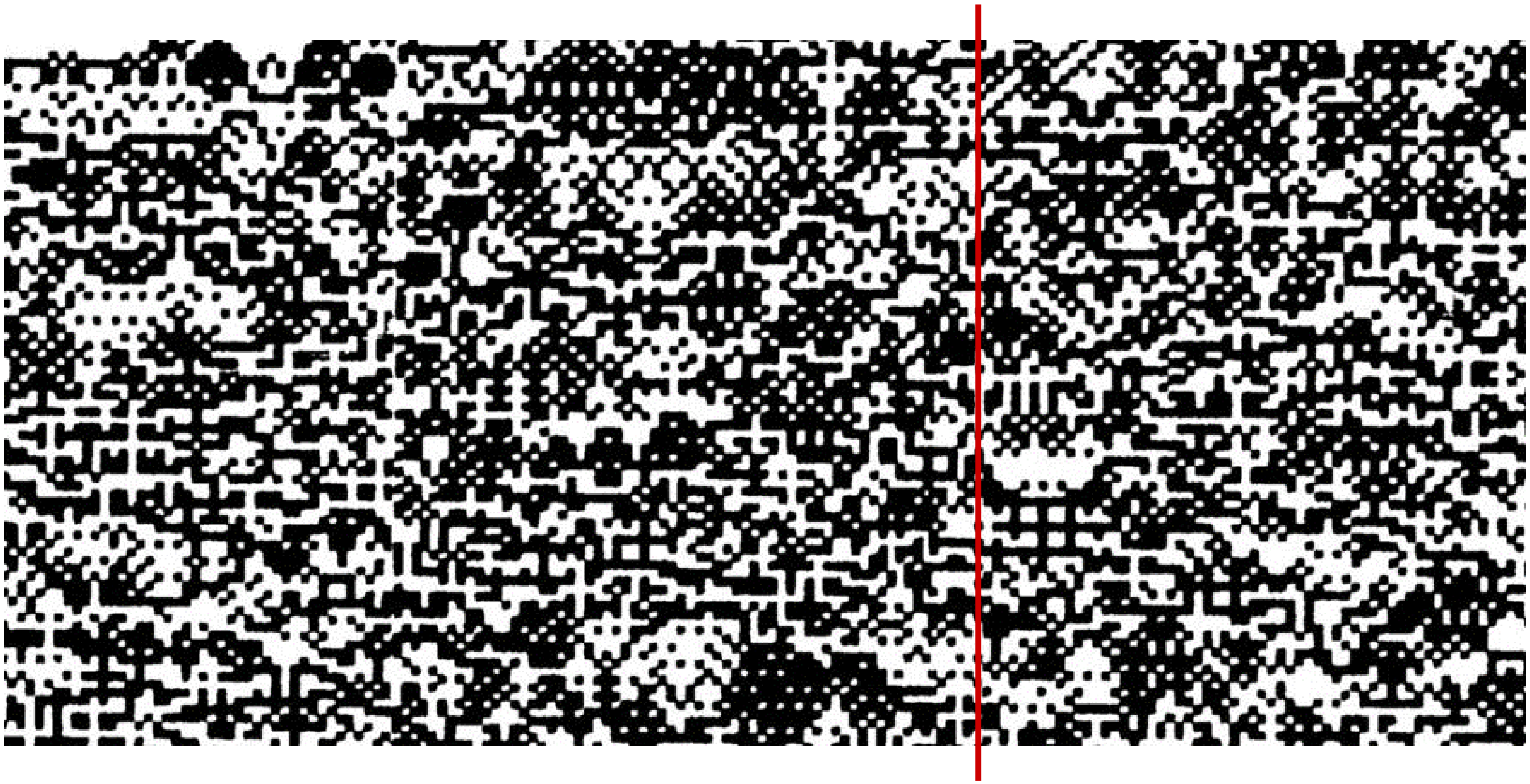
Trivial semi-orderly sequences, such as an alternating sequence of 0's and random bits, are also shallow, since they are rapidly computable from their random part.

(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. Algorithmic information measures a sequence's 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 run 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 for this time slice does not work by retracing its actual long history, but rather a short computation short-circuiting it.

Why is the true history no longer plausible?

Because to specify the state via a simulation of its actual history would involve naming the exact **number** of steps to run the simulation.

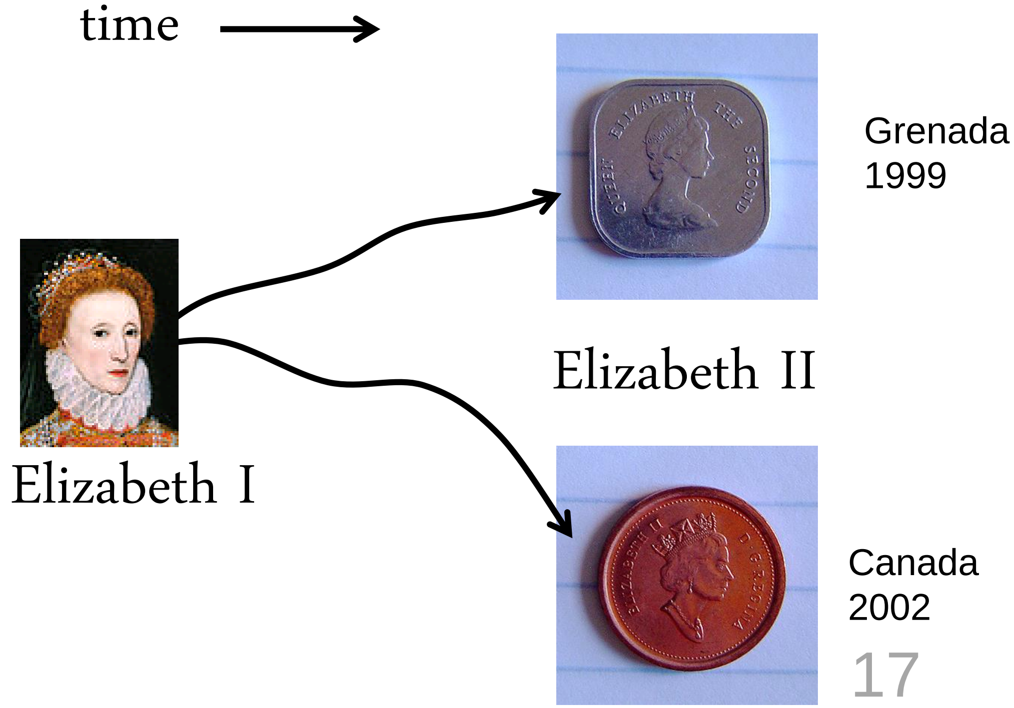
This number is typically very large, requiring about n bits to describe.

Therefore the actual history is no more plausible (in terms of Occam’s razor) than a “print program” that simply outputs the state from a verbatim description.

In a world at thermal equilibrium, with local interactions, correlations are generically local and mediated through the present.

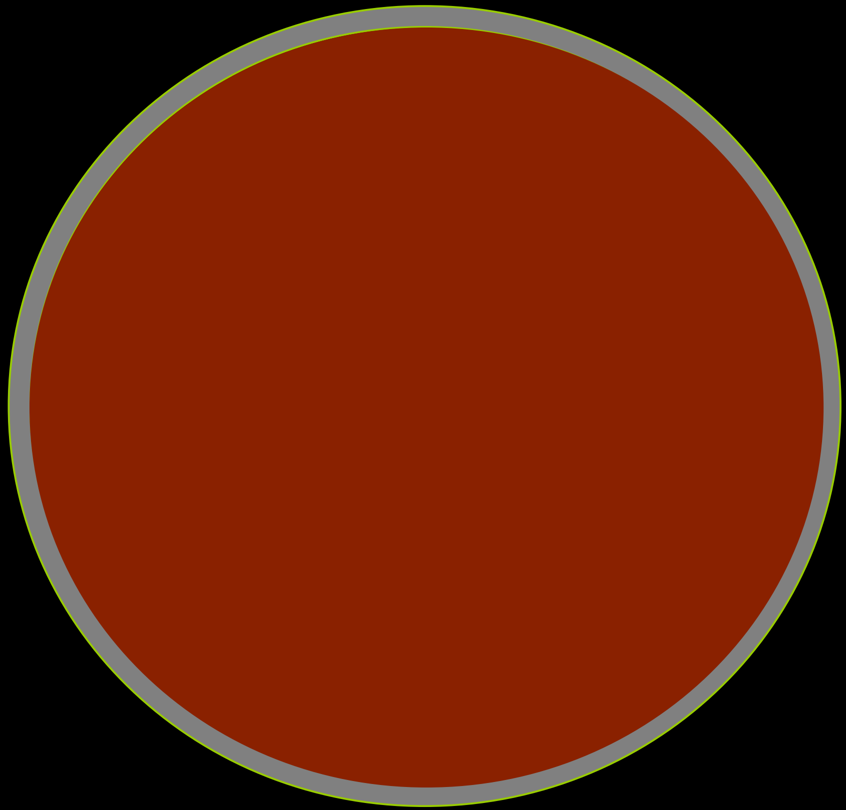
Correlations mediated through present only

By contrast, in a non-equilibrium world, local dynamics can generically give rise to long range correlations, mediated through a V-shaped path in space-time representing a common history.



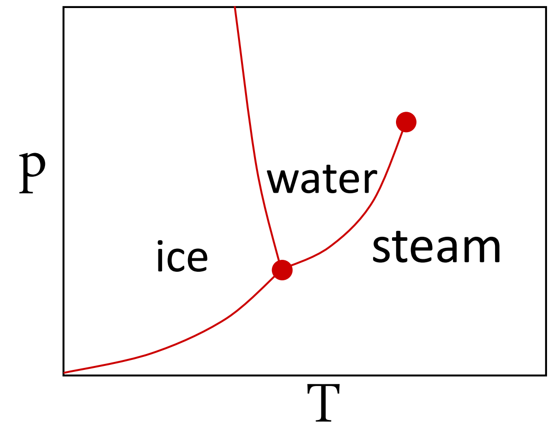
The cellular automaton is a classical toy model, but real systems with fully quantum dynamics behave similarly, losing their complexity, their long-range correlations and even their classical phenomenology as they approach equilibrium.

If the Earth were put in a large reflective box and allowed to come to equilibrium, its state would no longer be complex or even phenomenologically classical.



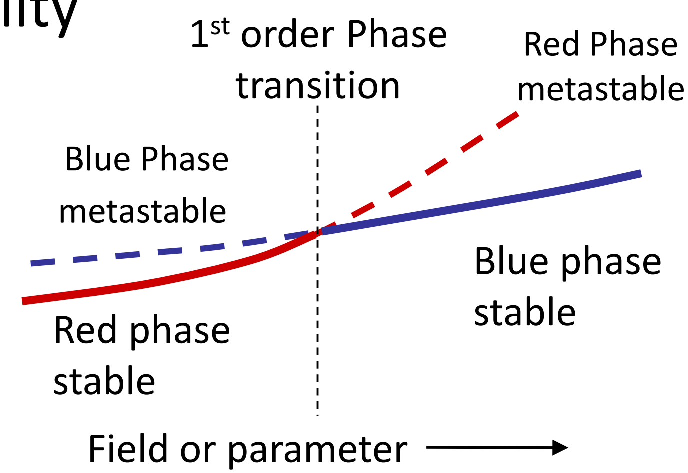
The entire state in the box would be a microcanonical superposition of near-degenerate energy eigenststates of the closed system. Such states are typically highly entangled and contain only short-range correlations.

How strong is the connection between disequilibrium and complexity, in the sense of logical depth?

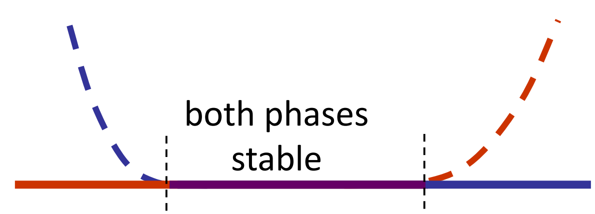


Are thermal equilibrium states generically shallow? Yes, by the Gibbs phase rule.
For generic parameter values, a locally interacting classical system, of finite spatial dimensionality and at finite temperature, relaxes to a unique phase of lowest bulk free energy.

=> no long term memory
=> depth remains bounded in large N limit
(Quantum topological exception, in 3 or more dimensions)



Classical **dissipative** systems can evade the Gibbs phase rule (cf BG85)



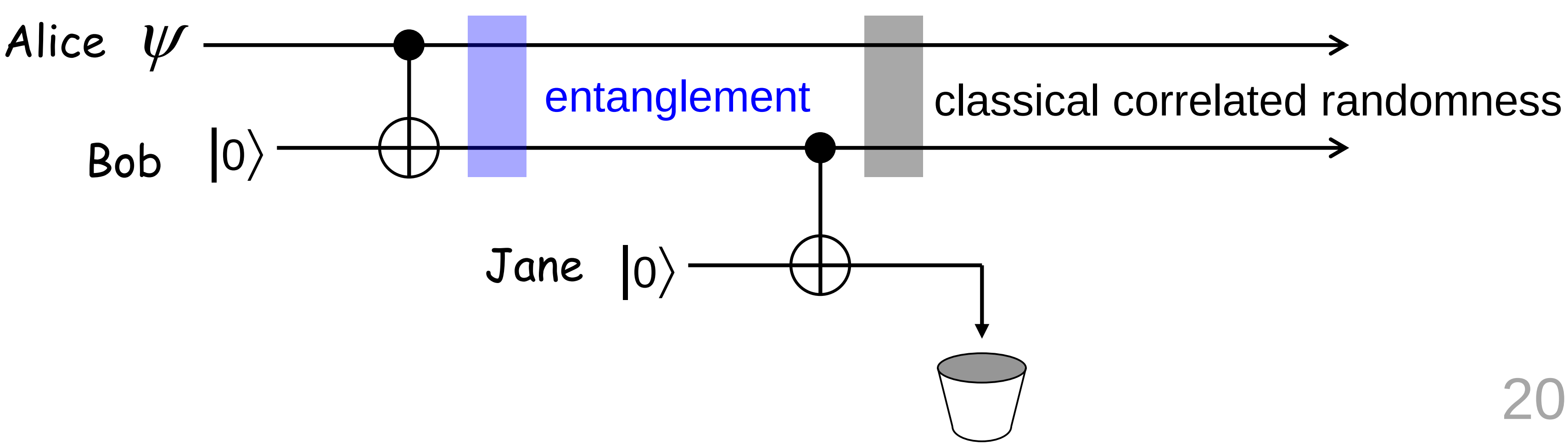
d+1 dimensional "free energy" from system's transfer matrix

Monogamy of Entanglement

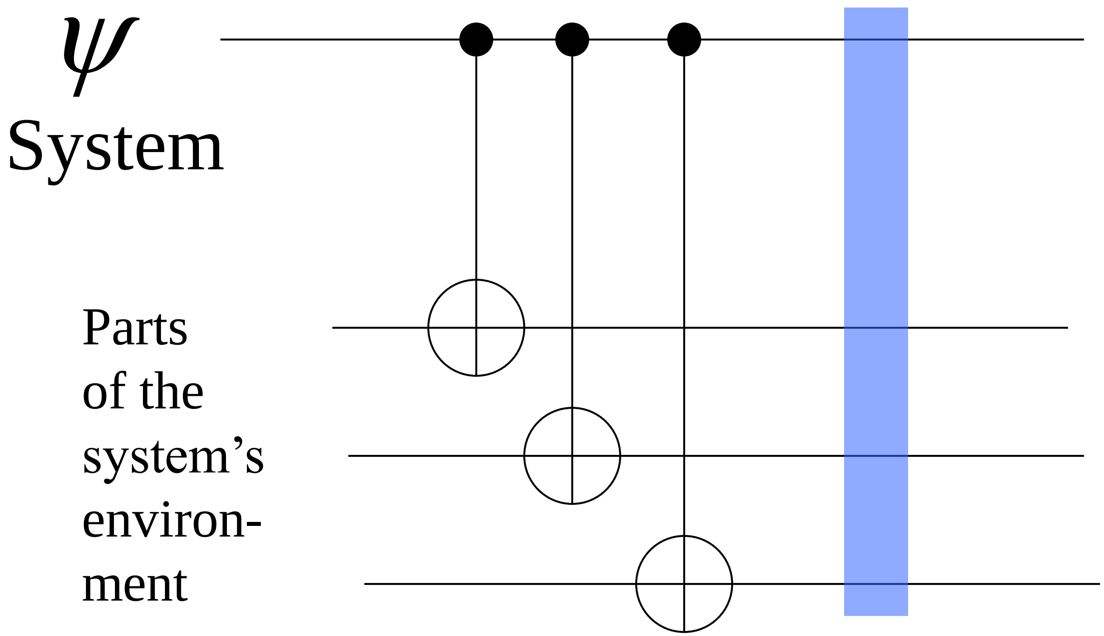
• If A and B are maximally entangled with each other, not only can't they be entangled with anyone else, they can't even be classically correlated with anyone else.

• Indeed, classical correlation typically arises from decoherence, which may viewed as what happens when you try to entangle yourself with multiple other parties.

"Two is a couple, three is a crowd."



How entanglement hides, creating a classical-appearing world



Massive eavesdropping causes the system to get classically correlated with many parts of its environment. But because of ***monogamy,*** *it remains entangled only with the whole environment.*

Information becomes classical by being replicated redundantly throughout the environment. (Zurek, Blume-Kohout et al)

"Quantum Darwinism" Maybe "Quantum Spam" would be a better name.

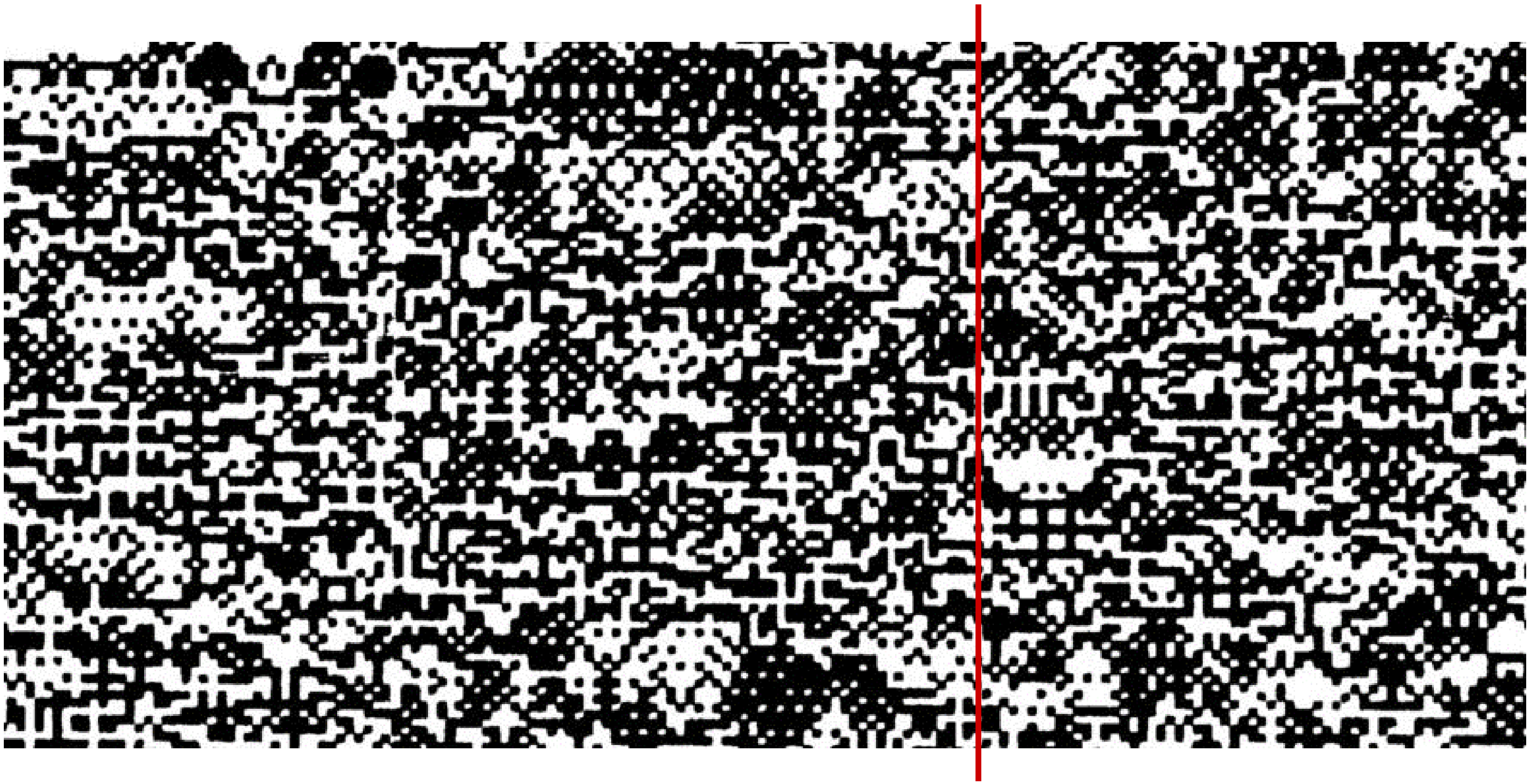
(This typically happens when the environment is not at thermal equilibrium, and when it contains many subsystems that interact more strongly with the system than with each other and… The earth's environment is like that.)

Riedel and Zurek have pointed out the role of non-thermal illumination in creating classical correlations in everyday life, e.g. photons from the sun reflecting off objects on the surface of the Earth to produce massively redundant records of their positions.

If these photons continue to propagate away in free space, the system will never equilibrate and the redundant record will be permanent, though inaccessible, even outliving the Earth.

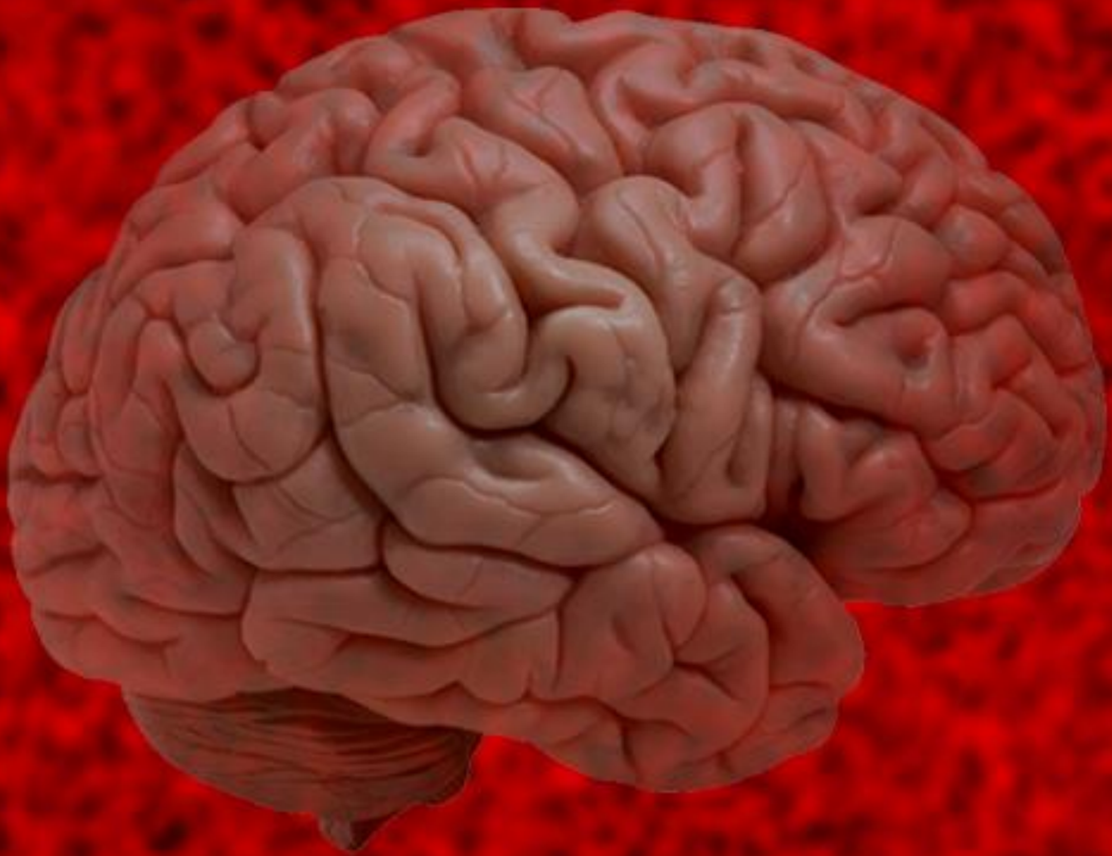
But if the reflected photons were instead trapped inside a reflective box, they would be repeatedly absorbed and re-emitted from the Earth, becoming featureless black body radiation and obfuscating the former redundant correlations, thereby rendering the system no longer classical.

Recall that if a system's dynamics is allowed to run for a long time after equilibration (comparable to the system's Poincaré recurrence time) its actual history can no longer be reliably inferred from its present state.

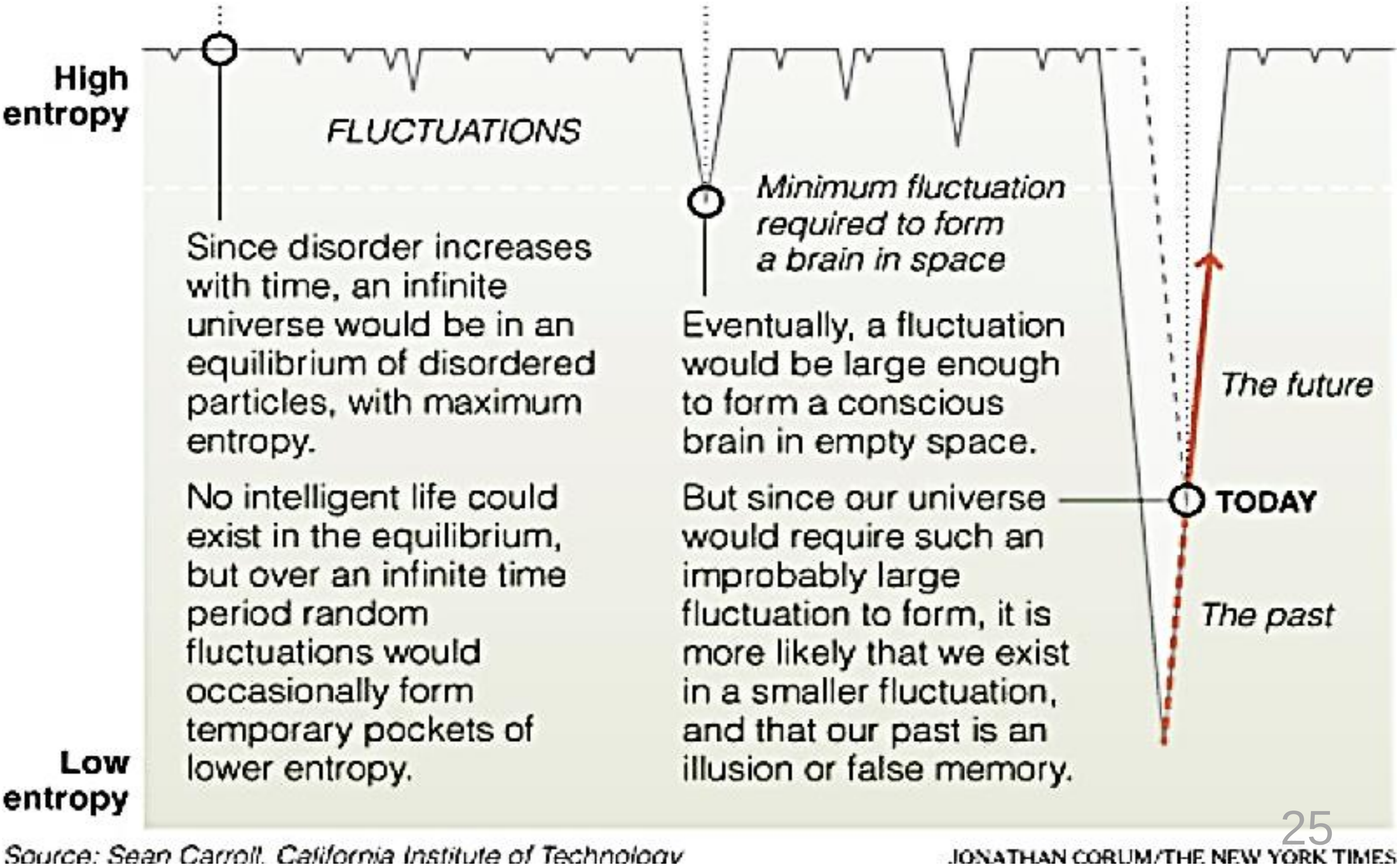


Conversely, a deep structure, one that seems to have had a long history, might just be the result of an unlikely thermal fluctuation, a so-called Boltzmann Brain.

A friend of Boltzmann proposed that the low-entropy world we see may be merely a thermal fluctuation in a much larger universe. “Boltzmann Brain” has come to mean a fluctuation just large enough to produce a momentarily functioning human brain, complete with false memories of a past that didn’t happen, and perceptions of an outside world that doesn’t exist. Soon the BB itself will cease to exist.



A diabolical conundrum: Boltzmann fluctuations nicely explain the low entropy state of our world, and the arrow of time, but they undermine the scientific method by implying that our picture of the universe, based on observation and reason, is ***false.***



Diabolical Conundrum Continued: People began worrying about equilibration in the 19th Century, calling it the "heat death of the universe", but thought of it as a problem for the far future.

Boltzmann showed us that it is already a problem in the present, undermining our ability to make inferences make about conditions in the past or elsewhere, based on those here and now. The inhabitants of any universe that will ultimately equilibrate, either microcanonically or canonically, must make the additional postulate, unsupported by observation, that they are situated atypically early in its history. Otherwise, their "scientific" inferences are no better than those of the inhabitants of Borges' fictional *Library of Babel* (which contained, randomly shelved, one copy of every possible 410 page book).

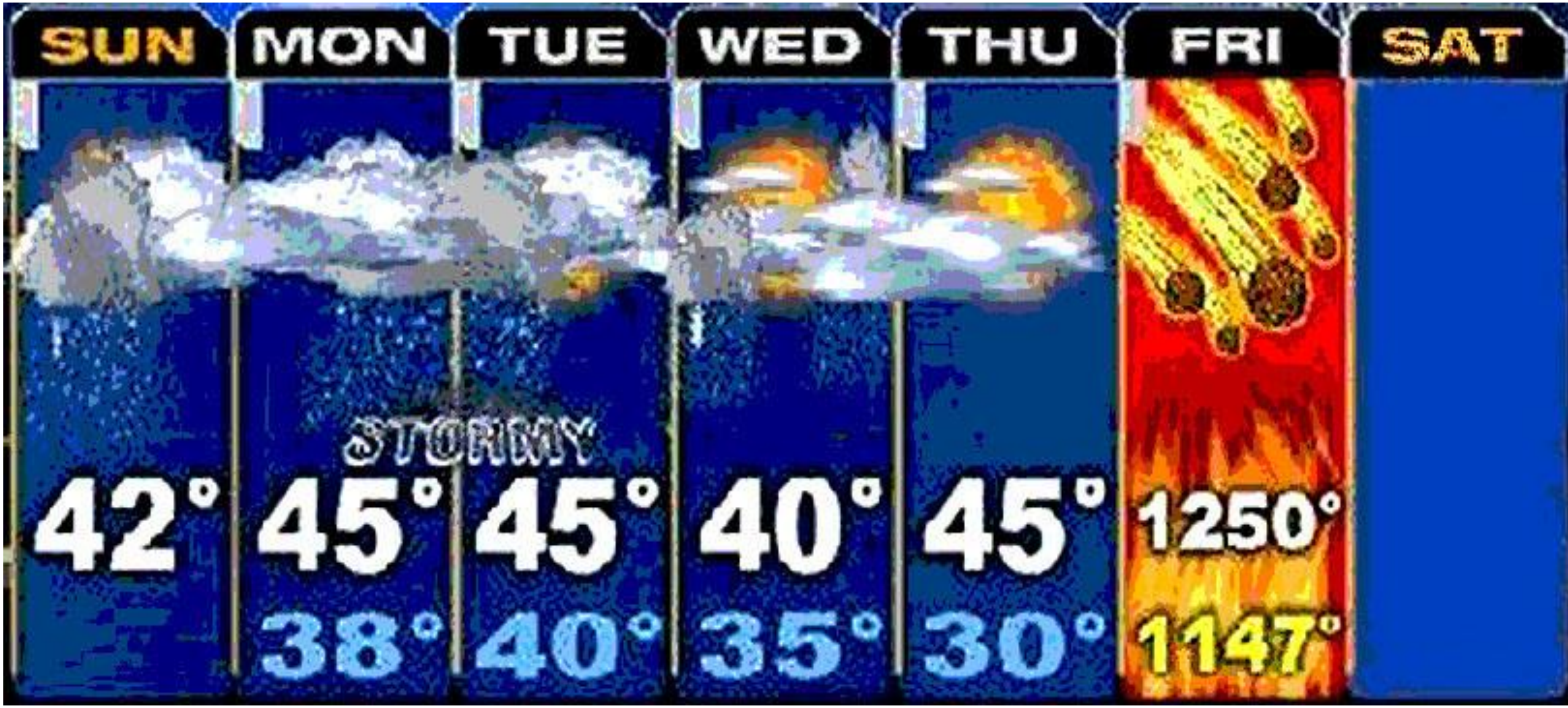
Nowadays serious cosmologists worry about Boltzmann Brains

e.g. arxiv:1308.4686

Can the Higgs Boson Save Us From the Menace of the Boltzmann Brains?

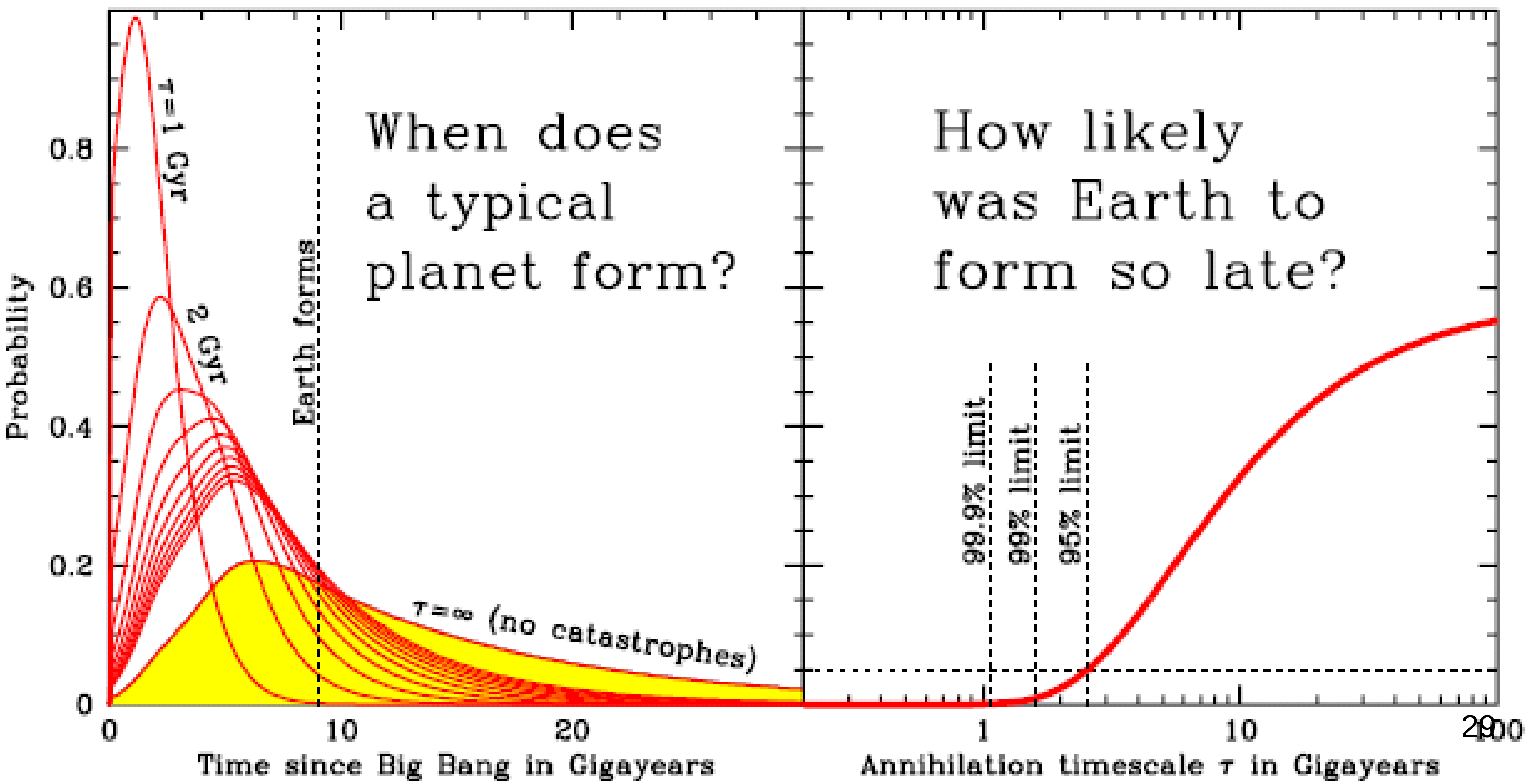
Kimberly K. Boddy and Sean M. Carroll

In other words, current cosmological models predict that the far future of our universe will be an equilibrium thermal state at positive temperature and infinite duration, giving infinitely many opportunities for Boltzmann brains to form. This seems to make it infinitely less likely that we are inhabitants of a young live universe than an old dead one. To forestall this violation of typicality, they propose that the universe will end in around 100 billion years.



Four years ago, superstitious people thought the world would end at the wraparound of the Mayan Calendar. My then 4 year old granddaughter said, “That’s silly. The world isn’t going to end.” Despite this common sense idea, it is tricky to reason about world-ending phenomena that haven’t happened yet, especially ones like Vacuum Phase Transitions that would be too sudden to notice, like dying in one’s sleep.

For example, could it be that apocalypses are intrinsically rather likely, and we've just been extraordinarily lucky so far? Tegmark and Bostrom (Nature 2005, 438, 754) argue **No**, on the grounds that potentially habitable planets were being formed for several billion years before the Earth.



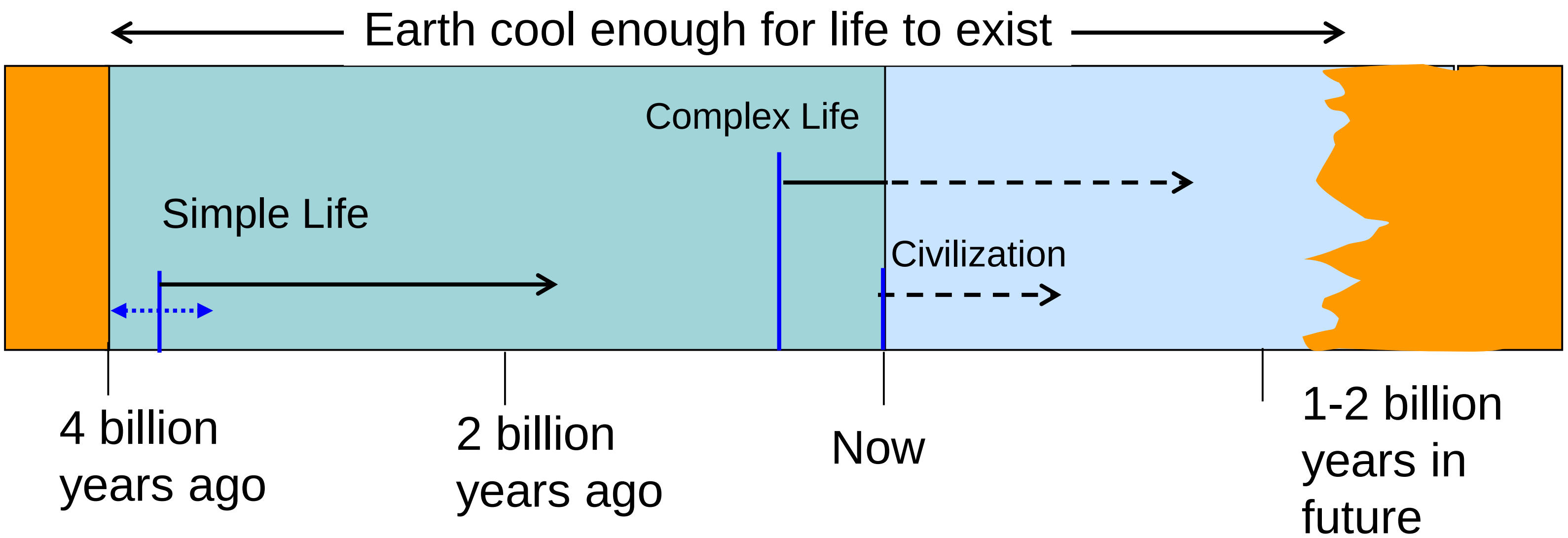
Doomsday arguments illustrate undisciplined thinking based on assumed typicality of the observer, without considering ways in which the observer may be atypical.

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

Carlton Caves' birthday party rebuttal the doomsday argument arxiv:0806.3538: Imagine wandering into a birthday party and learning that the celebrant is 50 years old. Then there is a 1/2 chance they will live to be 100 years old and a 1/3 chance to 150. Conversely, upon encountering a one day old baby, would it be fair to warn the parents that their child will probably only live a few weeks?

In both cases the person's body contains internal evidence of their life expectancy that invalidates the assumption of typicality.

A more severe doomsday question occurs in connection with *civilization*, which has existed only a few *millionths* of the time potentially available for it (e.g. before the sun gets too hot).



Why is civilization so atypically new?

- **VPTs?** No. By Tegmark and Bostrom's argument (if you believe it), VPTs don't happen often enough to explain such extreme newness.
- **Intrinsic Instability?** Maybe civilization, especially technological civilization, is unstable, tending to destroy itself within a few thousand years.
    - Why can't we protect ourselves from this, e.g. by becoming more peaceful and cooperative, or colonizing space?
    - Why don't we see the remains of previous civilizations? Maybe they're too rare, less than 1 per galaxy, which would also explain Fermi's paradox (the lack of contact with extraterrestrials).
- **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 doomsday reasoning life ***as they know it*** may be about to disappear. (Cf. David Deutsch "The Beginning of Infinity")

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

A building, dating from Year VII of the Fascist Era (1922-43), which turned out to be less atypical than Mussolini hoped.

Returning to the more pessimistic hypothesis of self-destruction, Arthur Schopenhauer made perhaps the first anthropic argument in his rebuttal of Leibniz' "best of all possible worlds." He argued that instead we should expect to find ourselves in the worst of all possible worlds. By this he meant not a world full of nastiness and evil, but one on the brink of self-destruction:

"…individual life is a ceaseless battle for existence itself; while at every step destruction threatens it. Just because this threat is so often fulfilled provision had to be made, by means of the enormous excess of the germs, that the destruction of the individuals should not involve that of the species, for which alone nature really cares. The world is therefore as bad as it possibly can be if it is to continue to be at all. Q. E. D. The fossils of the entirely different kinds of animal species which formerly inhabited the planet afford us, as a proof of our calculation, the records of worlds the continuance of which was no longer possible, and which consequently were somewhat worse than the worst of possible worlds." 1844

Schopenhauer’s anthropic principle can be viewed as a natural manifestation of high-dimensional geometry:

Almost all the volume of a high-dimensional spherical ball is very near the surface; therefore almost all possible worlds will be near the boundary of instability.

An even more pessimistic notion is that our world is well beyond the boundary of spontaneous stability, and we are only here because we have been atypically lucky, like a pencil that has balanced on its point for hours in a moving train.

This is the question of finite versus infinite fine-tuning in cosmology.

Infinite fine tuning undermines science in much the same way as Boltzmann brains.   Can we a devise a plausible self-organizing cosmology that does not equilibrate and requires only finite fine tuning?

Wigner's Friend

Schrödinger's infamous cat is in a superposition of alive and dead before the box is opened.

Eugene Wigner imagined a gentler experiment, relevant to the Quantum Boltzmann Brain problem:

Wigner's friend performs a quantum measurement with two outcomes but only tells Wigner what happened later.

After the experiment, but before Wigner hears the result, Wigner regards his friend as being in a superposition of two states, but the friend perceives only one or the other of them.

In principle (and even in practice, for atom-sized friends) Wigner can contrive for the friend to undo the measurement and forget its result—a "quantum eraser" experiment.

Wigner's friend might have been viewed as no more than a philosophical conundrum, but it is relevant to the anthropic counting of observers.

In a 2014 sequel to their 2013 paper, Boddy and Carroll, joined by Pollack, argue that it is not necessary for the universe to self-destruct to avoid the menace of Boltzmann brains. They instead argue that the late thermal state of the universe doesn't generate any Boltzmann brains because there is no mechanism to **observe** them, in the strong sense of making a permanent external classical record.

But as Jess Riedel and I have argued, all our experience, like that of Wigner's friend, is potentially impermanent. Therefore I think it is unreasonable to insist that nothing happens until a permanent record of it is made. Moreover observership, in the anthropic sense, is an introspective property of a system, not a property of how it would behave if measured externally.

If a piece of our universe, centered on the sun, were put in a box with perfectly reflective walls, 1 million light years in diameter, it would take us a half a million years to notice any difference. Yet the long term evolution of this isolated system would be radically different from the evolution of the universe we believe we inhabit, lacking this box. The boxed universe would recur repeatedly to near its initial state, and, exponentially more frequently, to Boltzmann brain states, where the recurrence would be confined to a solar-system sized patch near the center, with the remaining volume being thermal and uncorrelated. Nevertheless, the central region would match the solar system as it is now, with all its classical equipment and storage media recording evidence of its supposed multi-billion-year history and the results of recent experiments, and conscious beings having thoughts like ours. So unless one is willing to push the moveable quantum-classical boundary out indefinitely far out, this system would experience what we experience now, but on its orbit false local recurrences would vastly outnumber true ones.

Similarly, we argue, in the thermal de Sitter state of an unboxed universe, false local recurrences would vastly outnumber full recurrences, and these would infinitely outnumber the single first-time occurrence of our solar system in the young expanding universe.

To think about this, it helps to review some basic facts about entanglement and quantum mixed states:

- A mixed state is completely characterized by its density operator ρ, which describes all that can be learned by measuring arbitrarily many specimens of the state. For an ensemble of pure states $\{p_j, \psi_j\}$, ρ is given by the weighted sum of the projectors onto these states.
- Ensembles with the same ρ are indistinguishable.
- A system **S** in a mixed state $\rho^{\mathbf{S}}$ can, without loss of generality, be regarded as a subsystem of a larger bipartite system **RS** in a pure state $\Psi^{\mathbf{RS}}$, where R denotes a non-interacting reference system.
- "Steering" Any ensemble $\{p_j, \psi_j\}$ compatible with ρ can be remotely generated by performing measurements on the R part of $\Psi^{\mathbf{RS}}$. Measurement outcome *j* occurs with probability p_j, leaving S in state ψ_j.

Jess Riedel's scenario suggesting why Boltzmann brains ought to be present in thermal states at any positive temperature, even though there is no external observer.

- Let π_{BB} be a projector onto some state representing a fluctuation, for example a copy of the Solar System pasted into a much larger patch of de Sitter vacuum.
- Any finite temperature thermal state $\boldsymbol{\rho}$ of this patch can be expressed as a weighted sum

$$\rho = \lambda\, \pi_{BB} + (1-\lambda)\, \boldsymbol{\sigma}$$

where $\boldsymbol{\sigma}$ is a thermal state "depleted" in $\boldsymbol{\pi}_{BB}$.

- An all-powerful Preparator tosses a λ-biased coin, and prepares π_{BB} or $\boldsymbol{\sigma}$ according to the outcome.
- Before departing, the Preparator takes away, in reference system **R**, a record of all this, including, for example, souvenir photos of the just-created Earth and its inhabitants.

Since this is a valid preparation of the thermal state, and keeping in mind that it is impossible in principle to distinguish different preparations of the same mixed state, it is hard to see why the inhabitants of the de Sitter patch do not have some small probability of experiencing a life resembling our own, at least for a while.

Jason Pollack's reply to this argument: their 2014 paper, alleging the absence of such fluctuations, does not apply to all thermal states, but only those purified by a reference system **R** of a particular form, so that state $\Psi^{\mathbf{RS}}$ is a Bunch-Davies pure state of the universe whose local patches $\rho^{\mathbf{S}}$ are all in thermal de Sitter states.

This may be viewed as an Occam-type argument from simplicity, favoring simplicity not of the accessible system **S**, but of the inaccessible purifying system **R**.

Internal vs External views: Our suggested internal criterion for a state $\boldsymbol{\rho}$ to have nonzero participation of a Boltzmann brain state π_{BB}, namely

$$\exists \boldsymbol{\sigma}, \lambda > 0: \quad \rho = \lambda\, \pi_{BB} + (1-\lambda)\, \boldsymbol{\sigma}$$

is more restrictive than the usual criterion that $\boldsymbol{\rho}$ have a positive expectation when subjected to an external measurement of π_{BB}, namely,

$$\mathrm{tr}(\rho\, \pi_{BB}) > 0.$$

Even a zero temperature vacuum state (the Lorentz vacuum) would have a positive Boltzmann brain probability when measured externally. The energy for creating the Boltzmann brain out of the ground state would come from the measuring apparatus. This is a further reason we think an external measuring apparatus is an encumbrance in a cosmological setting, when reasoning about a system’s internal experiences.

Open questions

- Wigner's Friend's experiences, if any
- Are there plausible cosmologies (e.g. eternal inflation) that never equilibrate, and thus avoid the Boltzmann brain problem?
- Does entanglement enable a non-dissipative route to fault-tolerant self-organization?
- Conjectured emergence of spacetime from entanglement, or vice-versa (e.g. "ER=EPR")

Workshop on "Quantum Foundations of a Classical Universe," IBM Research Aug 11-14, 2014
http://www.jessriedel.com/conf2014/conf2014.html or
http://researcher.watson.ibm.com/researcher/view_group.php?id=5661

C. J. Riedel and W. H. Zurek, "Quantum Darwinism in an Everyday Environment: Huge Redundancy in Scattered Photons," *Phys. Rev. Lett.* **105**, 020404 (2010). [arXiv:1001.3419] cf also longer treatment in [arxiv:1102.31793v3]

C.J. Riedel, Classical branch structure from spatial redundancy in a many-body wave function, arXiv:1608.05377.

C.H. Bennett blog post on logical depth versus other complexity measures
http://dabacon.org/pontiff/?p=5912

CH Bennett, blog post on *Schopenhauer and the Geometry of Evil,*
https://quantumfrontiers.com/2016/05/29/schopenhauer-and-the-geometry-of-evil/

C.H. Bennett "Logical Depth and Physical Complexity" in *The Universal Turing Machine– a Half-Century Survey*, edited by Rolf Herken Oxford University Press 227-257, (1988)
http://researcher.ibm.com/researcher/files/us-bennetc/UTMX.pdf

C.H. Bennett and G. Grinstein "On the Role of Dissipation in Stabilizing Complex and Non-ergodic Behavior in Locally Interacting Discrete Systems" *Phys. Rev. Lett.* **55,** 657-660 (1985).
http://researcher.ibm.com/researcher/files/us-bennetc/BG85%20with%20Toom%20snapshotsq.pdf

Peter Gacs, "Reliable Computation with Cellular Automata" *J. Computer and System Science* **32**, 15-78 (1986) http://www.cs.bu.edu/~gacs/papers/GacsReliableCA86.pdf

With happy memories of HRI and of YouQu 2015, which I was privileged to attend in person,

Thank you,

and best wishes for YouQu 2017

Extra slides

Classical **dissipative** systems can evade the Gibbs phase rule, storing information indefinitely and performing error-correcting computations despite hostile noise.

VOLUME 55, NUMBER 7 PHYSICAL REVIEW LETTERS 12 AUGUST 1985

phase region is a narrow metastable zone (demarcated on the right by dashed lines), in which large islands of the favored phase grow but small islands shrink. A critical exponent of 3.0 ± 0.4 was found for the vertical width of the two-phase region as a function of noise amplitude $(p+q)$ below the critical point. Other runs on one-phase systems with unbiased noise $p=q$ yielded the value $\beta = 0.122 \pm 0.01$ for the exponent describing magnetization as a function of noise amplitude below the critical point.

Besides being irreversible, the NEC model differs from conventional kinetic Ising models in having synchronous updating. However, preliminary runs in

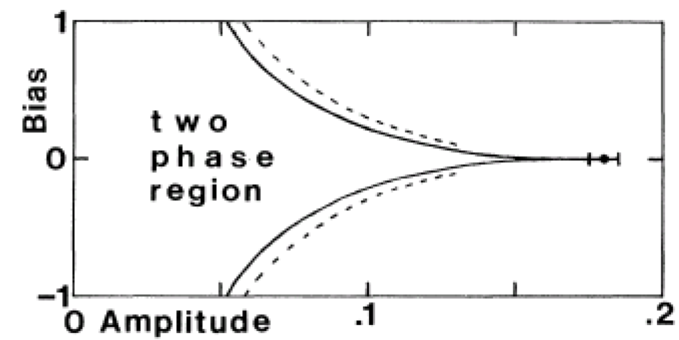
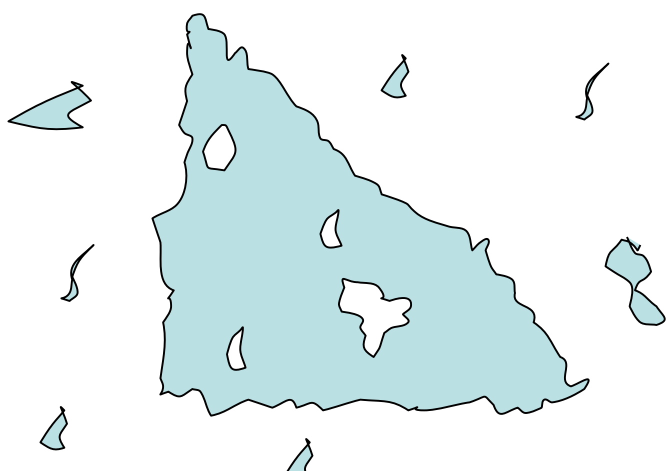
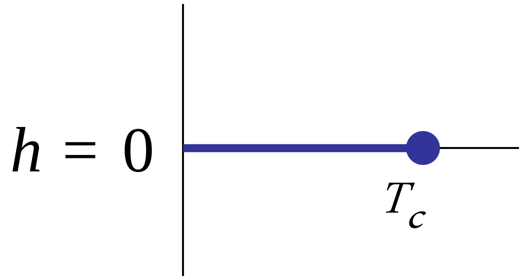


FIG. 1. Phase diagram of the NEC system, for noise parameters p and q, with amplitude $= p+q$ and bias $= (p-q)/(p+q)$.

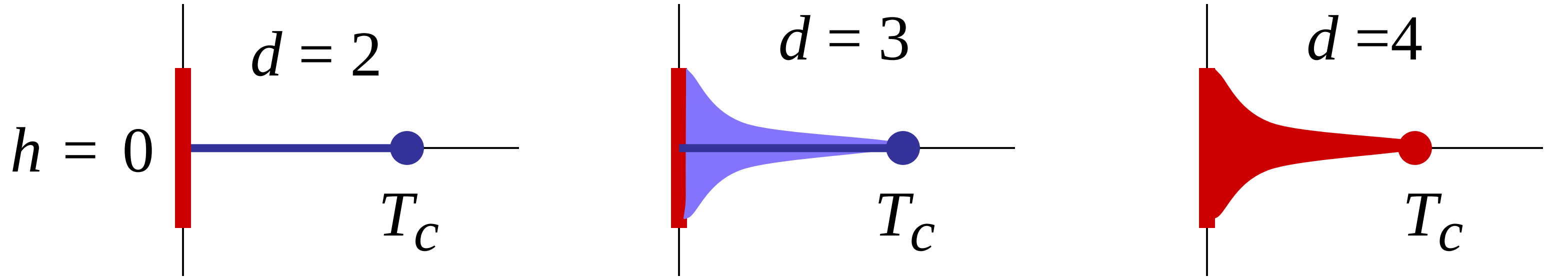


Toom’s NEC rule, a dissipative noisy cellular automaton rule, *Future := Maj{North, East, Center} +Noise,* in contrast to Ising model, is stable against symmetry-breaking field in 2d, and gives Gacs- Reif fault tolerant cellular automaton in 3D



Phase Diagram of Classical Ising model in $d > 1$ dimension. Stores a classical bit reliably when h=0 and $T<T_c$

Phase diagrams for local quantum models (Toric code)



Degenerate ground state stores a qubit reliably at T=0, even for nonzero h. For T>0, stores a bit reliably only at h=0

Stores a qubit at T=0. For T>0, can store a quantum-encoded classical bit, even when h is nonzero

Stores a quantum-encoded qubit even at nonzero T and h.

Cf S. Chesi, D. Loss, S. Bravyi, and B. Terhal, arXiv:0907.2807

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"

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 if this doesn't work, due to the above "measure problem", could we not fall back on defining the set of "all possible universes" in a purely mathematical way, untainted by physics?

Yes– use the universal probability defined by the Monkey Tree, despite its being only semicomputable. (cf Juergen Schmidhuber *Algorithmic Theories of Everything* arXiv:quant-ph/0011122)

But that gives too easy an answer to the question of Self-organization: By virtue of its computational universality, a positive measure fraction of the Monkey Tree is devoted to self-organizing behavior, according to any computable definition thereof.

But before going so far, do we want to include any "universal" *physical* principles in the universal prior?

- Reversibility? (very physical, but tends to lead to equilibrium)
- Superposition – quantum mechanics
- Locality / field theories? (Lloyd and Dryer 's universal path integral arxiv:1302.2850)
- Fault-tolerance, stability w.r.t.
  - Noise = positive temperature
  - Variation of the model's continuous parameters, e.g. interaction energies, transition probabilities

Conway's game of life is irreversible, computationally universal, but doesn't look very physical or noise-tolerant

The 1-d Ising cellular automaton shown earlier is reversible, looks to be computationally universal, but is not noise-tolerant

Gacs' 1-d probabilistic cellular automaton is irreversible (does not obey detailed balance) but is universal and fault tolerant

Probabilistic cellular automata that are irreversible (i.e. do not obey detailed balance) are reasonable models for parts of the universe, such as our earth, with equilibration-preventing environments, environments that keep them classical (in the quantum Darwinism sense), or universes that have a live youth and a cold dead old age, preventing Boltzmann fluctuations.

Peter Gacs has shown that there are automata of this sort even in one dimension that are computationally universal, noise-tolerant (all local transition probabilities positive) and stable with respect to generic small perturbations of these transition probabilities. Moreover they can self-organize into a hierarchically encoded computation starting from a translationally invariant initial condition. The encoded computation receives its input via the transition probabilities, and is stable with respect to small perturbations of them. (cf Gacs 1985 JCSS paper and 2014 workshop talk)

To make the quantitative definition of logical depth more stable with respect small variations of the string *x* and the universal machine *U*, the definition needs to be refined to take weighted account of all programs for computing the object, not just the smallest.

The *s*-significant depth of a string *x*, denoted $D_s(x)$, is defined as the least run time of any *s*-incompressible program to compute *x*:

$$D_s(x) = \min\{T(p): U(p)=x \;\&\; |p|-|p^*|<s\}.$$

Here *p* ranges over bit strings treated as self-delimiting programs for the universal computer *U*, with $|p|$ denoting the length of *p* in bits, and p^* denoting the minimal program for *p*, i.e. $p^* = \min\{q: U(q)=p\}$.

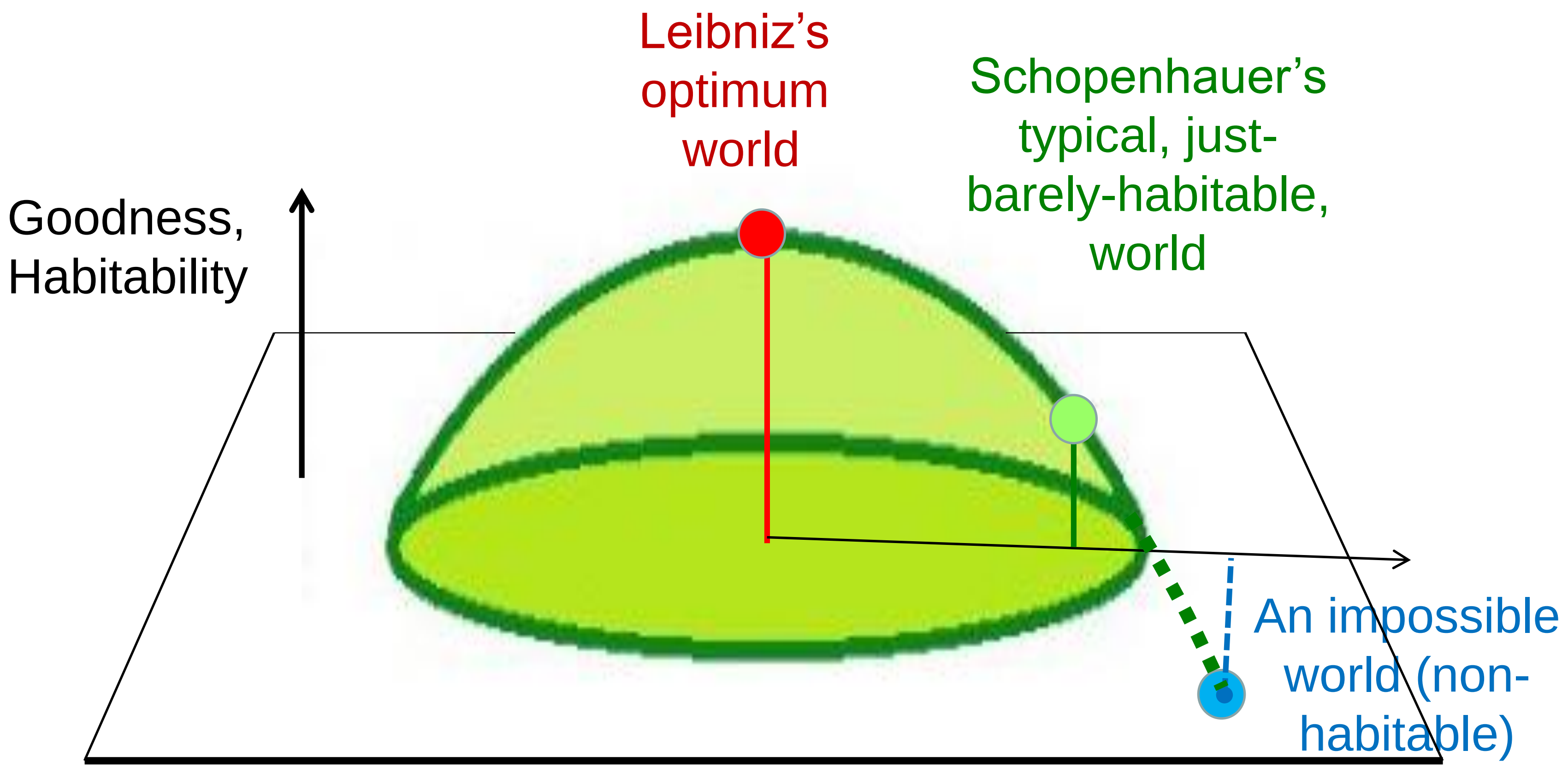
This formalizes the notion that all hypotheses for producing *x* in fewer than *d* steps suffer from at least *s* bits worth of ad-hoc assumptions. Informally, this means they suffer from at least *s* bits worth of Donald-Duckness.

What Wigner's friend thinks before he tells Wigner

Some believe that nothing can be said to classically happen until it is fixed by an act of irreversible measurement. I dislike this view because everything, even our own experiences, can potentially be undone.
For example, if our galaxy were enclosed in a reflective box, it would eventually reach thermal equilibrium and lose its classical character.

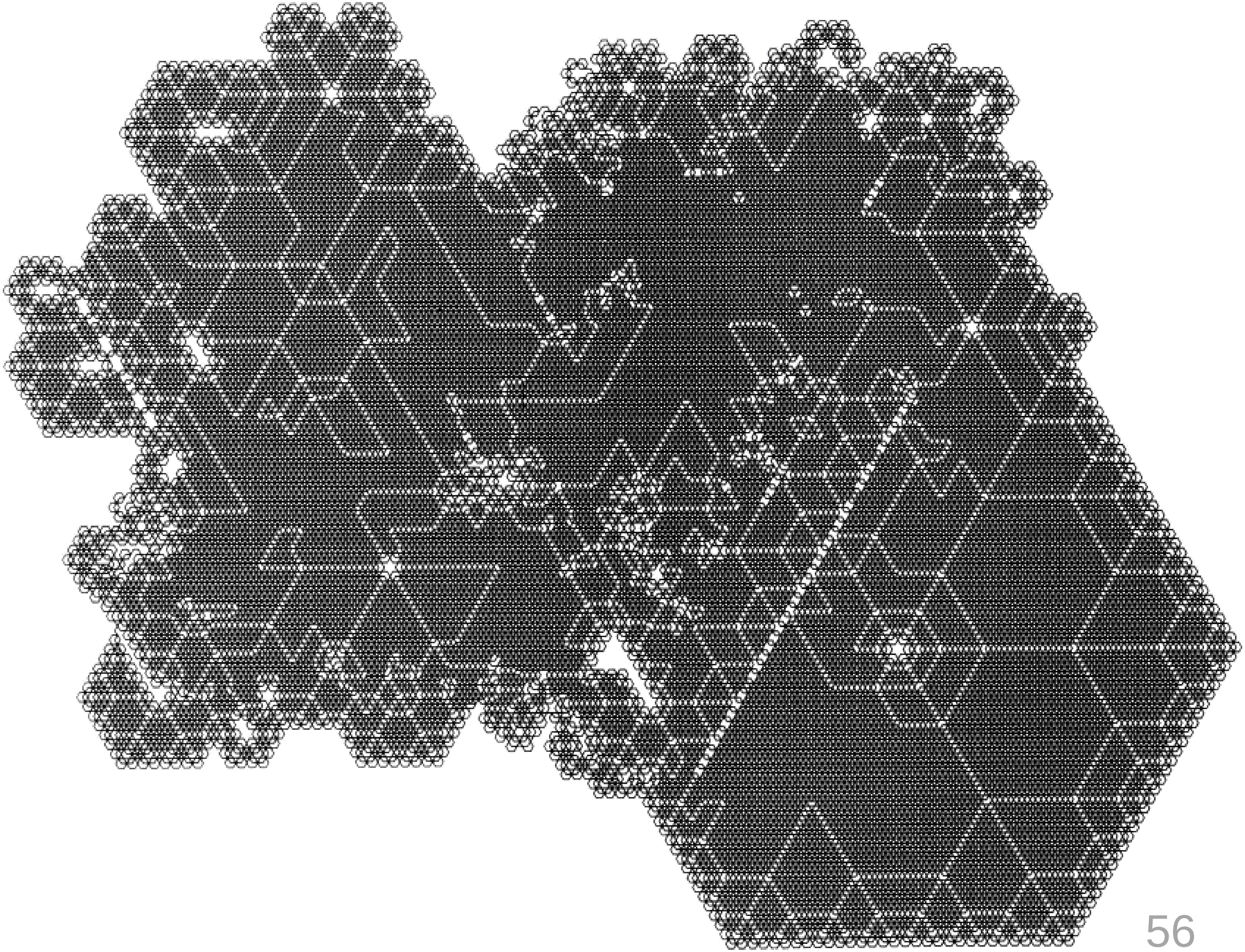
Riedel and Zurek have pointed out the important role of illumination by a non-equilibrium radiation field in bringing about decoherence and classicization in everyday life. Therefore Riedel and I propose to characterize the emergence and decay of classical correlations in a simplified three-stage scenario:

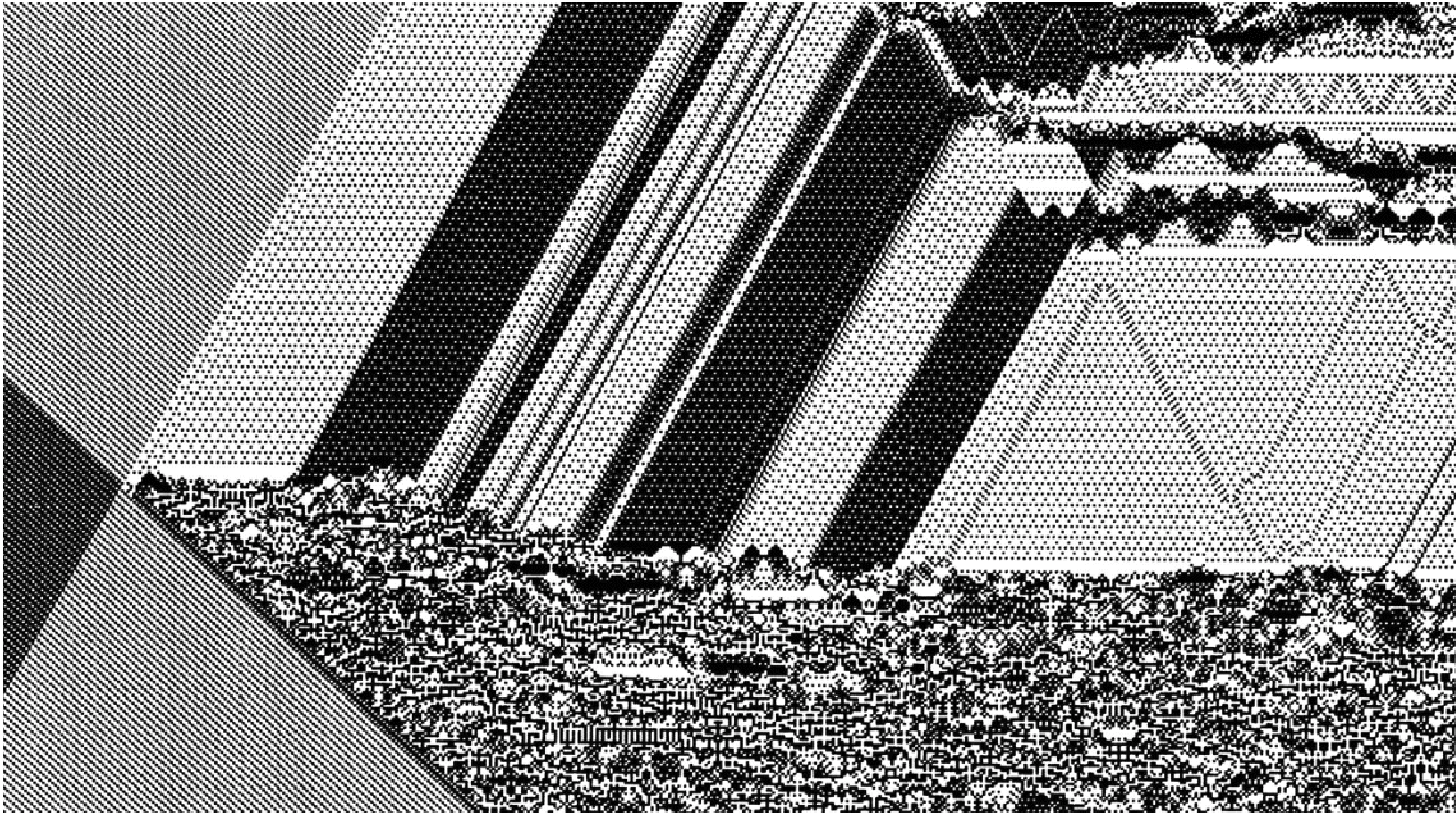
- A macroscopic or mesoscopic system S, such as a dust grain, is initially at thermal equilibrium at temperature T1.
- It is placed in box filled with thermal radiation at temperature T2.
- Eventually box and system come to equilibrium at an intermediate temperature.



In a parameter space of *d* dimensions (here only two are shown), the average goodness of worlds with positive goodness is only 2/(d+2) of Leibniz's optimum, tending to zero (just barely habitable) for large *d.* Non-habitable worlds (blue) have negative goodness, and by anthropic selection are excluded from the sample.

I cannot ascribe to the Théodicée as a methodical and broad unfolding of optimism, any other merit than this, that it gave occasion later for the immortal “Candide” of the great Voltaire; whereby certainly Leibniz s often-repeated and lame excuse for the evil of the world, that the bad sometimes brings about the good, received a confirmation which was unexpected by him… But indeed to the palpably sophistical proofs of Leibniz that this is the best of all possible worlds, we may seriously and honestly oppose the proof that it is the worst of all possible worlds. For possible means, not what one may construct in imagination, but what can actually exist and continue. Now this world is so arranged as to be able to maintain itself with great difficulty; but if it were a little worse, it could no longer maintain itself. Consequently a worse world, since it could not continue to exist, is absolutely impossible: thus this world itself is the worst of all possible worlds. For not only if the planets were to run their heads together, but even if any one of the actually appearing perturbations of their course, instead of being gradually balanced by others, continued to increase, the world would soon reach its end. Astronomers know upon what accidental circumstances principally the irrational relation to each other of the periods of revolution this depends, and have carefully calculated that it will always go on well; consequently the world also can continue and go on. We will hope that, although Newton was of an opposite opinion, they have not miscalculated, and consequently that the mechanical perpetual motion realised in such a planetary system will not also, like the rest, ultimately come to a standstill. Again, under the firm crust of the planet dwell the powerful forces of nature which, as soon as some accident affords them free play, must necessarily destroy that crust, with everything living upon it, as has already taken place at least three times upon our planet, and will probably take place oftener still. The earthquake of Lisbon, the earthquake of Haiti, the destruction of Pompeii, are only small, playful hints of what is possible. A small alteration of the atmosphere, which cannot even be chemically proved, causes cholera, yellow fever, black death, &c., which carry off millions of men; a somewhat greater alteration would extinguish all life. A very moderate increase of heat would dry up all the rivers and springs. The brutes have received just barely so much in the way of organs and powers as enables them to procure with the greatest exertion sustenance for their own lives and food for their offspring; therefore if a brute loses a limb, or even the full use of one, it must generally perish. Even of the human race, powerful as are the weapons it possesses in understanding and reason, nine-tenths live in constant conflict with want, always balancing themselves with difficulty and effort upon the brink of destruction. Thus throughout, as for the continuance of the whole, so also for that of each individual being the conditions are barely and scantily given, but nothing over. The individual life is a ceaseless battle for existence itself; while at every step destruction threatens it. Just because this threat is so often fulfilled provision had to be made, by means of the enormous excess of the germs, that the destruction of the individuals should not involve that of the species, for which alone nature really cares. The world is therefore as bad as it possibly can be if it is to continue to be at all. Q. E. D. The fossils of the entirely different kinds of animal species which formerly inhabited the planet afford us, as a proof of our calculation, the records of worlds the continuance of which was no longer possible, and which consequently were somewhat worse than the worst of possible worlds.*





Physics of Computation Conference Endicott House MIT May 6-8, 1981

1 Freeman Dyson
2 Gregory Chaitin
3 James Crutchfield
4 Norman Packard
5 Panos Ligomenides
6 Jerome Rothstein
7 Carl Hewitt
8 Norman Hardy
9 Edward Fredkin
10 Tom Toffoli
11 Rolf Landauer
12 John Wheeler

13 Frederick Kantor
14 David Leinweber
15 Konrad Zuse
16 Bernard Zeigler
17 Carl Adam Petri
18 Anatol Holt
19 Roland Vollmar
20 Hans Bremerman
21 Donald Greenspan
22 Markus Buettiker
23 Otto Floberth
24 Robert Lewis

25 Robert Suaya
26 Stan Kugell
27 Bill Gosper
28 Lutz Priese
39 Madhu Gupta
30 Paul Benioff
31 Hans Moravec
32 Ian Richards
33 Marian Pour-El
34 Danny Hillis
35 Arthur Burks
36 John Cocke

37 George Michaels
38 Richard Feynman
39 Laurie Lingham
40 Thiagarajan
41 ?
42 Gerard Vichniac
43 Leonid Levin
44 Lev Levitin
45 Peter Gacs
46 Dan Greenberger